DYNAMIC STRUCTURAL EQUATION MODELS: ESTIMATION AND INFERENCE

Dario Ciraki

London School of Economics and Political Science
University of London

Submitted in Fulfilment of the Requirement
for the Degree of Doctor of Philosophy

Abstract

The thesis focuses on estimation of dynamic structural equation models in which some or all variables might be unobservable (latent) or measured with error. Moreover, we consider the situation where latent variables can be measured with multiple observable indicators and where lagged values of latent variables might be included in the model. This situation leads to a dynamic structural equation model (DSEM), which can be viewed as dynamic generalisation of the structural equation model (SEM). Taking the mismeasurement problem into account aims at reducing or eliminating the errors-in-variables bias and hence at minimising the chance of obtaining incorrect coefficient estimates. Furthermore, such methods can be used to improve measurement of latent variables and to obtain more accurate forecasts. The thesis aims to make a contribution to the literature in four areas. Firstly, we propose a unifying theoretical framework for the analysis of dynamic structural equation models. Secondly, we provide analytical results for both panel and time series DSEM models along with the software implementation suggestions. Thirdly, we propose non-parametric estimation methods that can also be used for obtaining starting values in maximum likelihood estimation. Finally, we illustrate these methods on several real data examples demonstrating the capabilities of the currently available software as well as importance of good starting values.

List of Tables

List of Figures

Contents

Acknowledgements

I would like to express my gratitude to my supervisors, Dr Martin Knott for his support and guidance at all stages of this work and Professor Anders Skrondal for valuable advice and suggestions.

Further acknowledgements go to the participants of the LSE Research Seminars and colleagues from the LSE Statistics Department.

Financial support and contribution from the LSE Teaching Assistantships in 2004–2006 and Department of Statistics Scholarship in 2005 is also acknowledged.

Finally, I would like to thank my wife Sue for her continuing support and patience during my research work.

Chapter 1

Background

1.1 Introduction

There is nothing like a latent
variable to stimulate the imagination.[1]

Arthur S. Goldberger

Latent or unobserved variables have the role of reducing dimensionality in multivariate analysis or representing quantities of substantive interest that are themselves not directly measurable.[2] Our focus here is on structural models with latent variables where term "structural" implies specific, theoretically implied relationships among observable (manifest) and latent variables. A similar definition of structural equation models, due to A.S. Goldberger, refers to "stochastic models in which each equation represents a causal link, rather then mere empirical association" (Goldberger 1972a, p. 979).

Structural equation latent variable models (SEM) emerged from an increasingly popular but never entirely completed merger of econometric and psychometric methods, namely structural or simultaneous equation models and factor analysis. The specifics and common aspects of these two traditions are reviewed in historical context by Goldberger (1972a). Psychometrics contributed factor-analytic measurement models hence enabling empirical measurement of latent variables, while econometrics developed structural equation models that incorporate causal, possibly simultaneous relationships among the modelled variables. Combined, these two approaches yield methods for modelling causal relationships among latent variables.

Joint estimation of a classical simultaneous equations system where all modelled variables are latent (unobserved), but measured by factor-analytic measure-

[1] Cited from Chamberlain (1990), p. 126.

[2] Bartholomew and Knott (1999) clarify this by making a distinction between a ' realist' and an ' instrumentalist' view of latent variables depending on whether latent variables are regarded as existing in the real (but possibly unmeasurable) world or as merely means of reducing complex multivariate data to smaller number of more easily interpretable dimensions.

ment models, was proposed by Keesling (1972), Wiley (1973), and Jöreskog (1973). Jöreskog (1973) furthermore gave a full theoretical analysis of this (general) model and pioneered its computer implementation in the still leading structural equation programme "LISREL".

The "Jöreskog-Keesling-Wiley model" (commonly known as *SEM* or *LISREL*) can be feasibly estimated in the full-information Gaussian maximum likelihood framework by minimising the distance between the model-implied (theoretical) and the empirical covariance matrices. Jöreskog has shown that with independent, identically distributed (i.i.d.) Gaussian data the covariance matrix has Wishart distribution with the known theoretical covariance structure. Hence, given the parameters of interests are identified, their maximum likelihood estimates can be obtained using iterative optimisation techniques such as quasi Newton or scoring algorithms. This does not necessarily holds for dynamic models and time series or panel data, which is likely the main reason why SEM models found considerably more applications in the psychometric and social science literature then in econometrics where dynamic models and time series data are standard.

Latent variables in econometrics first appeared in the "errors-in-variable" models and were broadly divided into models with truly latent variables and those with observable variables that are measured with error (Ainger et al. 1984, Wansbeek and Meijer 2000).

The gap between econometrics and psychometrics was partially bridged by the development of the general structural equation model with latent variables due to Jöreskog (1973). However, dynamic structural equation models with latent variables are rarely used in the empirical literature, in contrast to the static models. This is largely due to estimation problems and lack of appropriate statistical software.

The existing methods for estimation of dynamic models in econometrics mainly focus on errors-in-variable models (Ghosh 1989, Tercerio Lomba 1990). These models are characterized by the assumption of unobservability due to errors in measurement, hence the observable variables are considered as proxies containing measurement error. This is a different and less general assumption from the one in the factor analytic tradition where the observable variables are assumed to be generated by the latent variables and thus multiple observable indicators are considered.

There is a large literature on static SEM models with multiple indicators for independent (cross-section) data (Bartholomew and Knott 1999, Skrondal and Rabe-Hesketh 2004). The extension of these methods to longitudinal data (i.e. repeated measurement on the same cross-section) was suggested by Jöreskog and Sörbom (1977). However, longitudinal models were not used as widely in the empirical literature and the majority of the exiting applications are limited on static models using data with very small time dimension (see e.g. Jansen and Oud (1995).

More complex dynamic SEM models (DSEM) using time series and panel data[3] were not extensively researched or applied in the literature despite their considerable applicability.

In the most general framework, DSEM model encompass most dynamic linear models including dynamic simultaneous equations models, where all variables might be unobservable but measured by multiple observable indicators. Such general setting turns out to be tedious from both theoretical and empirical side. Consequently, the literature on dynamic models with latent variables focuses on various special cases of the most general model. However, numerous substantive applications call for more general framework. The most common case that motivates a general DSEM model is when the relationship among the modelled variables is dynamic and simultaneous while at the same time the error of measurement (or unobservability) is present in all variables. If multiple observable indicators are available for each unobservable variable, this situation naturally leads to a dynamic SEM specification we consider here.

DSEM modelling can be used to address the problems caused by simultaneity and measurement errors in multivariate models by making use of the information contained in the observable indicators of latent, or erroneously measured variables.

In particular, we are concerned with time series and panel models that are characterised by the following three points.

(i) All variables in the structural (simultaneous) model may be unobservable (latent);

(ii) Each latent variable is measured by one or more observable indicators;

(ii) Structural relationship can be simultaneous and dynamic, thus lags of both endogenous and exogenous latent variables are possible.

In the next section (§1.2) we review the contemporary literature on structural and dynamic latent variable models and suggest that general DSEM models satisfying the criteria (i)–(iii) above have not been fully treated.

[3]We distinguish "panel" from "longitudinal" data insofar the former comprises multiple observations on a time series process of length T that is observed N times, while the later is made of T repeated observations on the cross-section of size N. In practice, typical econometric panel data might have larger T in comparison to N then typical longitudinal data sets used in social sciences, but the true distinction is in the stochastic properties of the process that generated the data, which tends to have pronounced dynamic properties in econometric panels.

1.2 Literature review

1.2.1 Structural equation model (SEM)

The general structural equation model with latent variables (SEM) has its roots in the regression models with unobservable independent variables considered by Zellner (1970) and models discussed in Goldberger (1972b). Pagan (1973) proposed a estimation procedure for the models with composite disturbance terms while the general structural equation model with latent variables, though still without multiple indicators of the latent variables was introduced by Jöreskog (1973). Jöreskog and Goldberger (1975) analysed a special case of the SEM model known as MIMIC (multiple-indicators-multiple-causes) which is a latent variable model with perfectly observed exogenous variables ("causes"). Wansbeek and Meijer (2000) currently gives the most comprehensive review of the static structural equation models with the focus on the models for independent data.

The SEM model was introduced in the literature by Keesling (1972), Wiley (1973), and Jöreskog (1973) and is thus also known as the *Jöreskog-Keesling-Wiley model.* The first computer implementation is due to Jöreskog and Sörbom (1996b) who developed the LISREL[4] computer programme (see Cziráky (2004) for a review). The basic SEM model is specified by three matrix equations as

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1.1}$$

$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{1.2}$$

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \tag{1.3}$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_m)$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_g)$ are vectors of latent variables, $\mathbf{y} = (y_1, y_2, \ldots, y_l)$ and $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ are vectors of observable variables, and $\mathbf{B}$ $(m \times m)$, $\boldsymbol{\Gamma}$ $(m \times g)$, $\boldsymbol{\Lambda}_x$ $(k \times g)$, and $\boldsymbol{\Lambda}_y$ $(l \times m)$ are coefficient matrices. The vectors of errors in measurement in $\mathbf{y}$ and $\mathbf{x}$ are denoted by $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ and assumed to be uncorrelated with $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, and $\boldsymbol{\zeta}$.

Without loss of generality we assume that variables are measured in deviation from their mean. Note that (1.1) is a structural equation with latent variables and (1.2) and (1.3) are measurement models for the endogenous ($\boldsymbol{\eta}$) and exogenous latent variables ($\boldsymbol{\xi}$), respectively.

The specification (1.1)–(1.3), however, is not the only one and several alternatives have been suggested in the literature. The "Bentler-Weeks" and "reticular action model" specifications are the best known equivalent alternatives and both have covariance structure identical to that of the specification (1.1)–(1.3). Wansbeek and

[4]The abbreviation stands for LInear STructural RELations.

Meijer (2000) give a more detail discussion on these alternative specifications from a comparative angle. We will use the specification (1.1)–(1.3).

Historically, the SEM model (1.1)–(1.3) emerged partly from the econometrics tradition and it is easy to see the resemblance between (1.1) and the classical econometric simultaneous equation models. The other part came from psychometrics (factor analysis) tradition thus the measurement models (1.2) and (1.3) have a classical factor analytic form.

The key characteristic of the SEM model is the joint estimation of both the structural and the measurement models, which is most commonly done in the covariance structure analysis (CSA) framework. There is some discussion in the literature of whether CSA should be considered as the umbrella family of methods that include SEM as a special case or vice versa. Originally, the method for the analysis of covariance structures based on maximum likelihood was suggested as a fairly general approach encompassing factor analytic and related models by Jöreskog (1970).

The CSA approach is based on fitting a discrepancy function that minimises the difference between the model-implied (theoretical) and data-implied (empirical) covariance matrices. The SEM model (1.1)–(1.3) has a theoretical covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, expressed in terms of the model parameters, of the form

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Lambda}_y \boldsymbol{\Pi}\left(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi}\right)\boldsymbol{\Pi}'\boldsymbol{\Lambda}'_y + \boldsymbol{\Theta}_\varepsilon & \boldsymbol{\Lambda}_y \boldsymbol{\Pi}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}'_x + \boldsymbol{\Theta}_{\varepsilon\delta} \\ \boldsymbol{\Lambda}_x \boldsymbol{\Phi}\boldsymbol{\Gamma}'\boldsymbol{\Pi}'\boldsymbol{\Lambda}'_y + \boldsymbol{\Theta}_{\varepsilon\delta} & \boldsymbol{\Lambda}_x \boldsymbol{\Phi}\boldsymbol{\Lambda}'_x + \boldsymbol{\Theta}_\delta \end{pmatrix}, \tag{1.4}$$

where $\boldsymbol{\Pi} \equiv (\mathbf{I} - \mathbf{B})^{-1}$, $E[\boldsymbol{\xi}\boldsymbol{\xi}'] \equiv \boldsymbol{\Phi}$, $E[\boldsymbol{\zeta}\boldsymbol{\zeta}'] \equiv \boldsymbol{\Psi}$, $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] \equiv \boldsymbol{\Theta}_\varepsilon$, $E[\boldsymbol{\delta}\boldsymbol{\delta}'] \equiv \boldsymbol{\Theta}_\delta$, and $E[\boldsymbol{\varepsilon}\boldsymbol{\delta}'] \equiv \boldsymbol{\Theta}_{\varepsilon\delta}$. Letting $\mathbf{z} \equiv (\mathbf{y}' : \mathbf{x}')'$ where $\mathbf{Z} \equiv (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$ is a sample with N observations on $\mathbf{z}$, and $\bar{\mathbf{z}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i$ is the sample mean vector, we can further define the sample covariance matrix as

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})' \tag{1.5}$$

Let $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathrm{S})$ be a fitting function. While the ACS method is generally applicable to any form of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the SEM covariance structure (1.4) encompasses many common static linear models as special cases.

Most commonly used CSA fitting function used to obtain coefficient estimates that minimise the discrepancy between (1.4) and (1.5) is the Wishart maximum likelihood function Jöreskog (1981). It is given by

$$F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathrm{S})_{ML} = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}\ \mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \ln|\mathrm{S}| - l - k. \tag{1.6}$$

In addition to the maximum likelihood discrepancy function, there are several nonparametric alternatives. The most frequently used are the generalised least squares

(GLS) and unweighted least squares (ULS) criterion (Jöreskog and Goldberger 1972, Anderson 1973). The GLS discrepancy function is given by

$$F(\boldsymbol{\Sigma}(\boldsymbol{\theta}))_{GLS} = (\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))'\mathbf{W}^{-1}(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})), \tag{1.7}$$

where $\mathbf{W}^{-1}$ is a general matrix of weights. It is commonly taken that $\mathbf{W}^{-1} = \mathbf{S}^{-1}$ in which case (1.7) simplifies to

$$F(\boldsymbol{\Sigma}(\boldsymbol{\theta}))_{GLS} = \frac{1}{2}\mathrm{tr}\,(\mathbf{I} - \mathbf{S}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\theta}))^2,$$

or if no weights are used, i.e., $\mathbf{W}^{-1} = \mathbf{I}$, we obtain unweighted least squares (ULS) criterion

$$F(\boldsymbol{\Sigma}(\boldsymbol{\theta}))_{ULS} = \frac{1}{2}\mathrm{tr}\,(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))^2. \tag{1.8}$$

However, unlike ML and GLS the ULS criterion does not posses scale-invariance property.[5] On the other hand while ML and GLS methods require positive definite covariance matrix the ULS has no such requirement (Jöreskog 1981). Note that under some simple assumptions when $\mathrm{plim}\,\mathbf{S} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ then $\mathrm{plim}\,\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ where $\hat{\boldsymbol{\theta}} = \arg\min F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S})$ (Anderson 1989). The same results holds for $\hat{\boldsymbol{\theta}}$ that minimises $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}))_{GLS}$ or $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}))_{ULS}$. In general, $\hat{\boldsymbol{\theta}} = \arg\min F(\boldsymbol{\Sigma}(\boldsymbol{\theta}))$ will be a consistent estimator of $\boldsymbol{\theta}_0$ if $F(\boldsymbol{\Sigma}, \mathbf{S}) \to 0 \Rightarrow \boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1} \to \mathbf{I}$ (Shapiro 1983, Shapiro 1984, Kano 1986, Anderson 1989).

1.2.2 Dynamic latent variable models

Latent variable models specifically designed for dependent data (i.e. time series) were introduced in the econometric literature in the eighties and could be divided into various extensions of the classical factor analysis model and dynamic extensions of certain special cases of the static SEM model. Unlike in psychometrics, where the development of the SEM model followed an elegant path of merging the already existing simultaneous equation models with the factor analysis, the development of dynamic latent variable models did not follow such path. The SEM model has not been directly generalised to dynamic cases in its full generality and the factor analytic model was reoccurring in the literature often in less general settings then previously considered. In this section we give a brief overview of the main developments in the literature on dynamic latent variable models and show that most of these approaches stream from different traditions and thus fail to provide a unified treatment of the topic.

[5]Scale invariance implies that the value of the fit function remains unchanged regardless of the changes in the measurement scale.

Dynamic factor analysis models

Geweke (1977), Geweke and Singleton (1981) and Singleton (1980) proposed frequency-domain methods for estimation of a dynamic confirmatory factor model (DCFM). The DCFM model relates an observed vector $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tn})'$ to a linear distributed lag function of a latent vector of common factors $\boldsymbol{\xi}_t = (\xi_{t1}, \xi_{t2}, \ldots, \xi_{tk})'$, and a vector of specific factors $\boldsymbol{\delta}_t = (\delta_{t1}, \delta_{t2}, \ldots, \delta_{tn})$, where $n > k$. The model is specified as

$$\mathbf{x}_t = \sum_{k=-\infty}^{\infty} \boldsymbol{\Lambda}_k \boldsymbol{\xi}_{t-k} + \boldsymbol{\delta}_t, \tag{1.9}$$

and it is assumed that $E[\mathbf{x}] = \mathbf{0}$, $E[\boldsymbol{\xi}] = \mathbf{0}$, $\mathrm{Cov}\,(\delta_{it}, \delta_{jt}) = 0$ for $i \neq j$, $\mathrm{Cov}\,(\boldsymbol{\delta}_t, \boldsymbol{\xi}_{t-j}) = 0\ \forall j$, $\mathrm{Cov}\,(\boldsymbol{\delta}_{t-i}, \boldsymbol{\delta}_{t-j})$ is diagonal for $i = j$, and $\mathrm{Cov}\,(\xi_{it}, \xi_{jt}) = 0\ \forall i \neq j$. Both $\boldsymbol{\xi}_t$ and $\boldsymbol{\delta}_t$ are allowed to be serially correlated, and in addition $\boldsymbol{\xi}_t$ can be mutually correlated.

Denoting the autocovariance function of $\mathbf{x}_t$ by $\mathbf{R}_x(r) = E(\mathbf{x}_t \mathbf{x}'_{t+r})$, for $t = \ldots, -1, 0, 1, \ldots$, the autocovariance function of the DCFM model (1.9) is given by

$$\mathbf{R}_{\mathbf{x}}(r) = \sum_{k=-\infty}^{\infty} \boldsymbol{\Lambda}_k \sum_{l=-\infty}^{\infty} \mathbf{R}_\xi(r + k - l)\boldsymbol{\Lambda}'_l + \mathbf{R}_\varepsilon(r),$$

and thus the spectral density function $\mathbf{S}_x(\omega)$ can be obtained by taking the Fourier transform of $\mathbf{R}_x(r)$, i.e., $\mathbf{S}_x(\omega) = \sum_{r=-\infty}^{\infty} \mathbf{R}_x(r) e^{-i\omega r}$, for $|\omega| \leq \pi$. This gives

$$\begin{aligned}
\mathbf{S}_x(\omega) &= \sum_{r=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \boldsymbol{\Lambda}_k \sum_{l=-\infty}^{\infty} \mathbf{R}_\xi(r + k - l)\boldsymbol{\Lambda}'_l e^{-i\omega r} + \sum_{r=-\infty}^{\infty} \mathbf{R}_\varepsilon(r) e^{-i\omega r} \\
&= \sum_{k=-\infty}^{\infty} \boldsymbol{\Lambda}_k e^{-i\omega k} \sum_{u=-\infty}^{\infty} \mathbf{R}_\varepsilon(u) e^{-i\omega u} \sum_{z=-\infty}^{\infty} \boldsymbol{\Lambda}'_z e^{-i\omega z} + \mathbf{S}_\varepsilon(\omega)
\end{aligned}$$

The DCFM model can be estimated in frequency domain by firstly computing the finite Fourier transform of the vector $\mathbf{x}$

$$\tilde{\mathbf{x}}(\omega_j(T)) = (2\pi T)^{-1/2} \sum_{t=1}^{T} \mathbf{x}_t e^{it\omega_j(T)},$$

for $j = 1, 2, \ldots, T$ and $\omega_j(T) \equiv 2\pi j/T$. The likelihood is given by

$$L\left(\tilde{\mathbf{x}}(\omega_q^h), \mathbf{S}_x(\omega^q)\right) = (2\pi)^{-nl_q} |\mathbf{S}_x(\omega^q)|^{-l_q} \exp\left(-\sum_{h=1}^{l_q} \tilde{\mathbf{x}}(\omega_q^h)' \mathbf{S}_x(\omega^q)^{-1} \tilde{\mathbf{x}}(\omega_q^h)\right),$$

where ω_h^q, $h = 1, \ldots, l_q$ is the q-th band with l_q adjacent harmonic frequencies that splits the interval $[0, \pi]$ into Q disjoint intervals. An unconstrained quasi-maximum likelihood estimator of the spectral density matrix is given by

$$\hat{\mathbf{S}}_x(\omega^q) = l_q^{-1} \sum_{h=1}^{l_q} \tilde{\mathbf{x}}(\omega_q^h) \tilde{\mathbf{x}}(\omega_q^h)'$$

Geweke and Singleton (1981) suggest a goodness-of-fit statistic based on the likelihood ratio principle as

$$\lambda = \frac{\left|\hat{\mathbf{S}}_x(\omega)_A\right|^{-1} e^{-\sum_{i=1}^m \tilde{\mathbf{x}}_i' \hat{\mathbf{C}}_A^{-1} \tilde{\mathbf{x}}_i}}{\left|\hat{\mathbf{S}}_x(\omega)\right| e^{-\sum_{i=1}^m \tilde{\mathbf{x}}_i' \hat{\mathbf{C}}^{-1} \tilde{\mathbf{x}}_i}}, \tag{1.10}$$

where $\hat{\mathbf{S}}_x(\omega)_A$ and $\hat{\mathbf{S}}_x(\omega)$ are unconstrained and constrained ML estimators of $\mathbf{S}_x(\omega)$, respectively. Asymptotically, $2 \ln \lambda \sim \chi_d^2$ for d the number of distinct elements in $\mathbf{S}_x(\omega)$ less the number of free parameters.

Note that Geweke and Singleton (1981) methods for estimating DCFM models are in fact based on classical Wishart-likelihood CSA approach where time series data is initially spectrally decomposed and "prewhitened" to eliminate seasonality and serial autocorrelation. To see the similarity with the CSA approach described in section §1.2, note that for a finite Fourier transform of $\mathbf{x}$ at the m harmonic frequencies $(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots \tilde{\mathbf{x}}_m)$ the (complex) likelihood is of the form

$$L\left(\mathbf{S}_x(\omega), \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m\right) = (2\pi)^{-nm} |\mathbf{S}_x(\omega)|^{-m} \exp\left(-\sum_{i=1}^m \tilde{\mathbf{x}}_i' \mathbf{S}_x(\omega)^{-1} \tilde{\mathbf{x}}_i\right),$$

thus the log of the likelihood is

$$\ln L_1\left(\mathbf{S}_x(\omega), \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m\right) = -nm \ln(2\pi) - m \ln |\mathbf{S}_x(\omega)| - \operatorname{tr} \mathbf{C}\mathbf{S}_x(\omega)^{-1}, \tag{1.11}$$

where $\mathbf{C} \equiv \frac{1}{m} \sum_{i=1}^m \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i'$. Geweke and Singleton (1981) re-scale (1.11) by multiplying it by $-m^{-1}$ and add terms that do not include any unknown parameters to obtain

$$\ln L_2\left(\mathbf{S}_x(\omega), \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m\right) = \ln |\mathbf{S}_x(\omega)| + \operatorname{tr} \mathbf{C}\mathbf{S}_x(\omega)^{-1} - \ln |\mathbf{C}| - n. \tag{1.12}$$

Hence (1.12) is of the same form as the Wishart maximum likelihood fitting function (1.6). The main difference between the procedure for estimation of DCFM models (Geweke 1977, Singleton 1980, Geweke and Singleton 1981) and the Wishart-likelihood approach (Jöreskog 1970, Jöreskog 1981, Anderson 1989) is in initial transformation of the data with spectral methods that aims at rendering data serially uncorrelated and seasonality-free, thus satisfying the i.i.d. assumptions required for the CSA method. Similarly, the likelihood-ratio goodness-of-fit statistic (1.10) resembles the usual χ^2 test used in the CSA framework with a transformed data matrix.

An additional complication due to the use of spectral methods is that the Fourier transform $\tilde{\mathbf{x}}_i$ generally includes complex values, hence Geweke and Singleton (1981) additionally transform the data vector to make it real, and similarly define real-transform of the parameter vector.

Dynamic multiple indicator multiple causes model

Engle and Watson (1981) proposed a dynamic version of the MIMIC model (Zellner 1970, Goldberger 1972a, Goldberger 1972b, Jöreskog and Goldberger 1975) and suggested a maximum-likelihood procedure for estimation of a dynamic MIMIC model (DYMIMIC).[6] The model they consider is written in the state-space form to facilitate application of the Kalman filer algorithm and is specified by a state and a measurement equation, respectively as

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\phi}\mathbf{x}_{t-1} + \boldsymbol{\gamma}\mathbf{z}_t + \mathbf{v}_t \\ \mathbf{y}_t &= \boldsymbol{\alpha}\mathbf{x}_t + \boldsymbol{\beta}\mathbf{z}_t + \mathbf{e}_t \end{aligned} \tag{1.13}$$

where $\mathbf{x}_t$ $(J \times 1)$ is unobservable and $\mathbf{y}_t$ $(P \times 1)$, and $\mathbf{z}_t$ $(K \times 1)$ are observable vectors. The error vectors $\mathbf{v}_t$ and $\mathbf{e}_t$ are assumed to be normally distributed and mutually independent, i.e.,

$$\begin{pmatrix} \mathbf{v}_t \\ \mathbf{e}_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{Q}_t & 0 \\ 0 & \mathbf{R}_t \end{pmatrix}\right] \tag{1.14}$$

The DYMIMIC model (1.13) is more general then the Geweke and Singleton (1981)'s dynamic factor analysis model insofar it allows for the effects of the exogenous variables measured without error ($\mathbf{z}_t$). However, (1.13) does not allow for simultaneous relationships among latent variables ($\mathbf{x}_t$) and also assumes that exogenous variables ($\mathbf{z}_t$) are perfectly observed. Nevertheless, the DYMIMIC model includes as special cases several important time series models and it can be easily shown that models such as ARIMA, time-varying regression models, multivariate ARIMA, and dynamic factor analysis are all special cases of DYMIMIC model (Watson and Engle 1983).

Engle and Watson (1981) and Watson and Engle (1983) propose an estimation approach based on the scoring algorithm and the Kalman filter. The key statisti-

[6]See also Watson and Kraft (1984) and Engle et al. (1985).

cal assumption required is multivariate normality and mutual independence of the measurement and state error vectors (1.14). Note that (1.13) can be re-written as

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\alpha}\left(\boldsymbol{\phi}\mathbf{x}_{t-1}+\boldsymbol{\gamma}\mathbf{z}_t+\mathbf{v}_t\right)+\boldsymbol{\beta}\mathbf{z}_t+\mathbf{e}_t \\ &= \boldsymbol{\alpha}\boldsymbol{\phi}\mathbf{x}_{t-1}+\left(\boldsymbol{\alpha}\boldsymbol{\gamma}+\boldsymbol{\beta}\right)\mathbf{z}_t+\left(\boldsymbol{\alpha}\mathbf{v}_t+\mathbf{e}_t\right),\end{aligned}$$

which now has composite error structure. If we furthermore let $\boldsymbol{\alpha}\mathbf{v}_t+\mathbf{e}_t\equiv\mathbf{u}_t$ the model can be written in terms of innovations as

$$\mathbf{u}_t=-\boldsymbol{\alpha}\boldsymbol{\phi}\mathbf{x}_{t-1}-\left(\boldsymbol{\alpha}\boldsymbol{\gamma}+\boldsymbol{\beta}\right)\mathbf{z}_t \tag{1.15}$$

Denoting the contemporaneous covariance matrix of the innovations $E[\mathbf{u}_t\mathbf{u}'_t]\equiv\mathbf{H}_t$[7], it follows that the log-likelihood function of the DYMIMIC model is of the form

$$\ln L_t\left(\boldsymbol{\theta}\right)=-\frac{T}{2}\ln\left|\mathbf{H}_t\right|-\frac{1}{2}\sum_{t=1}^{T}\mathbf{u}'_t\mathbf{H}_t^{-1}\mathbf{u}_t. \tag{1.16}$$

The model parameters ($\boldsymbol{\theta}$) can be estimated recursively, using the Kalman filter, where the recursion is given by

$$\boldsymbol{\theta}^{k+1}=\boldsymbol{\theta}^k+\lambda_k\Im_k^{-1}\frac{\partial\ln L}{\partial\boldsymbol{\theta}}\bigg|\,\boldsymbol{\theta}^k. \tag{1.17}$$

Engle and Watson (1981) show that Kalman filer algorithm can be applied to the DYMIMIC model when the initial state $\boldsymbol{x}_0$ is treated as either an unknown constant (fixed) or a random variable. The former assumption allows estimation of the models containing non-stationary variables.

While relatively simple to implement, the scoring algorithm might be slow to converge thus Watson and Engle (1983) propose an additional estimation procedure based on the expectation maximisation (EM) algorithm. The EM algorithm is particularly convenient for estimation of latent variable models because the unknown values of the latent variables can be treated as missing observations. In the DYMIMIC context, Watson and Engle (1983) implement the two steps of the EM algorithm through a multivariate regression (maximisation step) and by calculation of sample moments of the smoothed values of $\mathbf{x}_t$ (estimation step). However, unlike the scoring algorithm, the EM algorithm does not produce an estimate of the information matrix. Another problem with the EM algorithm is in its insensitivity to underidentification, thus Watson and Engle (1983) suggest that EM and scoring algorithms should be combined.

[7]Hence we have $\mathbf{H}_t=E\left[\left(-\boldsymbol{\alpha}\boldsymbol{\phi}\mathbf{x}_{t-1}-\left(\boldsymbol{\alpha}\boldsymbol{\gamma}+\boldsymbol{\beta}\right)\mathbf{z}_t\right)\left(-\boldsymbol{\alpha}\boldsymbol{\phi}\mathbf{x}_{t-1}-\left(\boldsymbol{\alpha}\boldsymbol{\gamma}+\boldsymbol{\beta}\right)\mathbf{z}_t\right)'\right]$.

Dynamic shock-error model

Dynamic shock-error (DSE) model is a single equation version of an autoregressive distributed lag model with latent variables (Aigner et al. 1984, Ghosh 1989, Terceiro Lomba 1990). In the DSE model both endogenous and exogenous variables are measured with error thus a static version of the DSE model is a special case of the SEM model with only one structural equation. Furthermore, the DSE's measurement model for the latent variables allows one observable indicator per latent variable with unit loadings. The DSE model is specified with a single structural equation and two measurement equations as

$$\eta_t = \sum_{i=1}^{p} \beta_i \eta_{t-i} + \sum_{i=1}^{q} \gamma_i \xi_{t-i} + \zeta_t \tag{1.18}$$

$$y_t = \eta_t + \varepsilon_t \tag{1.19}$$

$$x_t = \xi_t + \delta_t \tag{1.20}$$

where η_t and ξ_t are scalars. Therefore, (1.18) is a single equation autoregressive distributed lag model in latent variables. The variables η_t and ξ_t are not observed, instead y_t and x_t are observed with error in the form of (1.19) and (1.20).

Ghosh (1989) proposed an estimation procedure for the DSE model (1.18) based on the state-space approach of Engle and Watson (1981) and Watson and Engle (1983). Ghosh (1989) and Terceiro Lomba (1990) suggest a maximum likelihood approach to estimation of the DSE model, which would be possible if the the model could be written in the state-space form (SSF).

Similarly to the assumptions required for estimation of the DYMIMIC models, Ghosh (1989) assumes normal and mutually independent errors in the DSE model, i.e.,

$$E \begin{bmatrix} \zeta_t^2 & \zeta_t \varepsilon_t & \zeta_t \delta_t \\ \varepsilon_t \zeta_t & \varepsilon_t^2 & \varepsilon_t \delta_t \\ \delta_t \zeta_t & \delta_t \varepsilon_t & \delta_t^2 \end{bmatrix} = \begin{pmatrix} \sigma_\zeta^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\delta^2 \end{pmatrix}$$

The DSE model thus allows for the measurement error in the exogenous variables, hence latent exogenous variables are permitted in the model unlike in the DYMIMIC model of Engle and Watson (1981). However, the Ghosh (1989) model is univariate and each latent variable is measured by a single observable indicator.

Nonparametric principal components

Bai and Ng (2002) consider estimation of the number of the unobserved factors with dependent data with a focus on applications in finance, which extends the methods introduced by Stock and Watson (1998) and Forni et al. (2000). They analyse a simple linear factor model using principal components estimator in a " panel" with $i = 1, 2, \ldots, N$ cross-section units (or variables) observed over $t = 1, 2, \ldots, T$ time periods. The model is given as

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Nt} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{Nr} \end{pmatrix} \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \\ \vdots \\ \xi_{rt} \end{pmatrix} + \begin{pmatrix} \delta_{1t} \\ \delta_{2t} \\ \vdots \\ \delta_{Nt} \end{pmatrix}, \tag{1.21}$$

where x_{it}'s are observed while λij, ξ_{it} and δ_{it} are unobserved. In full-sample notation (1.21) can be written as

$$\mathbf{X} = \Xi \mathbf{\Lambda} + \mathbf{E}, \tag{1.22}$$

where $\mathbf{X}$ and $\mathbf{E}$ are $T \times N$, Ξ is $T \times r$, and Λ is $r \times N$. This notation and terminology is somewhat unorthodox since N denotes either the number of variables or the number of cross-section units. However, the confusion between cross-sections and variables lessens in the typical finance applications where e.g. asset returns of individual firms might be considered either as observations on individuals (e.g. different firms each with specific asset return) or as variables (e.g. different asset returns coming from specific firms), and the stock market as a whole can be seen as driven by a smaller number of unobserved factors that account for much of the variability in numerous observed asset returns. This nevertheless does not cover the classical panel case with both multiple individuals and multiple variables observed over a given time period.

The methods proposed by Bai and Ng (2002) typically cover multivariate time series models where N denotes the number of variables, while both multiple individuals and multiple variables across time are not allowed, thus the model (1.21) cannot be considered a classical panel model. Note, however that multiple variables and multiple individuals can considered if the time dimension is absent in which case N would denote the number of variables (e.g. types of goods) and T would denote the number of individuals (e.g. households).

Nevertheless, an important distinction can be drawn between classical factor analysis where either N or T must be fixed and the model (1.21) which allows both N and $T \to \infty$.

The assumption in (1.21) is that the dimension r of the latent vector $\boldsymbol{\xi} = (\xi_{1t}, \xi_{2t}, \ldots, \xi_{rt})$ does not depend on N or T. There is no restriction regarding serial and cross-sectional dependence and homoscedasticity of the errors is not assumed. The degree dependence in the errors (idiosyncratic component) is however limited and in its presence the model will have an ' approximate factor structure'.

The key contribution of these methods relates to the situation when both N and T are allowed to go to infinity in which case the classical eigenvalue and maximum likelihood estimation methods tend to produce an estimate of r that increases with N, while the true r might be fixed in the population.

Bai and Ng (2002) estimate the model (1.21) using the asymptotic principal component method, which minimises the criterion function

$$V(k, \mathbf{\Lambda}, \mathbf{F}^k) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \lambda_i^{k\prime} F_i^k \right)^2, \tag{1.23}$$

subject to the constraint $1/N \mathbf{\Lambda}^{k\prime} \mathbf{\Lambda}^k = \mathbf{I_k}$ or the constraint $1/T \Xi^{k\prime} \Xi^k T = \mathbf{I_k}$, where $k < \min\{N, T\}$ is an arbitrary integer. Bai and Ng (2002) proposed several non-parametric information criteria for estimating k on the basis of the principal components solution.

Bai (2003) developed an inferential framework for the asymptotic analysis of the factor model (1.21) suitable for the cases when both N and T are large and when nether N nor T are fixed. In addition, Bai (2003) allows non-diagonal error covariance matrix and serial dependence in the latent variables, which can be treated as either fixed or random, thus extending the work of Chamberlain and Rothschild (1983), Connor and Korajzcyk (1993), and Forni et al. (2000).

The contribution of the Bai (2003) is the asymptotic distribution of both the factors and the factor loadings. In both cases it turns out that the asymptotic distribution is normal, and to obtain this result a specific linear transformation of factors and factor loadings was applied. In particular if we let $\mathbf{H} = (\mathbf{\Lambda}'\mathbf{\Lambda}/N) \left(\Xi' \tilde{\Xi} \big/ T \right) \mathbf{V}_{NT}$, where $\tilde{\Xi}$ denotes an estimate of the factor matrix given by the $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of the matrix $\mathbf{XX}'$, it follows that $\underset{T,N\to\infty}{\text{plim}} \left(\Xi' \tilde{\Xi} \big/ T \right) = \mathbf{Q}$ for an invertible matrix $\mathbf{Q}$. Subsequently, the asymptotic distribution of the linear functions $\sqrt{T} \left(\tilde{\Xi}_t - \mathbf{H}' \Xi_t \right)$ and $\sqrt{T} \left(\tilde{\mathbf{\Lambda}}_t - \mathbf{H}' \mathbf{\Lambda}_t \right)$ will be multivariate normal.

The methods suggested by Bai and Ng (2002) and Bai (2003) follow a recent tend in the financial econometrics literature on latent variable models for dependent data, however they are limited to static factor analytic models and non-parametric principal components estimation methods. Thus, the applicability of these methods to more complex dynamic models is only possibly by using the estimated factor scores. This approach can extended to models for the non-stationary data (Bai 2004), though the same limitations regarding more complex dynamic models still apply.

SEM models for time series

Structural equation models with latent variables are widely used in longitudinal analysis with repeated-measurement data and standard structural equation software for covariance structure analysis can be used to fit such models (Jöreskog and Sörbom 1977, Jansen and Oud 1995). Longitudinal studies usually do not treat repeated measurement as stochastic processes (i.e. time series) and commonly focus on static models thereby avoiding statistical complications arising from modelling dynamic structure of the data. A as an example of a typical longitudinal model consider a sample of $i = 1, \ldots, N$ individuals observed at two time points. Suppose y_{it} is brand preference and x_{it} is personal income where we wish to estimate a simple model of product-brand loyalty of the form

$$y_{i2} = \alpha + \beta y_{i1} + \gamma x_{i2}, \tag{1.24}$$

where current brand preference (y_{i2}) is affected by personal income and previous brand preference. This model does not treat brand preference as a stochastic process having specific statistical properties, rather it hypothesises that current preference toward a particular product brand might depend on the personal income but also it can be affected by person's past brand preference.

In the simplest case with no measurement error and y and x being metrical variables we would typically estimate the coefficient vector $\boldsymbol{\theta} = (\alpha, \beta, \gamma)$ by ordinary least squares as $\hat{\boldsymbol{\theta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}_2$, where $\boldsymbol{X} = (\boldsymbol{y}_1 : \boldsymbol{x}_2)$, $\boldsymbol{y}_1 = (y_{11}, \ldots, y_{N1})'$, $\boldsymbol{x}_2 = (x_{12}, \ldots, x_{N2})'$, and $\boldsymbol{Y}_2 = (y_{12}, \ldots, y_{N2})'$. When $T > 2$ repeated measurements are available on the same N individuals, the usual way to arrange the data would be into an $N \times T$ matrix $\boldsymbol{X} = (\boldsymbol{y}_1 : \cdots : \boldsymbol{x}_T)$, thus $\boldsymbol{X'X} \sim T \times T$. On the other hand, in econometric literature on panel data analysis it is common to stack all individuals into an $N \times T$ vector $\boldsymbol{Z} = (\boldsymbol{y}_1 : \cdots : \boldsymbol{x}_T)'$, which gives $\boldsymbol{Z'Z} \sim 1 \times 1$, a scalar. The usual approach in econometrics literature is to use p repeated, lagged, values of $\boldsymbol{Z}$ arranged as

$$\boldsymbol{W} = \begin{pmatrix} y_{i1} & - & - \\ y_{i2} & y_{i1} & - \\ y_{i3} & y_{i2} & y_{i1} \\ y_{i4} & y_{i3} & y_{i2} \\ y_{j1} & - & - \\ y_{j2} & y_{j1} & - \\ y_{j3} & y_{j2} & y_{j1} \\ y_{j4} & y_{j3} & y_{j2} \end{pmatrix},$$

where i and j are two different individuals (we assume $N > T$ individuals are in the sample, but show matrices for $N = 2$ to simplify the exposition). Hence, for $p = 2$ it follows that $\boldsymbol{W'W} \sim 2 \times 2$. Note that $(\boldsymbol{W'W})^{-1}\boldsymbol{W'w}$, where $\boldsymbol{w}$ is the first column of $\boldsymbol{W}$, is the vector of OLS coefficients $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ from the autoregression

$y_{it} = \beta_1 y_{it} + \beta_1 y_{it-1} + \beta_1 y_{it-2} + \varepsilon_{it}$. On the contrary, if we arranged the data (for individuals i and j) as

$$\boldsymbol{W}^* = \begin{pmatrix} y_{i1} & y_{i2} & y_{i3} \\ y_{j1} & y_{j2} & y_{j3} \end{pmatrix},$$

then computing $(\boldsymbol{W}\ '\boldsymbol{W}\ )^{-1}\boldsymbol{W}\ '\boldsymbol{w}^*$ will not produce OLS estimates of the autoregressive coefficients computed above by $(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{w}$. The data structure of the $\boldsymbol{W}^*$ matrix can be termed as "un-stacked" or "wide format", hence $\boldsymbol{W}$ would be the matrix with "stacked" or "long format" data.

The "wide format" is a natural way to arrange independent data where each column corresponds to different variable, as it would be the case with cross-section data. Once the time dimension is introduced, the "wide format" is still a natural if temporal dynamics are ignored and hence if the observations taken on the same variable in different points in time are treated as different (independent) variables. Structural equation models such as those considered by Jöreskog and Sörbom (1977) require an empirical estimate of the covariance matrix such as $(N-1)^{-1}\boldsymbol{W}\ '\boldsymbol{W}\ $, when $N > T$, though the empirical literature is inconclusive regarding SEM estimation when $N = 1$. In such case we can still compute $(T-1)^{-1}\boldsymbol{W}'\boldsymbol{W}$, but as remarked above, generally this will not lead to identical estimates.

Nevertheless, a number of empirical papers attempted to use the covariance structure analysis as implemented in standard SEM software packages such as LISREL to model pure time series data ($N = 1$) using $(T-1)^{-1}\boldsymbol{W}'\boldsymbol{W}$ in place of the empirical covariance matrix and using a fitting function such as Wishart likelihood. MacCallum and Ashby (1986) suggested using the SEM approach to fit time series models to cross-lagged (quasi) covariance matrix with data matrix arranged as

$$\begin{pmatrix} y_1 & - & - \\ y_2 & y_1 & - \\ y_3 & y_2 & y_1 \\ \vdots & \vdots & \vdots \\ y_T & y_{T-1} & y_{T-2} \end{pmatrix},$$

which after deleting rows with missing values becomes

$$\boldsymbol{Y} = \begin{pmatrix} y_3 & y_2 & y_1 \\ y_4 & y_3 & y_2 \\ y_5 & y_4 & y_3 \\ \vdots & \vdots & \vdots \\ y_T & y_{T-1} & y_{T-2} \end{pmatrix}. \tag{1.25}$$

Multiple time series and lags > 2 can be arranged as a straightforward extension of (1.25). Using (1.25) to compute an empirical covariance matrix gives

$$\frac{1}{T-1}\boldsymbol{Y}'\boldsymbol{Y} = \frac{1}{T-1}\begin{pmatrix} \sum_{t=3}^{T} y_t^2 & \sum_{t=3}^{T} y_t y_{t-1} & \sum_{t=3}^{T} y_t y_{t-2} \\ \sum_{t=3}^{T} y_t y_{t-1} & \sum_{t=2}^{T-1} y_t^2 & \sum_{t=2}^{T-1} y_t y_{t-1} \\ \sum_{t=3}^{T} y_t y_{t-2} & \sum_{t=2}^{T-1} y_t y_{t-1} & \sum_{t=1}^{T-2} y_t^2 \end{pmatrix},$$

which will converge to a Toeplitz matrix as $T \to \infty$ for stationary time series y_t, $t = 1, \ldots, T$.

Molenaar (1985) and Molenaar et al. (1992) considered estimation of dynamic factor models using empirical matrices such as $\frac{1}{T-1}\boldsymbol{Y}'\boldsymbol{Y}$. Hamaker et al. (2002) and Hamaker et al. (2003) investigated this approach to fitting univariate ARMA models and reported simulation results which implies SEM estimates differs from maximum likelihood estimates for ARMA(p, q) models with $q > 0$. These approaches are similar to the frequency-domain methods of Geweke (1977), Geweke and Singleton (1981), and Singleton (1980) and differ in terms of whether the data is pre-whitened using Fourier transform or not before the covariance matrix is computed. Further review of these and similar approaches is given in Oud (2001) and Oud (2004).

1.3 Conclusion and aims for further research

The literature on dynamic latent variable models so far considered several special cases of what could be seen as a dynamic generalisation of the static SEM model. Simple static factor analysis models can be estimated under certain restrictive assumptions about the errors using static SEM methods (Amemiya and Anderson 1990, Anderson and Amemiya 1988, Browne 1984, Shapiro 1983, Shapiro 1984, Shapiro and Brown 1987).

Dynamic models in the empirical literature are primarily limited to dynamic factor analysis models (Chamberlain and Rothschild 1983, Connor and Korajzcyk 1993, Dhrymes et al. 1984, Donald 1997, Forni et al. 2000, Forni and Reichlin 1998, Geweke 1977, Geweke and Singleton 1981, Singleton 1980), or simple static factor analysis models estimated by principal components methods (Bai and Ng 2002, Bai 2003, Bai 2004).

Generalisations of the SEM models to dynamic structural models are generally limited to dynamic version of the Jöreskog and Goldberger (1975) MIMIC model. The dynamic MIMIC model (DYMIMIC) is a structural equation model with exogenous variables, however the dynamics and measurement errors are limited to the endogenous variables (Engle et al. 1985, Engle and Watson 1981, Stock and Watson 1989, Stock and Watson 1998, Watson and Engle 1983, Watson and Kraft 1984).

Longitudinal models or models for repeated measurement were considered in the dynamic SEM context by McArdle (1988) and McArdle (2001). A similar model was proposed by Dunson (2003) for categorical variables. These models are, however, quasi-dynamic since they treat repeated measures on the same variable as distinct variables and thus formulate standard SEM models with repeatedly measured variables.

The extensions that consider lagged exogenous latent variables or exogenous variable measured with errors along with the lagged endogenous latent variables were focused on "dynamic shock-error" models and dynamic errors-in-variables models, which are essentially single equation (univariate) models with univariate measurement models, i.e. the unobserved variables are proxied by a single indicator variable (Bloch 1989, Deistler and Anderson 1989, Ghosh 1989, Terceiro Lomba 1990). The approach taken by Ghosh (1989) and Terceiro Lomba (1990) to estimation of the dynamic shock-error models is based on their re-writing in the state-space form. However, even with the simplest univariate models the state-space form is difficult to obtain. Ghosh (1989) solves this problem by introducing an additional autoregressive equation for the exogenous latent variable. On the other hand, Terceiro Lomba (1990) considers models with contemporaneous exogenous latent variables hence avoiding the problem with lagged exogenous variables which cannot be easily written in the state-space form.

The existing literature is scarce in respect to dynamic structural equation model with latent endogenous and latent exogenous variables, multiple simultaneous equations, and measurement models for the latent variables with multiple indicators. Such models would present a dynamic generalisation of the static multi-indicator SEM model, hence, theoretical and practical consideration of dynamic SEM models would be an important extension of the literature.

In summary, we can identify three main problem areas where further research is needed.

Unifying theoretical framework Different traditions in the literature deal with various special cases of dynamic structural equation models, such as errors-in-variables, state-space, and latent variable models. There is a notable divergence in the literature and lack of cross-referencing. Consequently, methods that focus on errors-in-variables, latent variable models, or state-space models appear to be concerned with different models rather then special cases of dynamic structural equation models.

Feasible theoretical analysis Standard theoretical and asymptotic analysis is virtually intractable using classical approaches and methods; Consequently, basic results such as the analytic derivatives, the Hessian matrix, or the Cramer-Rao lower bound are difficult or impossible to obtain for complex multivariate models such as DSEM using standard methods.

Software implementation The currently available SEM software packages such as LISREL, AMOS, M-Plus, Mx, and EQS were designed for cross-sectional models and independent data. While specification of dynamic models in these packages is possible, it typically requires specifying the modelled relationship for each time point and subsequently imposing equality restrictions across all time points. Moreover, with the exception of one package (Mx), dynamic models can be estimated only with panel data (time series cross-section) but not with pure time series data. In addition, the Mx package also requires equality restrictions and does not make use of analytical derivatives in the estimation. These packages also require very good starting values for estimation; there are currently no specific methods for obtaining starting values in dynamic models, which is an additional obstacle for empirical implementation.

1.4 Outline of the thesis

The thesis focuses on estimation of dynamic structural (i.e. simultaneous) equation models in which some or all variables might be unobservable (latent) or measured with error. Moreover, we consider the situation where latent variables can be measured with multiple observable indicators and where lagged values of latent variables might be included in the model. This situation leads to a dynamic structural equation model (DSEM), which can be viewed as dynamic version of the structural equation model (SEM). Our focus is on obtaining coefficient estimates using both parametric and non-parametric methods. Post-estimation diagnostics and measures of overall fit are beyond the scope of the present work and are thus left for further research.

Taking the mismeasurement problem into account aims at reducing or eliminating the errors-in-variables bias and hence at minimising the chance of obtaining incorrect coefficient estimates. Furthermore, such methods can be used to improve measurement of latent variables and to obtain more accurate forecasts.

The literature on dynamic latent variable models can be divided into several different traditions emerging from fields such as econometrics, psychometrics, and engineering. Certain special cases of dynamic structural equation models, such as dynamic factor model, have been extensively analysed in the time series literature. There is a close link between these methods and the unobservable states models estimated in the state-space form. In chapter §1 we give an overview of the literature by addressing the key developments and pointing out to the areas requiring further research.

Latent variable models have been traditionally analysed as errors-in-variable models using instrumental variables methods in the mainstream econometrics literature, as covariance structure models in the psychometrics literature, and as state space models in both engineering and econometrics literature. Chapter §2 addresses the a lack of a unifying theoretical framework for dynamic models with latent variables and suggests such framework based on DSEM model, which can be shown to encompass numerous specific models considered in the literature. The approach taken here uses the idea of a Gaussian vector likelihood and specifies the theoretical covariance structure implied by the DSEM model for a multivariate time series process that started at $t = 1$ and was observed till $t = T$. It is shown that different approaches to errors-in-variables and latent variables can be viewed as different forms of the DSEM model, hence giving rise to specific multivariate likelihoods, whose parametrisations can be compared within a unifying statistical framework.

Chapter §3 considers maximum likelihood estimation of DSEM models in multivariate Gaussian context for the $N > T$ (panel) case and gives the analytical expressions for the score and the Hessian matrix along with a closed-form (theoretical) covariance matrix. The closed-form covariance matrix is obtained by making certain assumptions about the pre-sample values, which requires large-T asymptotics. Some of the existing theoretical results for the SEM model, namely the analytical first derivatives, were implemented in SEM software packages such as LISREL, which can be used to estimate certain DSEM models for panel data with relatively small T.

The analytical results obtained in chapter §3 differ from the existing results in two respects. Firstly, the model is formulated for the time series process, which eliminates the necessity to specify a separate SEM model for each time point and then impose cross-equation restrictions across all time points, as it is necessary in LISREL and similar SEM software packages. Secondly, the analytical results are obtained using modern matrix calculus methods based on zero-one matrices that enable derivation of fully vectorised expressions for the first and second derivatives. Moreover, the obtained score vectors contain derivatives for individual DSEM coefficients thus no equality or symmetry restrictions need to be imposed on the score vector. Fully vectorised expressions make standard asymptotic analysis straightforward and facilitate computer implementation in modern matrix languages such as S, R, or Ox.

Chapter §4 considers maximum likelihood estimation of DSEM models with pure time series data using a "raw data" maximum likelihood (RD-ML). In this chapter we obtain the closed-form expressions for the likelihood and analytical derivatives of the pure time series DSEM model thus providing the analytical inputs for the RD-ML estimation. Moreover, we outline some S code for estimation of such models using quasi-Newton optimisers in S-Plus and R environments.

In chapter §5 we propose non-parametric methods for estimation of DSEM models suitable for both pure time series and panel data. Generalised instrumental variables (GIVE) and full information instrumental variable (FIVE) methods are considered for the estimation of DSEM models in the "observed form", i.e., as errors-in-variable models with composite error terms.

These methods are specific in terms of model specification and choice of instruments, which are here interdependent. Namely, we specify the latent variable model as a DSEM model in which measurement errors need to satisfy certain statistical criteria. These criteria are similar to those in the classical factor analysis and are based on the validity of observable indicators as measures of the unobservable (latent) variables. Valid measurement models should have uncorrelated measurement errors, which can be generalised to the time series context by further requiring zero lagged covariances of the measurement errors. We show that basic specification of the DSEM model implies lags of the observable indicators as potentially valid instruments. Empirical validity of such instruments can be tested using standard validity of instruments tests. Instrumental variables methods have a well known advantage of not imposing any distributional assumptions on the data. They also provide non-iterative estimators that are very easy to compute using standard general purpose statistical software. An additional purpose of these methods is in obtaining good starting values for maximum likelihood estimation using standard SEM software packages such as LISREL.

In chapter §6 the above methods are applied to real-data empirical examples with two main aims. The first aim is to demonstrate how DSEM models can be estimated using standard econometric and SEM software packages when starting values are obtained using the methods suggested in chapter §5. Both fixed and random effects dynamic panel models are considered in the context of specific empirical applications: a model of financial development and economic growth and a micro-consumption model. The second aim is to investigate the limits of the existing SEM software on data size and model complexity in estimation of empirical DSEM models.

DSEM models can be easily estimated using GIVE/FIVE methods with standard econometric software packages, which holds for both pure time series and panel models and for very large data sets. Moreover, these methods provide estimates that can be used as starting values in standard SEM software packages. Using the LISREL package, we show that even for relatively simple DSEM models convergence cannot be achieved without starting values that are very close to the maximum likelihood estimates. Nevertheless, we show that the starting values obtained with GIVE/FIVE methods can be successfully used as starting values in LISREL estimation.

The ability of SEM software to handle panels with large T is, however, very limited. Along with the need to specify the model for each time point and subsequently impose equality constraints on all parameters across T time points, we also report computing difficulties associated even with relatively small T. The largest model we estimate using LISREL in combination with GIVE/FIVE starting values uses a panel data set with $N = 5152$ and $T = 13$. Using these data, we estimate a DSEM model with three structural equations including dynamics of up to five lags, with 37 coefficients, estimated as 13×37 coefficients with equality constraints across $T = 13$ time periods, which might be one of the largest models estimated with LISREL. It seems unlikely that similar models could be estimated for much larger T using standard SEM software such as LISREL. This suggests two limitations of the currently available SEM software for estimation of panel DSEM models. First is dependence on externally provided starting values. The second is the "small T problem", namely, values of T in the vicinity of 13 (our largest estimated model) are too small to satisfy the large-T asymptotics we needed to obtain the close-form theoretical covariance matrix. While it is plausible that somewhat larger panels might still be estimated using the available software, this will be unlikely with sufficiently large T needed to justify the asymptotic assumptions.

Chapter 2

Statistical framework

2.1 Introduction

The literature on dynamic latent variable models can be broadly classified into three traditions. The first tradition emerged from econometrics literature on the errors-in-variable models and regression with measurement error (Cheng and Van Ness 1999, Wansbeek and Meijer 2000). The second one is closely linked to covariance structure methods and generalised method of moments, streaming from the psychometrics and multivariate statistics (Jöreskog 1981, Bartholomew and Knott 1999, Skrondal and Rabe-Hesketh 2004). Finally, the third tradition based on estimation of the models written in "state-space form" emerged from control engineering and was adopted in econometrics owing to the suitability of the Kalman filter algorithm for estimation of various econometric models written in the " state space form" (Harvey 1989, Durbin and Koopman 2001).

This threefold and apparently diverging developments did not facilitate advance of dynamic latent variable models matching the expanding literature on static latent variable models (see e.g. Skrondal and Rabe-Hesketh (2004) for a comprehensive review). Consequently, specific empirical applications became linked with particular estimation methods and a lack of a more general framework hindered estimation of more elaborate empirical models. For example, the DYMIMIC model of Engle et al. (1985) permits dynamics in the endogenous latent variables but does not allow exogenous latent variables, which facilitated a number of empirical applications in which substantive problems had to be limited to static, perfectly observable exogenous variables.

Aside of seemingly diverging and specific directions in the development of particular estimation methods, a notable lack of cross-referencing among the three main traditions can be observed in different streams of literature. In summary, an encompassing statistical framework that unifies different traditions in development of estimation methods would facilitate both developments of estimation methods and implementation of more general empirical models.

We suggest a unifying statistical framework for dynamic latent variable models based on the general dynamic structural or simultaneous equation model (DSEM). DSEM model is general in the sense it subsumes many dynamic (and static) linear models under a common parametric form.

We develop a statistical framework by making distributional assumptions about the exogenous components and the measurement errors in the general DSEM model. We then show how the general model can be formulated following the three main traditions and compare the models resulting from such formulations by referring to their stochastic properties. In particular, we show that different approaches do not necessarily result in identical reparametrisation of the general model, rather some additional or different statistical assumptions need to be made to make different models equivalent. Finally, we suggest that some forms are suitable for particular estimation methods and briefly discuss the implications for the development of such methods.

2.2 General dynamic structural equation model (DSEM)

In this section we consider a dynamic simultaneous equation model with latent variables (DSEM). A DSEM(p, q) model at any time period t using the " t-notation" as

$$\boldsymbol{\eta}_t = \sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{t-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{\zeta}_t \tag{2.1}$$

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_y \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t \tag{2.2}$$

$$\boldsymbol{x}_t = \boldsymbol{\Lambda}_x \boldsymbol{\xi}_t + \boldsymbol{\delta}_t \tag{2.3}$$

where $\boldsymbol{\eta}_t = (\eta_t^{(1)}, \eta_t^{(2)}, \ldots, \eta_t^{(m)})'$ and $\boldsymbol{\xi}_t = (\xi_t^{(1)}, \xi_t^{(2)}, \ldots, \xi_t^{(g)})'$ are vectors of possibly unobserved (latent) variables, $\boldsymbol{y}_t = (y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(n)})'$ and $\boldsymbol{x}_t = (x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(k)})'$ are vectors of observable variables, and $\boldsymbol{B}_j$ $(m \times m)$, $\boldsymbol{\Gamma}_j$ $(m \times g)$, $\boldsymbol{\Lambda}_x$ $(k \times g)$, and $\boldsymbol{\Lambda}_y$ $(n \times m)$ are coefficient matrices. The contemporaneous and simultaneous coefficients are in $\boldsymbol{B}_0$, and $\boldsymbol{\Gamma}_0$, while $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, $\ldots$, $\boldsymbol{B}_p$, and $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, $\ldots$, $\boldsymbol{\Gamma}_q$ contain coefficients of the lagged variables.

The DSEM model (2.1)–(2.3) can be viewed either as a dynamic generalisation of the static structural equation model with latent variables (SEM) or a generalised dynamic simultaneous equation model with unobservable variables. The static SEM (LISREL) model (Jöreskog 1970, Jöreskog 1981) is thus a special case of (2.1)–(2.3) with $\boldsymbol{B}_j = \boldsymbol{\Gamma}_j = \boldsymbol{0}$, for $j > 0$. Moreover, the general DSEM encompasses virtually all static or dynamic linear models, which can be specified by imposing zero restrictions on its parameter matrices. Table 2.1 lists the most common multivariate models and shows how they can be specified as special (restricted) cases of the general DSEM model (2.1)–(2.3).

Table 2.1: Special cases of the DSEM model

Model	Restrictions
Multivariate regression[a]	$\boldsymbol{B}_j = \boldsymbol{0}\ (\forall j)$, $\boldsymbol{\Gamma}_j = \boldsymbol{0}\ (j > 0)$, $\boldsymbol{\Lambda}_y = \boldsymbol{\Lambda}_x = \boldsymbol{I}$, $\boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{0}$
VAR(p)[a]	$\boldsymbol{\Lambda}_y = \boldsymbol{I}$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Gamma}_j = \boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{0}\ (\forall j)$
VMA(q)[a]	$\boldsymbol{B}_j = \boldsymbol{0}\ \forall j$, $\boldsymbol{\Lambda}_y = \boldsymbol{I}$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Psi} = \boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{0}$
VARMA(p, q)[a]	$\boldsymbol{\Gamma}_0 = \boldsymbol{\Lambda}_y = \boldsymbol{I}$, $\boldsymbol{B}_0 = \boldsymbol{\Lambda}_x = \boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \boldsymbol{\Psi} = \boldsymbol{0}$
Factor analysis[b]	$\boldsymbol{B}_j = \boldsymbol{0}\ (\forall j)$, $\boldsymbol{\Gamma}_j = \boldsymbol{0}\ (\forall j)$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Theta}_\delta = \boldsymbol{\Psi} = \boldsymbol{0}$
Dynamic factor analysis[c]	$\boldsymbol{\Gamma}_j = \boldsymbol{0}\ (\forall j)$, $\boldsymbol{\Lambda}_x = \boldsymbol{\Theta}_\delta = \boldsymbol{0}$
SEM (LISREL)[a]	$\boldsymbol{B}_j = \boldsymbol{0}\ (j > 0)$, $\boldsymbol{\Gamma}_j = \boldsymbol{0}\ (j > 0)$
DYMIMIC[d]	$\boldsymbol{\Lambda}_x = \boldsymbol{I}$, $\boldsymbol{\Theta}_\delta = \boldsymbol{0}$
Dynamic shock-error model[e]	$\boldsymbol{B}_j = \beta_j$, $\boldsymbol{\Gamma}_j = \gamma_j$, $\boldsymbol{\Lambda}_y = 1$, $\boldsymbol{\Lambda}_x = 1$, $\boldsymbol{\Psi} = \psi$, $\boldsymbol{\Theta}_\varepsilon = \theta$, $\boldsymbol{\Theta}_\delta = \delta$

[a] Hamilton (1994), Giannini (1992).
[b] Bartholomew and Knott (1999), Skrondal and Rabe-Hesketh (2004).
[c] Geweke (1977), Geweke and Singleton (1981), Engle and Watson (1981).
[d] Engle et al. (1985), Watson and Engle (1983).
[e] Ghosh (1989), Terceiro Lomba (1990).

The idea behind the SEM model was to combine multiple-indicator factor-analytic measurement model for the latent variables with a structural equation model thus allowing for the measurement error in all variables in the structural model (Jöreskog 1970, Jöreskog 1981, Bartholomew and Knott 1999, Skrondal and Rabe-Hesketh 2004). The static SEM model can be written as a special case of (2.1)–(2.3), i.e.,

$$\boldsymbol{\eta}_t = \boldsymbol{B}_0\boldsymbol{\eta}_t + \boldsymbol{\Gamma}_0\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t \tag{2.4}$$

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_y\boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t \tag{2.5}$$

$$\boldsymbol{x}_t = \boldsymbol{\Lambda}_x\boldsymbol{\xi}_t + \boldsymbol{\delta}_t. \tag{2.6}$$

Since both $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ are unobservable some reduction or elimination of the unobservables would be necessary. An econometric interpretation would consider (2.4) a simultaneous equation model in the structural form (see e.g. Judge et al. (1988)). Here, by " structural" we refer to the model with endogenous variables on both sides of the equation as opposite to the " reduced" model, which has endogenous variables only on the left-hand side. We can easily obtain the reduced form of (2.4) as[1] $\boldsymbol{\eta}_t = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1}(\boldsymbol{\Gamma}_0\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t)$, which can be further substituted into (2.5) to obtain the " reduced " form of the model

[1]We assume that $\boldsymbol{I} - \boldsymbol{B}_0$ is of full rank, hence $(\boldsymbol{I} - \boldsymbol{B}_0)^{-1}$ exists.

$$
\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\Lambda}_y (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} (\boldsymbol{\Gamma}_0 \boldsymbol{\xi}_t + \boldsymbol{\zeta}_t) + \boldsymbol{\varepsilon}_t && (2.7) \\
\boldsymbol{x}_t &= \boldsymbol{\Lambda}_x \boldsymbol{\xi}_t + \boldsymbol{\delta}_t, && (2.8)
\end{aligned}
$$

with has only observable variables on the left-hand side. This enables derivation of the closed-form covariance matrix of $\boldsymbol{w}_i \equiv (\boldsymbol{y}_t' : \boldsymbol{x}_t')'$ in terms of the model parameters. For instance, if $\boldsymbol{w}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, it follows that $(T-1)\boldsymbol{S} \sim W(T-1, \boldsymbol{\Sigma})$, where $\boldsymbol{S} = \frac{1}{T-1}\sum_{i=1}^{T} \boldsymbol{w}_i \boldsymbol{w}_i'$ is the empirical covariance matrix, and W denotes the Wishart distribution.[2]

However, the same approach cannot be straightforwardly applied to the DSEM model (2.1)–(2.3), which contains lagged latent variables. Namely, the reduction from (2.4)–(2.6) to (2.7)–(2.8) would not eliminate the lagged values of $\boldsymbol{\eta}_t$.

The likelihood function for a sample of T observations generated by a dynamic model specified for a typical time point t (i.e. in " t-notation), such as (2.1)–(2.3), can be obtained recursively by sequential conditioning (Hamilton 1994, p. 118). In this approach we would write down the probability density function of the first sample observation ($t = 1$) conditional on the initial $r = \max(p, q)$ observations and then obtain the density for the second sample observation ($t = 2$), conditional on the the first, etc. until the last observation ($t = T$). The likelihood function would then be obtained as a product of the T sequentially derived conditional densities, assuming conditional independence of the successive observations. However, this approach is not feasible for complex multivariate dynamic models with latent variables as sequential conditioning soon becomes intractable.

An alternative approach leading to an equivalent expression for the likelihood function would be to assume that the observed sample came from a T-variate (e.g. Gaussian) distribution, having multivariate density function, from which the sample likelihood immediately follows (Hamilton 1994, p. 119). This approach might not be easily applicable to dynamic latent variable models for which we generally wish to obtain the likelihood in separated form, i.e., with all unknown parameters placed in the covariance matrix, separated from the observed data vectors. Without such

[2] The Wishart distribution has the likelihood function of the form

$$
f_W(\boldsymbol{S}) = \frac{|\boldsymbol{S}|^{\frac{1}{2}(T-1-n-k)} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\right]}{\pi^{\frac{1}{4}T(T-1)} 2^{\frac{1}{2}(T(n+k))} |\boldsymbol{\Sigma}|^{\frac{1}{2}(n+k)} \prod_{j=1}^{p} \Gamma\left(\frac{T+1-j}{2}\right)}
$$

where T is the sample size; see e.g. Anderson (1984). When a closed form of the model-implied covariance matrix $\boldsymbol{\Sigma}$ is available, assuming the model is identified or overidentified and the data is multinormal, it is straightforward to obtain the maximum likelihood estimates of the parameters by maximising the logarithm of the Wishart likelihood. In the later case, a measure of the overall fit can be obtained as −2 times the Wishart log likelihood, which is asymptotically χ^2 distributed; see e.g. Amemiya and Anderson (1990).

separation we would be left with T "missing" observations on the latent vectors $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ instead of only their unknown second moment matrices.

We can solve this problem by specifying a DSEM model (2.1)–(2.3) for the time series process that started at time $t = 1$ and was observed till time $t = T$ using a " T-notation" defined in Table 2.2. The vector $\{*\}_1^T$ can then be taken as a single realization from a T-variate distribution.

Table 2.2: T-notation

Symbol	Definition	Dimension
$\boldsymbol{H}_T$	$\text{vec}\{\boldsymbol{\eta}_t\}_1^T = (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_T')'$	$mT \times 1$
$\boldsymbol{Z}_T$	$\text{vec}\{\boldsymbol{\zeta}_t\}_1^T = (\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_T')'$	$mT \times 1$
$\boldsymbol{\Xi}_T$	$\text{vec}\{\boldsymbol{\xi}_t\}_1^T = (\boldsymbol{\xi}_{1)}', \ldots, \boldsymbol{\xi}_T')'$	$gT \times 1$
$\boldsymbol{Y}_T$	$\text{vec}\{\boldsymbol{y}_t\}_1^T = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$	$nT \times 1$
$\boldsymbol{E}_T$	$\text{vec}\{\boldsymbol{\varepsilon}_t\}_1^T = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_T')'$	$nT \times 1$
$\boldsymbol{X}_T$	$\text{vec}\{\boldsymbol{x}_t\}_1^T = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T')'$	$kT \times 1$
$\boldsymbol{\Delta}_T$	$\text{vec}\{\boldsymbol{\delta}_t\}_1^T = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_T')'$	$kT \times 1$

Working with the model in T-notation will enable us to " reduce" the model (2.1)–(2.3) and obtain a closed form covariance structure and hence a closed form likelihood of the general DSEM model.

We make the following simplifying assumption about the pre-sample (initial) observations.

Assumption 2.2.0.1 (Initial observations) *We assume that* $r = \max(p, q)$ *pre-sample observations are equal to their expectation, i.e.,* $\boldsymbol{\eta}_{i(-r)} = \boldsymbol{\eta}_{i(-r+1)} = \cdots = \boldsymbol{\eta}_{i0} = \boldsymbol{0}$ *and* $\boldsymbol{\xi}_{i(-r)} = \boldsymbol{\xi}_{i(-r+1)} = \cdots = \boldsymbol{\xi}_{i0} = \boldsymbol{0}$.

Anderson (1971) suggested that such treatment of the pre-sample (initial) values allows considerable simplification of the covariance structure and gradients of the Gaussian log-likelihood. More recently, Turkington (2002) showed that making such assumption allows more tractable mathematical treatment of complex multivariate models by using the shifting and zero-one matrices. In addition, we require covariance stationarity as follows.

Assumption 2.2.0.2 (Covariance stationarity) *The observable and latent variables are mean (or trend) stationary and covariance stationary.*

Letting $s = \ldots, -1, 0, 1, \ldots$, *we require the following*

1. $E[\boldsymbol{\eta}_t] = E[\boldsymbol{\xi}_t] = \boldsymbol{0} \Rightarrow E[\boldsymbol{y}_t] = E[\boldsymbol{x}_t] = \boldsymbol{0}$.[3]

[3]The cases with deterministic trend can be incorporated in the present framework by considering detrended variables, e.g. if $\boldsymbol{z}_t$ contains deterministic trend, we can define $\tilde{\boldsymbol{z}}_t \equiv \boldsymbol{z}_t - t$, which is trend-stationary.

2. *The structural equation (2.1) is stable, and the roots of the equations*

$$|\boldsymbol{I} - \lambda \boldsymbol{B}_1 - \lambda^2 \boldsymbol{B}_2 - \cdots - \lambda^p \boldsymbol{B}_p| = 0 \text{ and } |\boldsymbol{I} - \lambda \boldsymbol{\Gamma}_1 - \lambda^2 \boldsymbol{\Gamma}_2 - \cdots - \lambda^q \boldsymbol{\Gamma}_q| = 0$$

are greater then one in absolute value.

3. $E\left[\boldsymbol{\xi}_t \boldsymbol{\xi}'_{t-s}\right] \equiv \boldsymbol{\Phi}_s$, *so that* $\boldsymbol{\Phi}_{-s} = \boldsymbol{\Phi}'_s$.

By Assumption 2.2.0.2 it follows that the observable variables generated by the latent variables are also covariance stationary, i.e., $\forall s, k \in \mathbb{Z}$, $E\left[\boldsymbol{y}_t \boldsymbol{y}'_{t-s}\right] = E\left[\boldsymbol{y}_t \boldsymbol{y}'_{t-k}\right]$, $E\left[\boldsymbol{x}_t \boldsymbol{x}'_{t-s}\right] = E\left[\boldsymbol{x}_t \boldsymbol{x}'_{t-k}\right]$, and $E\left[\boldsymbol{y}_t \boldsymbol{x}'_{t-s}\right] = E\left[\boldsymbol{y}_t \boldsymbol{x}'_{t-k}\right]$. Next, by Assumption 2.2.0.1 the pre-sample (initial) observations are zero thus we can ignore them and write the DSEM model (2.1)–(2.3) for the time series process that started at time $t = 1$ and was observed until $t = T$ in the " T-notation" as $\{\boldsymbol{\eta}_t\}_1^T \equiv (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T)$, or

$$\{\boldsymbol{\eta}_t\}_1^T = \begin{pmatrix} \eta_1^{(1)} & \cdots & \eta_T^{(1)} \\ \vdots & \cdots & \vdots \\ \eta_1^{(m)} & \cdots & \eta_T^{(m)} \end{pmatrix}, \tag{2.9}$$

and similarly, $\{\boldsymbol{\xi}_t\}_1^T \equiv (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T)$ and $\{\boldsymbol{\zeta}_t\}_1^T \equiv (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_T)$. The structural equation (2.1) can thus be written for the time series process as

$$\{\boldsymbol{\eta}_t\}_1^T = \sum_{j=0}^{p} \boldsymbol{B}_j \{\boldsymbol{\eta}_t\}_1^T \boldsymbol{S}'^j_T + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \{\boldsymbol{\xi}_t\}_1^T \boldsymbol{S}'^j_T + \{\boldsymbol{\zeta}_t\}_1^T, \tag{2.10}$$

where we made use of a $T \times T$ shifting matrix $\boldsymbol{S}_T$ given by

$$\boldsymbol{S}_T \equiv \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}. \tag{2.11}$$

By definition, we take $\boldsymbol{S}_T^0 \equiv \boldsymbol{I}_T$. The structural equation (2.10) can be vectorised using the vec operator that stacks the $e \times f$ matrix $\boldsymbol{Q}$ into an $ef \times 1$ vector vec$\boldsymbol{Q}$, i.e., vec$\boldsymbol{Q} = (\boldsymbol{q'}_1, \ldots, \boldsymbol{q'}_f)'$ where $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_f)$. Therefore, from (2.10) we can obtain the structural equation in the reduced form as

$$\begin{aligned} \operatorname{vec}\{\boldsymbol{\eta}_t\}_1^T &= \left(\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right) \operatorname{vec}\{\boldsymbol{\eta}_t\}_1^T + \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \operatorname{vec}\{\boldsymbol{\xi}_t\}_1^T + \operatorname{vec}\{\boldsymbol{\zeta}_t\}_1^T \\ &= \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \operatorname{vec}\{\boldsymbol{\xi}_t\}_1^T + \operatorname{vec}\{\boldsymbol{\zeta}_t\}_1^T\right) \end{aligned} \tag{2.12}$$

where

$$
\begin{aligned}
\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j &= \left(\boldsymbol{S}_T^0 \otimes \boldsymbol{B}_0\right) + \left(\boldsymbol{S}_T^1 \otimes \boldsymbol{B}_1\right) + \ldots + \left(\boldsymbol{S}_T^p \otimes \boldsymbol{B}_p\right) \\
&= \begin{pmatrix}
\boldsymbol{B}_0 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{B}_1 & \boldsymbol{B}_0 & \boldsymbol{0} & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{B}_p & \cdots & \boldsymbol{B}_1 & \boldsymbol{B}_0 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{B}_p & \cdots & \boldsymbol{B}_1 & \boldsymbol{B}_0 & \ddots & \boldsymbol{0} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{B}_p & \cdots & \boldsymbol{B}_1 & \boldsymbol{B}_0
\end{pmatrix},
\end{aligned}
\tag{2.13}
$$

and hence

$$
\left(\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right) \operatorname{vec}\{\boldsymbol{\eta}_t\}_1^T = \begin{pmatrix}
\boldsymbol{B}_0 \boldsymbol{\eta}_{i1} \\
\sum_{j=0}^{1} \boldsymbol{B}_j \boldsymbol{\eta}_{(2-j)} \\
\vdots \\
\sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{(p+1-j)} \\
\sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{(p+2-j)} \\
\vdots \\
\sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{(T-j)}
\end{pmatrix}.
\tag{2.14}
$$

Similarly, note that

$$
\begin{aligned}
\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) &= \left(\boldsymbol{S}_T^0 \otimes \boldsymbol{\Gamma}_0\right) + \left(\boldsymbol{S}_T^1 \otimes \boldsymbol{\Gamma}_1\right) + \ldots + \left(\boldsymbol{S}_T^q \otimes \boldsymbol{\Gamma}_p\right) \\
&= \begin{pmatrix}
\boldsymbol{\Gamma}_0 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \boldsymbol{0} & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \ddots & \boldsymbol{0} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0
\end{pmatrix},
\end{aligned}
\tag{2.15}
$$

which implies that

$$\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \mathrm{vec}\,\{\boldsymbol{\xi}_t\}_1^T = \begin{pmatrix} \boldsymbol{\Gamma}_0 \boldsymbol{\xi}_{i1} \\ \sum_{j=0}^{1} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(2-j)} \\ \vdots \\ \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(q+1-j)} \\ \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(q+2-j)} \\ \vdots \\ \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{(T-j)} \end{pmatrix}. \tag{2.16}$$

Now let $\boldsymbol{\iota}_r$ be an $r \times 1$ vector of ones, i.e., $\boldsymbol{\iota}_r \equiv (1, 1, \ldots, 1)'$, so that we can write the $mT \times m$ block-vector of identity matrices of order m as $(\boldsymbol{I}_m, \boldsymbol{I}_m, \ldots, \boldsymbol{I}_m)' = (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)$. Note that $(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)' = \frac{1}{T}(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_m)$ and $(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m)'(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_m) = T\boldsymbol{I}_m$.

Writing the measurement equations (2.2) and (2.3) for the process vectors $\{\boldsymbol{y}_t\}_T^1$ and $\{\boldsymbol{x}_t\}_T^1$ we have the equations $\{\boldsymbol{y}_t\}_T^1 = \boldsymbol{\Lambda}_y \{\boldsymbol{\eta}_t\}_1^T + \{\boldsymbol{\varepsilon}_t\}_1^T$ and similarly $\{\boldsymbol{x}_t\}_T^1 = \boldsymbol{\Lambda}_x \{\boldsymbol{\xi}_t\}_1^T + \{\boldsymbol{\delta}_t\}_1^T$, which after applying the vec operator become

$$\mathrm{vec}\,\{\boldsymbol{y}_t\}_T^1 = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\,\mathrm{vec}\,\{\boldsymbol{\eta}_t\}_1^T + \mathrm{vec}\,\{\boldsymbol{\varepsilon}_t\}_1^T \tag{2.17}$$

$$\mathrm{vec}\,\{\boldsymbol{x}_t\}_T^1 = (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\,\mathrm{vec}\,\{\boldsymbol{\xi}_t\}_1^T + \mathrm{vec}\,\{\boldsymbol{\delta}_t\}_1^T. \tag{2.18}$$

Finally, using the notation from Table 2.2, the DSEM model DSEM model (2.1)-(2.3) can now be written as

$$\begin{aligned}
\underbrace{\boldsymbol{H}_T}_{mT\times 1} &= \underbrace{\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}}_{mT\times mT} \left[\underbrace{\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right)}_{gT\times gT} \underbrace{\boldsymbol{\Xi}_T}_{gT\times 1} + \underbrace{\boldsymbol{Z}_T}_{mT\times 1}\right] && (2.19)\\
\underbrace{\boldsymbol{Y}_T}_{nT\times 1} &= \underbrace{(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)}_{nT\times mT} \underbrace{\boldsymbol{H}_T}_{mT\times 1} + \underbrace{\boldsymbol{E}_T}_{nT\times 1} && (2.20)\\
\underbrace{\boldsymbol{X}_T}_{kT\times 1} &= \underbrace{(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)}_{kT\times gT} \underbrace{\boldsymbol{\Xi}_T}_{gT\times 1} + \underbrace{\boldsymbol{\Delta}_T}_{kT\times 1}. && (2.21)
\end{aligned}$$

It follows that (2.19) can be substituted into (2.20) to obtain a system of equations with observable variables on the left-hand side

$$\begin{aligned}
\boldsymbol{Y}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_T + \boldsymbol{Z}_T\right] \\
&\quad + \boldsymbol{E}_T && (2.22)\\
\boldsymbol{X}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\, \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T. && (2.23)
\end{aligned}$$

We will refer to (2.22) and (2.23) as the *reduced form* specification.

2.3 Statistical framework

The DSEM model (2.1)–(2.3) specifies a dynamic relationship among latent and observable variables. Furthermore, we can view the reduced form model (2.22)–(2.23) as a mechanism that generated the observed data $\boldsymbol{V}_T' \equiv (\boldsymbol{Y}_T' : \boldsymbol{X}_T')'$, whose distribution will be our main focus.

Derivation of the density function of $\boldsymbol{V}_T$ can be approached in several ways. Bartholomew and Knott (1999) describe a general theoretical framework for describing the density of the observables given latent variables. Skrondal and Rabe-Hesketh (2004) term this conditional distribution *reduced form distribution* and point out to two general ways of deriving it. In the first approach, the observable variables are assumed to be conditionally independent given latent variables. The second approach specifies multivariate joint density for the observables given latent variables (Skrondal and Rabe-Hesketh 2004, 127).

We take an approach to formal derivation of the joint density of the observable variables using the results from the multinormal theory on distribution of linear forms (Mardia et al. 1979). By considering (2.22)–(2.23) as the mechanism that generates the observable data, we will be able to fully characterize the distribution of $\boldsymbol{V}_T$ by making distributional assumptions only about the unobservable components in (2.22)–(2.23). We firstly make the following assumption.

Assumption 2.3.0.1 (Errors) *The vectors of measurement errors $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\delta}_t$ are homoscedastic Gaussian white noise stochastic processes, uncorrelated with $\boldsymbol{\zeta}_t$ (errors in the structural equation). For $l = \ldots, -1, 0, 1, \ldots$ and $s = \ldots, -1, 0, 1, \ldots$ we require that*

$$E[\boldsymbol{\zeta}_l \boldsymbol{\zeta}_s'] = \begin{cases} \boldsymbol{\Psi}, & l = s \\ \boldsymbol{0} & l \neq s \end{cases}, \quad E[\boldsymbol{\varepsilon}_l \boldsymbol{\varepsilon}_s'] = \begin{cases} \boldsymbol{\Theta}_\varepsilon, & l = s \\ \boldsymbol{0}, & l \neq s \end{cases},$$

$$E[\boldsymbol{\delta}_l \boldsymbol{\delta}_s'] = \begin{cases} \boldsymbol{\Theta}_\delta, & l = s \\ \boldsymbol{0}, & l \neq s \end{cases},$$

where $\boldsymbol{\Psi}$ $(m \times m)$, $\boldsymbol{\Theta}_\varepsilon$ $(n \times n)$, and $\boldsymbol{\Theta}_\delta$ $(k \times k)$ are symmetric positive definite matrices. We also require that $E\left[\boldsymbol{\zeta}_t \boldsymbol{\xi}_{t-s}'\right] = E\left[\boldsymbol{\varepsilon}_t \boldsymbol{\xi}_{t-s}'\right] = E\left[\boldsymbol{\delta}_t \boldsymbol{\xi}_{t-s}'\right] = E\left[\boldsymbol{\zeta}_t \boldsymbol{\varepsilon}_{t-s}'\right] = E\left[\boldsymbol{\zeta}_t \boldsymbol{\delta}_{t-s}'\right] = E\left[\boldsymbol{\delta}_t \boldsymbol{\varepsilon}_{t-s}'\right] = \boldsymbol{0}$, $\forall s$.

The joint distribution of the observable vector $\boldsymbol{V}_T$ (reduced form distribution) can be easily obtained if the observable variables are expressed as a linear function of the Gaussian unobservable random vectors $\boldsymbol{E}_T$, $\boldsymbol{\Delta}_T$, $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$. By Assumption 3.2.0.1 these vectors are mutually independent, hence we will refer to them as to *independent latent components*. The first two latent components of $\boldsymbol{L}_T \equiv (\boldsymbol{E}_T' : \boldsymbol{\Delta}_T' : \boldsymbol{\Xi}_T' : \boldsymbol{Z}_T')'$, i.e., $\boldsymbol{E}_T$ and $\boldsymbol{\Delta}_T$, are the measurement errors, while

$\boldsymbol{\Xi}_T$ contains independent or exogenous and conditioning variables. The status of $\boldsymbol{Z}_T$, the error vector in the structural equation, is less clear-cut. It is not uncommon to specify the structural equation without the error term specially if all variables in the equation are latent. Namely, if the structural equation is a theoretical relationship among unobservable variables, hence something that is assumed to be true in population but is not directly observable, then it might be dubious what is the source of such error. A reasonable explanation would be that $\boldsymbol{Z}_T$ contains all other un-modelled variables, hence it is itself a latent variable. Clearly, to justify the omission of such other variables we need to make very strict assumptions about $\boldsymbol{Z}_T$ requiring it to be a homoscedastic white noise process uncorrelated with independent variables and measurement errors. Thus, statistical properties of $\boldsymbol{Z}_T$ should be the same as those of a classical stochastic error term, though $\boldsymbol{Z}_T$ might be interpreted as a composite of " irrelevant" latent variables.

To fully characterize the distribution of the observable variables we only need to make additional assumptions about the marginal multinormal densities for the independent latent components.

Assumption 2.3.0.2 (Distribution) *Let* $\boldsymbol{\Xi}_T \sim N_{gT}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\Xi})$, $\boldsymbol{Z}_T \sim N_{mT}(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Psi})$, $\boldsymbol{E}_T \sim N_{nT}(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon)$, *and* $\boldsymbol{\Delta}_T \sim N_{kT}(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta)$. *Since* $\boldsymbol{E}_T$, $\boldsymbol{\Delta}_T$, $\boldsymbol{\Xi}_T$, *and* $\boldsymbol{Z}_T$ *are mutually independent,* $E[\boldsymbol{\Xi}_T \boldsymbol{Z}_T']$, $E[\boldsymbol{\Xi}_T \boldsymbol{E}_T']$, $E[\boldsymbol{Z}_T \boldsymbol{E}_T']$, $E[\boldsymbol{\Xi}_T \boldsymbol{\Delta}_T']$, *and* $E[\boldsymbol{Z}_T \boldsymbol{\Delta}_T']$ *are all zero with joint density*

$$\underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T} \sim N_{(n+k+g+m)T} \left[\boldsymbol{0}, \underbrace{\begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}}_{\boldsymbol{\Sigma}_L} \right]. \tag{2.24}$$

Given Assumptions 3.2.0.1 and 2.3.0.2 we can infer the distribution of any linear form in $\boldsymbol{L}_T$ using the following result from the multinormal theory.

Proposition 2.3.0.3 *If* $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and if* $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{c}$, *where* $\boldsymbol{A}$ *is any* $q \times p$ *matrix and* $\boldsymbol{c}$ *is any* q*-vector, then* $\boldsymbol{y} \sim N_q(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{c}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$.

Proof See Theorem 3.1.1. and Theorem 3.2.1 of Mardia et al. (1979, pg. 61-62).

Q.E.D.

Using the above result, and defining the following notation makes possible to obtain different versions of the general DSEM model as simple linear forms in $\boldsymbol{L}_T$.

Definition 2.3.0.4 (Parameters) *Using the simplifying notation*

$$\underbrace{\boldsymbol{A}_{\Xi}^{(1)}}_{nT\times mT} \equiv \underbrace{(\boldsymbol{I}_T \otimes \boldsymbol{A}_y)}_{(nT\times mT)} \underbrace{\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}}_{mT\times mT} \quad \text{and} \quad \underbrace{\boldsymbol{A}_{\Xi}^{(2)}}_{mT\times gT} \equiv \underbrace{\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j}_{mT\times gT} \Rightarrow \underbrace{\boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)}}_{nT\times gT},$$

we define the following matrices of parameters

$$\boldsymbol{P} \equiv \begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{I}_T \otimes \boldsymbol{A}_x & \boldsymbol{0} \end{pmatrix}, \quad \boldsymbol{K}_S \equiv \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{P} \\ \boldsymbol{0} & \boldsymbol{I}_{(g+m)T} \end{pmatrix}, \quad \boldsymbol{K}_R \equiv \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{P} \end{pmatrix}.$$

Denote a linear form by $\boldsymbol{F}_T^{(*)}$ and consider the following two forms

$$\boldsymbol{F}_T^{(S)} = \boldsymbol{K}_S \boldsymbol{L}_T \tag{2.25}$$
$$\boldsymbol{F}_T^{(R)} = \boldsymbol{K}_R \boldsymbol{L}_T. \tag{2.26}$$

It is easy to see that (2.26) corresponds to the reduced model (2.22)–(2.23) hence $\boldsymbol{F}_T^{(R)} = \boldsymbol{V}_T$ can be interpreted as the observable data generated by the linear form $\boldsymbol{K}_R \boldsymbol{L}_T$. On the other hand, $\boldsymbol{F}_T^{(S)}$ includes the latent variables $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ as endogenous or dependent. Models with both observable and latent variables treated as endogenous are commonly termed " structural" (Aigner et al. 1984, Cheng and Van Ness 1999, Wansbeek and Meijer 2000), though this can be easily confused with the structural form of the simultaneous equation system we referred to previously. To avoid confusion with terminology, we will refer to (2.26) as the *reduced structural latent form* (RSLF) model while we will term (2.25) *structural latent form* (SLF) model. The emphases on both models being " latent" will distinguish these forms from the errors-in-variables models that we will analyse in section §5.

We treat all variables except $\boldsymbol{\Xi}_t$ as random, while we will consider both cases with random and fixed $\boldsymbol{\Xi}_t$. The later case requires special consideration as it is obviously not encompassed by the Assumptions 3.2.0.1 and 2.3.0.2, which assume random $\boldsymbol{\Xi}_t$. The model with fixed $\boldsymbol{\Xi}_t$ is generally known as the *functional model* (Wansbeek and Meijer 2000, p. 11) in which no explicit assumptions regarding the distribution of $\boldsymbol{\Xi}_T$ are made and its elements are considered to be unknown fixed parameters or " incidental parameters" (Cheng and Van Ness 1999, p. 3).

Since we can assume that the observable data $\boldsymbol{V}_T$ were generated by linear forms (2.25) and (2.26), or equivalently by the reduced-form equations (2.22) and (2.23), we can let $\boldsymbol{F}_T^{(S)} \equiv (\boldsymbol{Y}_T' : \boldsymbol{X}_T' : \boldsymbol{\Xi}_T' : \boldsymbol{Z}_T')$ and $\boldsymbol{F}_T^{(R)} \equiv (\boldsymbol{Y}_T' : \boldsymbol{X}_T')$. Hence the distribution of the observable variables will be the same as the distribution of the linear form form $\boldsymbol{F}_T^{(R)}$. Now, by Proposition (2.3.0.3) it follows that

$$
\begin{aligned}
\boldsymbol{F}_T^{(S)} &\sim N_{(n+k+g+m)T}\left(\boldsymbol{0}, \boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K}'_S\right) && (2.27)\\
\boldsymbol{F}_T^{(R)} &\sim N_{(n+k)T}\left(\boldsymbol{0}, \boldsymbol{K}_R \boldsymbol{\Sigma}_L \boldsymbol{K}'_R\right) && (2.28)
\end{aligned}
$$

The difference between the structural (2.25) and the reduced (2.26) form is important insofar (2.26) does not model latent variables, i.e., it takes all latent components as independent or exogenous. It might be appealing to think of the reduced model (2.26) as conditional (on latent variables), however, this turns out to be a marginal model with $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ marginalized or integrated out of the likelihood, as we will show in section §2.3.1.

A common argument in the literature (Aigner et al. 1984, Wansbeek and Meijer 2000) used to justify this marginalization is unobservability of the latent variables that necessitates their removal from the model and focusing on (2.26) rather then on (2.25). This justification is apparently motivated by the choice of the estimation methods (e.g. Wishart maximum likelihood), which can handle only the reduced form model (2.26). However, recursive estimation methods using the Kalman filter (Kalman 1960) and the expectation maximisation (EM) algorithm (?) are potentially capable of handling models such as (2.25) and estimating the values of the unobservable variables (Harvey 1989, Durbin and Koopman 2001).

Therefore, marginalization of this kind might not be justified in general, and this matter requires a more formal approach. To tackle this issue, we firstly define the notion of *weak exogeneity* on the lines of Engle et al. (1983) as follows.

Definition 2.3.0.5 (Weak exogeneity) *Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be random vectors with joint density function $f_{xz}(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\omega})$, which can be factorised as the product of the conditional density function of $\boldsymbol{x}$ given $\boldsymbol{z}$ and the marginal density function of $\boldsymbol{z}$,*

$$
f_{xz}(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\omega}) = f_{x|z}(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{\omega}_1) f_z(\boldsymbol{z}; \boldsymbol{\omega}_2), \tag{2.29}
$$

where $\boldsymbol{\omega} \equiv (\boldsymbol{\omega}'_1 : \boldsymbol{\omega}'_2)'$ is the parameter vector and $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are parameter spaces of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$, respectively, with product parameter space

$$
\boldsymbol{\Omega}_1 \times \boldsymbol{\Omega}_2 = \{(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) : \boldsymbol{\omega}_1 \in \boldsymbol{\Omega}_1,\ \boldsymbol{\omega}_2 \in \boldsymbol{\Omega}_2\}
$$

such that $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ have no elements in common, i.e., $\boldsymbol{\omega}_1 \cap \boldsymbol{\omega}_2 = \phi$. Then, $\boldsymbol{z}$ is weakly exogenous for $\boldsymbol{\omega}_1$.

The practical implication of Definition 2.3.0.5 is that if $\boldsymbol{z}$ is weakly exogenous for $\boldsymbol{\omega}_1$, the joint density $f_{x|z}(\boldsymbol{x}|\boldsymbol{z}; \boldsymbol{\omega}_1)$ contains all information about $\boldsymbol{\omega}_1$ and thus the marginal density of $\boldsymbol{z}$ $f_z(\boldsymbol{z}; \boldsymbol{\omega}_2)$ is uninformative about $\boldsymbol{\omega}_1$. The following definition partitions the parameters of the DSEM model (2.1)–(2.3) into non-overlapping subvectors.

Definition 2.3.0.6 Parameters *Let the vector $\boldsymbol{\theta}$ include all unknown parameters of the DSEM model (2.1)-(2.3). We define the following partition*

$$\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Phi_j)} : \boldsymbol{\theta}'^{(\Psi)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_\delta)}\right)'. \tag{2.30}$$

where $\boldsymbol{\theta}^{(B_i)} \equiv \operatorname{vec} \boldsymbol{B}_i$, $\boldsymbol{\theta}^{(\Gamma_j)} \equiv \operatorname{vec} \boldsymbol{\Gamma}_j$, $\boldsymbol{\theta}^{(\Lambda_y)} \equiv \operatorname{vec} \boldsymbol{\Lambda}_y$, $\boldsymbol{\theta}^{(\Lambda_x)} \equiv \operatorname{vec} \boldsymbol{\Lambda}_x$, $\boldsymbol{\theta}^{(\Phi_j)} \equiv \operatorname{vech} \boldsymbol{\Phi}_j$, $\boldsymbol{\theta}^{(\Psi)} \equiv \operatorname{vech} \boldsymbol{\Psi}$, $\boldsymbol{\theta}^{(\Theta_\varepsilon)} \equiv \operatorname{vech} \boldsymbol{\theta}_\varepsilon$, *and* $\boldsymbol{\theta}^{(\Theta_\delta)} \equiv \operatorname{vech} \boldsymbol{\theta}_\delta$; $i = 0, \ldots, p$, $j = 0, \ldots q$.[4]

2.3.1 Structural latent form (SLF)

Given the linear form (2.25) or the SLF model, we are now interested whether the conditional model for the observable variables ($\boldsymbol{V}_T$) given the latent variables contains sufficient information to identify and estimate the model parameters.

By Assumption 2.3.0.2 and Proposition 2.3.0.3 the log-likelihood function of the SLF model is of the form

$$\ell_S\left(\boldsymbol{F}_T^{(S)}; \boldsymbol{\theta}\right) = \alpha - \frac{1}{2} \ln \left|\boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K}_S'\right| - \frac{1}{2} \operatorname{tr} \boldsymbol{L}_T'^{(L)} \left(\boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K}_S'\right)^{-1} \boldsymbol{F}_T^{(S)}, \tag{2.31}$$

where $\alpha \equiv -(n+k+g+m)\frac{T}{2}\ln(2\pi)$. The following proposition shows that the log-likelihood (2.31) can be decomposed into conditional and marginal log-likelihoods hence the likelihood can be expressed as the product of the form given in Definition 2.3.0.5.

Proposition 2.3.1.1 (Likelihood decomposition) *Let (2.31) be the log-likelihood of the structural model (2.25), i.e., the joint log-likelihood of the random vector $\boldsymbol{F}_T^{(S)}$. Denote the conditional log-likelihood of $\boldsymbol{V}_T$ given $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ by $\ell_{V|\Xi,Z}\left(\boldsymbol{V}_T | \boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1\right)$, and the marginal log-likelihoods of $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ by $\ell_\Xi\left(\boldsymbol{\Xi}_T; \boldsymbol{\theta}_2\right)$ and $\ell_Z\left(\boldsymbol{Z}_T; \boldsymbol{\theta}_3\right)$, respectively. Then (2.31) can be factorised as*

$$\ell_S\left(\boldsymbol{F}_T^{(S)}; \boldsymbol{\theta}\right) = \ell_{V|\Xi,Z}\left(\boldsymbol{V}_T | \boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1\right) + \ell_\Xi\left(\boldsymbol{\Xi}_T; \boldsymbol{\theta}_2\right) + \ell_Z\left(\boldsymbol{Z}_T; \boldsymbol{\theta}_3\right), \tag{2.32}$$

where $\boldsymbol{\theta}_1 \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_\delta)}\right)'$, $\boldsymbol{\theta}_2 \equiv \boldsymbol{\theta}^{(\Phi_j)}$, *and* $\boldsymbol{\theta}_3 \equiv \boldsymbol{\theta}^{(\Psi)}$. *Therefore, $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ are weakly exogenous for $\boldsymbol{\theta}_1$.*

Proof See Appendix §2A.

The Proposition 2.32 has interesting implications. Firstly, if all variables were observable, a conditional model with the log-likelihood $\ell_{V|\Xi,Z}\left(\boldsymbol{V}_T | \boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1\right)$ would

[4] We make use of the vech operator for the symmetric matrices, which stacks the columns on and below the diagonal.

provide all information about the the parameters of interest. As remarked above, some recursive algorithms might handle certain special cases with $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$ unobservable, hence by Proposition 2.32 methods based on the conditional likelihood might be justified.

However, larger models might contain too many unknowns which renders the conditional model unfeasible. The commonly used covariance structure and GMM estimators (Hall 2005) require a likelihood in the separated form since these methods aim at minimising the distance between the theoretical and empirical moments. Naturally, to make GMM-type of methods feasible, full separation of the latent and observable variables is necessary. This means the " modelled" variables must be observable and expressible as functions of unobservable variables and unknown parameters.

2.3.2 Reduced structural latent form (RSLF)

The log-likelihood of the RSLF model (2.26) is $(n+k)T$-dimensional Gaussian, thus of the same form as (2.31), though of a lower dimension. The other difference is that $\boldsymbol{F}_T^{(R)}$, unlike $\boldsymbol{F}_T^{(S)}$ in (2.31) does not contain any unobservables. Since $\boldsymbol{F}_T^{(R)} = \boldsymbol{V}_T$, the log-likelihood of the RSLF model is the log-likelihood of the observable data. It is given by

$$\ell_R\left(\boldsymbol{F}_T^{(R)};\boldsymbol{\theta}\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R| - \frac{1}{2}\text{tr}\,\boldsymbol{F}'^{(R)}_T\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\boldsymbol{F}_T^{(R)}. \tag{2.33}$$

It follows that (2.33) will be a closed-form log-likelihood of the RSLF model if a closed-form expression for $\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R$ can be obtained. This would make the RSLF model suitable for GMM-type of estimation.

The following proposition gives a closed form $\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R$, which in turn makes (2.33) a closed-form log-likelihood.

Proposition 2.3.2.1 *Let the covariance structure implied by the DSEM model (2.19)–(2.20) be partitioned as*

$$\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R \equiv \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{2.34}$$

where $\boldsymbol{\Sigma}_{11} \equiv E[\boldsymbol{Y}_T\boldsymbol{Y}'_T]$, $\boldsymbol{\Sigma}_{12} \equiv E[\boldsymbol{Y}_T\boldsymbol{X}'_T]$, *and* $\boldsymbol{\Sigma}_{22} \equiv E[\boldsymbol{X}_T\boldsymbol{X}'_T]$, *which is a function of the parameter vector*

$$\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}'^{(B_i)} : \boldsymbol{\theta}'^{(\Gamma_j)} : \boldsymbol{\theta}'^{(\Lambda_y)} : \boldsymbol{\theta}'^{(\Lambda_x)} : \boldsymbol{\theta}'^{(\Phi_j)} : \boldsymbol{\theta}'^{(\Psi)} : \boldsymbol{\theta}'^{(\Theta_\varepsilon)} : \boldsymbol{\theta}'^{(\Theta_\delta)}\right)',$$

where $\boldsymbol{\theta}^{(B_i)} \equiv \text{vec}\,\boldsymbol{B}_i$, $\boldsymbol{\theta}^{(\Gamma_j)} \equiv \text{vec}\,\boldsymbol{\Gamma}_j$, $\boldsymbol{\theta}^{(\Lambda_y)} \equiv \text{vec}\,\boldsymbol{\Lambda}_y$, $\boldsymbol{\theta}^{(\Lambda_x)} \equiv \text{vec}\,\boldsymbol{\Lambda}_x$, $\boldsymbol{\theta}^{(\Phi_j)} \equiv \text{vech}\,\boldsymbol{\Phi}_j$, $\boldsymbol{\theta}^{(\Psi)} \equiv \text{vech}\,\boldsymbol{\Psi}$, $\boldsymbol{\theta}^{(\Theta_\varepsilon)} \equiv \text{vech}\,\boldsymbol{\theta}_\varepsilon$, *and* $\boldsymbol{\theta}^{(\Theta_\delta)} \equiv \text{vech}\,\Theta_\delta$; $i = 0,\ldots,p$, $j =$

$0, \dots q$.[5] *Then the closed form of the block elements* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, *expressed in terms of the model parameters is given by*

$$\begin{aligned}
\boldsymbol{\Sigma}_{11} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \\
&\times \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right)\right. \\
&\times \left.\left(\sum_{j=0}^{q} \boldsymbol{S'}_T^j \otimes \boldsymbol{\Gamma}_j'\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right]\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon, \qquad (2.35)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{12} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right), \qquad (2.36)
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\Sigma}_{22} &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right) \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) + \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta\right), \qquad (2.37)
\end{aligned}$$

where $\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right) = E\left[\boldsymbol{\Xi}_T \boldsymbol{\Xi}_T'\right] \equiv \boldsymbol{\Sigma}_\Xi$.

Proof See Appendix §2B.

By Proposition 2.32 we have seen that the likelihood of the SLF model (2.25) can be factorised into conditional and marginal likelihoods rendering the latent components $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ weakly exogenous for the parameter sub-vector $\boldsymbol{\theta}_1$. Hence, if $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ were observable we would be able to ignore their marginal distributions without losing any information about $\boldsymbol{\theta}_1$. However, if $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ are not observed, the conditional log-likelihood $\ell_{V|\Xi,Z}\left(\boldsymbol{V}_T|\boldsymbol{\Xi}_T, \boldsymbol{Z}_T; \boldsymbol{\theta}_1\right)$ would not be feasible.

We have obtained the feasible likelihood by using the linear form (2.26) leading to the RSLF model with the log-likelihood (2.33), however, it is easy to see that this comes down to replacing the missing values of $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}$ with their second moment

[5] We make use of the vech operator for the symmetric matrices, which stacks the columns on and below the diagonal.

matrices, which involve the parameter sub-vectors $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$. Thus, obviously the likelihood of the RSLF model will depend on these two parameter sub-vectors. We can still invoke Proposition 2.32 noting that RSLF model (2.26) is a simple reducing linear transformation of the SLF model (2.25) to justify estimation of $\boldsymbol{\theta}_1$ using the RSLF likelihood. However, in this case we will also need to estimate $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$. In conclusion, while weak exogeneity in the sense of Definition 2.3.0.5 holds, we still need to estimate $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ along $\boldsymbol{\theta}_1$, which will require additional knowledge about $\boldsymbol{\theta}$ in the form of parametric restrictions, which cannot be inferred from data alone.

We can easily show that the likelihood of the RSLF model can be obtained by marginalizing the likelihood of the SLF model in respect to the unobservable variables. This can be seen by looking at the relationship between the covariance structures implied by these two models, which is sufficient for the purpose given the shape of their likelihoods is the the same (Gaussian). Thus we have

$$\begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{F}_T^{(S)} = \boldsymbol{F}_T^{(R)}$$

and

$$\boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K'}_S = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ \boldsymbol{\Sigma}_{\Xi} \boldsymbol{A'}_{\Xi}^{(2)} \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{\Sigma}_{\Xi} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda'}_x\right) & \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix},$$

thus it follows that

$$\begin{aligned}
\boldsymbol{K}_R \boldsymbol{\Sigma}_L \boldsymbol{K'}_R &= \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{K}_S \boldsymbol{\Sigma}_L \boldsymbol{K'}_S \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{A}_{\Xi}^{(2)} \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{A}_{\Xi}^{(1)} \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ \boldsymbol{\Sigma}_{\Xi} \boldsymbol{A'}_{\Xi}^{(2)} \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{\Sigma}_{\Xi} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda'}_x\right) & \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} \\ \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{A'}_{\Xi}^{(1)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix} \\
&\times \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.
\end{aligned}$$

The advantage of having the RSLF model with a closed-form covariance structure is in the potential to estimate its parameters by minimising some distance between the theoretical and empirical covariance matrices. On the other hand, the treatment of latent variables as exogenous and observable variables as multinormal, which justified this model in the first place, creates conceptual difficulties in special cases with perfectly observable variables or fixed $\boldsymbol{\xi}_t$. Then, the endogenous observable variables become identical to the exogenous latent variables, which contradicts the statistical assumptions behind the RSLF model. It is thus appealing to entertain the idea behind the errors-in-variables or measurement-errors models (Cheng and Van Ness 1999) where different approach is taken. We will consider this approach in the following section by firstly placing it in the same framework with the models discussed so far.

2.3.3 A restricted RSLF

The approach taken in the errors-in-variables literature is to estimate the structural model (2.1) by replacing each latent variable by a single noisy indicator or a " proxy" variable. Usually, some form of instrumental variables (e.g. other noisy indicators of the latent variables) are used in estimation with the aim of correcting the resulting errors-in-variables bias, and the focus is on evaluating and correcting the bias induced by the measurement error (Cheng and Van Ness 1999).

We will refer to the transformed model in which latent variables are replaced by observable but noisy indicators as the *observed form* (OF) model. To study the OF model we will firstly place it into the general DSEM framework, where each latent variable is measured by multiple indicators. Choosing one indicator per latent variable and normalizing its coefficient (loading) to unity leads to a restricted covariance structure and is thus a special case of the (unrestricted) DSEM covariance structure (2.34) considered above. Clearly, the unit-loading constrains can be used to fix the metric of the latent variable, which has only a re-scaling effect, without affecting the value of the likelihood function.

Imposing unit-loading (UL) restrictions thus leads to a UL-restricted covariance structure. The UL-restrictions are hence parametric restrictions that result in a special case of the general DSEM and so do not invoke a different model or assumptions. The UL-restriction rescales the measurement model for the exogenous latent variables whose indicators can be partitioned as

$$\underbrace{\boldsymbol{X}_T}_{kT\times 1} \equiv (\underbrace{\boldsymbol{X}'^{(\Lambda)}_T}_{(k-g)T\times 1} : \underbrace{\boldsymbol{X}'^{(U)}_T}_{gT\times 1})' \tag{2.38}$$

while the parametric restrictions are imposed as

$$\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x \\ \boldsymbol{I}_{gT} \end{pmatrix}, \tag{2.39}$$

thus resulting in the UL-restricted measurement model

$$\begin{pmatrix} \boldsymbol{X}^{(\Lambda)}_T \\ \boldsymbol{X}^{(U)}_T \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x \\ \boldsymbol{I}_{gT} \end{pmatrix} \boldsymbol{\Xi}_T + \begin{pmatrix} \boldsymbol{\Delta}^{(\Lambda)}_T \\ \boldsymbol{\Delta}^{(U)}_T \end{pmatrix}. \tag{2.40}$$

Therefore, the UL-restricted DSEM model can be written in the reduced form as

$$\begin{aligned}
\boldsymbol{Y}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right)\boldsymbol{\Xi}_T + \boldsymbol{Z}_T\right] \\
&\quad + \boldsymbol{E}_T & (2.41) \\
\boldsymbol{X}_T^{(\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(\Lambda)} & (2.42) \\
\boldsymbol{X}_T^{(U)} &= \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)}, & (2.43)
\end{aligned}$$

where we partitioned $\boldsymbol{X}_T$ into a $gT \times 1$ vector $\boldsymbol{X}_T^{(U)}$ and a $(k-g)T \times 1$ vector $\boldsymbol{X}_T^{(\Lambda)}$. Correspondingly, we have partitioned $\boldsymbol{\Delta}_T$ into sub-vectors $\boldsymbol{\Delta}_T^{(U)}$ and $\boldsymbol{\Delta}_T^{(\Lambda)}$. We partition the measurement error covariance matrix $\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta$ as

$$\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} \\ \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{pmatrix}, \tag{2.44}$$

so $\boldsymbol{\Sigma}_L$ is partitioned as

$$\boldsymbol{\Sigma}_L = \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}. \tag{2.45}$$

Now, if we define

$$\bar{\boldsymbol{K}}_R \equiv \begin{pmatrix} \boldsymbol{I}_{(n+k)T} & \bar{\boldsymbol{P}} \end{pmatrix}, \quad \bar{\boldsymbol{P}} \equiv \begin{pmatrix} \boldsymbol{A}_\Xi^{(1)} \boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \\ \boldsymbol{I}_{gT} & \boldsymbol{0} \end{pmatrix}, \tag{2.46}$$

it follows that

$$\bar{\boldsymbol{F}}_T^{(R)} = \bar{\boldsymbol{K}}_R \boldsymbol{L}_T, \tag{2.47}$$

thus the density of the restricted RSLF model is the same as of the unrestricted model but with different parametrisation, i.e.,

$$\bar{\boldsymbol{F}}_T^{(R)} \sim N_{(n+k)T}\left(\boldsymbol{0}, \bar{\boldsymbol{K}}_R \boldsymbol{\Sigma}_L \bar{\boldsymbol{K}}_R'\right), \tag{2.48}$$

hence we have the log-likelihood of the form

$$\ell_R\left(\bar{\boldsymbol{F}}_T^{(R)};\boldsymbol{\theta}\right) = -(n+k)\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\bar{\boldsymbol{K}}_R \boldsymbol{\Sigma}_L \bar{\boldsymbol{K}}_R'\right| - \frac{1}{2}\mathrm{tr}\, \bar{\boldsymbol{F}}_T'^{(R)}\left(\bar{\boldsymbol{K}}_R \boldsymbol{\Sigma}_L \bar{\boldsymbol{K}}_R'\right)^{-1}\bar{\boldsymbol{F}}_T^{(R)}. \tag{2.49}$$

Next, we partition the covariance matrix (2.34) corresponding to the partition of the data vector $\left(\boldsymbol{Y}_T : \boldsymbol{X}_T^{(\Lambda)} : \boldsymbol{X}_T^{(U)}\right)$ as

$$\bar{\boldsymbol{K}}'_R \bar{\boldsymbol{\Sigma}}_L \bar{\boldsymbol{K}}_R \equiv \begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{11} & \bar{\boldsymbol{\Sigma}}_{12} \\ \bar{\boldsymbol{\Sigma}}'_{12} & \bar{\boldsymbol{\Sigma}}_{22} \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{YY} & \bar{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} & \bar{\boldsymbol{\Sigma}}_{YX}^{(U)} \\ \bar{\boldsymbol{\Sigma}}_{XY}^{(\Lambda)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda,\Lambda)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)} \\ \bar{\boldsymbol{\Sigma}}_{XY}^{(U)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(U\Lambda)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(UU)} \end{pmatrix}. \tag{2.50}$$

The block-elements in (2.50) are as follows,

$$\begin{aligned}
\bar{\boldsymbol{\Sigma}}_{YY} &= \boldsymbol{A}_\Sigma^{(1)}\left(\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{\Sigma}_\Xi \boldsymbol{A}_\Sigma'^{(2)} + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A}_\Sigma'^{(1)} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon \\
\bar{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} &= \boldsymbol{A}_\Sigma^{(1)}\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\
\bar{\boldsymbol{\Sigma}}_{YX}^{(U)} &= \boldsymbol{A}_\Sigma^{(1)}\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{\Sigma}_\Xi \\
\bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} \\
\bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_\Xi + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} \\
\bar{\boldsymbol{\Sigma}}_{XX}^{(UU)} &= \boldsymbol{\Sigma}_\Xi + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}.
\end{aligned}$$

Note that the upper left block element remains the same, i.e.,

$$\boldsymbol{\Sigma}_{YY} = \bar{\boldsymbol{\Sigma}}_{12} = \boldsymbol{A}_\Sigma^{(1)}\left(\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{\Sigma}_\Xi \boldsymbol{A}_\Sigma'^{(2)} + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A}_\Sigma'^{(1)} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon,$$

while the $(1,2)$ block is

$$\begin{aligned}
\bar{\boldsymbol{\Sigma}}_{12} = \left(\bar{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} : \bar{\boldsymbol{\Sigma}}_{YX}^{(U)}\right) &= \boldsymbol{A}_\Sigma^{(1)}\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x : \boldsymbol{I}_{gT}\right) \\
&= \left(\boldsymbol{A}_\Sigma^{(1)}\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{\Sigma}_\Xi\left[\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right] : \boldsymbol{A}_\Sigma^{(1)}\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{\Sigma}_\Xi\right).
\end{aligned}$$

To derive the $(2,2)$ block partition the covariance matrix of the measurement errors as

$$\begin{pmatrix} E[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta}'^{(\Lambda)}_T] & E[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta}'^{(U)}_T] \\ E[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(\Lambda)}_T] & E[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T] \end{pmatrix} = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} \\ \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{pmatrix}$$

so we have:

$$\begin{aligned}
\bar{\boldsymbol{\Sigma}}_{22} &= \begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)} \\ \bar{\boldsymbol{\Sigma}}_{XX}^{(U\Lambda)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(UU)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x \\ \boldsymbol{I}_{gT} \end{pmatrix}\boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x : \boldsymbol{I}_{gT}\right) + \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta\right) \\
&= \begin{pmatrix} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) & \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_\Xi \\ \boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) & \boldsymbol{\Sigma}_\Xi \end{pmatrix} + \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} \\ \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{pmatrix} \\
&= \begin{pmatrix} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_\Xi + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} \\ \boldsymbol{\Sigma}_\Xi\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{\Sigma}_\Xi + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{pmatrix}.
\end{aligned}$$

Finally, note that the marginal covariance structure of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$, i.e.,

$$\begin{pmatrix} \bar{\boldsymbol{\Sigma}}_{YY} & \bar{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} \\ \bar{\boldsymbol{\Sigma}}_{XY}^{(\Lambda)} & \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}$$

is given by

$$\begin{pmatrix} \boldsymbol{A}_{\Sigma}^{(1)}\left(\boldsymbol{A}_{\Sigma}^{(2)}\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A'}_{\Sigma}^{(2)} + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A'}_{\Sigma}^{(1)} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\varepsilon} & \boldsymbol{A}_{\Sigma}^{(1)}\boldsymbol{A}_{\Sigma}^{(2)}\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A'}_{\Sigma}^{(2)}\boldsymbol{A'}_{\Sigma}^{(1)} & \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} \end{pmatrix}$$

2.3.4 Observed form (OF)

Suppose we wish to estimate the DSEM model with the unobservable $\boldsymbol{\Xi}_T$ but instead specify the model by replacing $\boldsymbol{\Xi}_T$ with its noisy indicators $\boldsymbol{X}_T^{(U)}$. This would lead to the model with errors in the variables (EIV). Such model can be interpreted in two ways. Firstly, we can arrive at such model if instead of the true $\boldsymbol{\Xi}_T$ we mistakenly include in the model its noisy indicators, thus introducing the additional error due to mis-measurement (noise), which gives

$$\begin{aligned} \boldsymbol{Y}_T &= \boldsymbol{A}_{\Xi}^{(1)}\left[\boldsymbol{A}_{\Xi}^{(2)}\underbrace{\left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{\Xi}_T} + \mathbf{Z}_T\right] + \boldsymbol{E}_T & (2.51) \\ \boldsymbol{X}_T^{(\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{\left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{\Xi}_T} + \boldsymbol{\Delta}_T^{(\Lambda)} & (2.52) \\ \boldsymbol{X}_T^{(U)} &= \boldsymbol{X}_T^{(U)}. & (2.53) \end{aligned}$$

Alternatively, we can specify the model in its latent form, and use a trivial identity and re-write it as an EIV model, i.e.,

$$\begin{aligned} \boldsymbol{Y}_T &= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{\left(\boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T & (2.54) \\ \boldsymbol{X}_T^{(\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\underbrace{\left(\boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}} + \boldsymbol{\Delta}_T^{(\Lambda)} & (2.55) \\ \boldsymbol{X}_T^{(U)} &= \underbrace{\left(\boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}} + \boldsymbol{\Delta}_T^{(U)}. & (2.56) \end{aligned}$$

In either case, we obtain a DSEM model in the *observed form* (OF), which can be seen as a linear transform

$$\bar{\boldsymbol{K}}_R \boldsymbol{L}_T = \begin{pmatrix} \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I}_{(k-g)T} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{gT} & \boldsymbol{I}_{gT} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \\ \boldsymbol{X}_T^{(U)} \end{pmatrix}. \tag{2.57}$$

We will inspect the OF model (2.57) by comparing its likelihood function to that of the UL-restricted latent form model considered in previous section. In order to do so we will need a simple result on the variance decomposition summarised in the following lemma.

Lemma 2.3.4.1 (Variance decomposition) *Let $\boldsymbol{X}_T^{(U)}$ be a $g \times 1$ vector containing observable indicators of a $g \times 1$ vector of latent variables $\boldsymbol{\Xi}_T$, such that each indicator relates to a single latent variable. We consider the measurement model*

$$\boldsymbol{X}_T^{(U)} = \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)}, \tag{2.58}$$

where $\boldsymbol{\Xi}_T$ can be either random or fixed, while $\boldsymbol{X}_T^{(U)}$ and $\boldsymbol{\Delta}_T^{(U)}$ are both random having some probability distributions $\boldsymbol{X}_T \sim \left(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX}^{(UU)}\right)$ and $\boldsymbol{\Delta}_T^{(U)} \sim \left(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)$, respectively. We make two different sets of assumptions depending on whether $\boldsymbol{\Xi}_T$ is random or fixed as follows.

Random $\boldsymbol{\Xi}_T$ *Suppose $\boldsymbol{\Xi}_T$ has a multivariate probability distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_\Xi$, i.e., $\boldsymbol{\Xi}_T \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_\Xi)$. We assume that*

$$E\left[\boldsymbol{\Xi}_T \boldsymbol{\Delta}_T'^{(U)}\right] = \boldsymbol{0}, \tag{2.59}$$

$$E\left[\boldsymbol{X}_T^{(U)} \boldsymbol{\Delta}_T'^{(U)}\right] \neq \boldsymbol{0}. \tag{2.60}$$

Note that Assumption (2.60) implies a classical rather than Berkson measurement model (Berkson 1950).[6]

Fixed $\boldsymbol{\Xi}_T$ *For non-random $\boldsymbol{\Xi}_T$ we state the Assumption (2.59) in terms of probability limits by treating $\boldsymbol{\Xi}_T$ as a vector of fixed but unobservable constants (incidental parameters). Thus we require that*

$$\underset{T\to\infty}{p\lim} \frac{1}{T} \boldsymbol{\Xi}_T \boldsymbol{\Delta}_T'^{(U)} = \boldsymbol{0}, \tag{2.61}$$

[6] In some cases an additional Assumption that $E\left[\boldsymbol{\Delta}_T^{(\Lambda)} \boldsymbol{\Delta}_T'^{(U)}\right] = \boldsymbol{0}$ can be made, which imposes weaker conditions on the measurement error covariance matrix than classical factor analysis by requiring block-diagonal rather then diagonal $\boldsymbol{\Theta}_\delta$.

In addition, we assume that $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T = \boldsymbol{\Sigma}_{\Xi}$, *hence in the fixed case we consider the unobservable sum of squares* $\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T$, *which is required to converge in probability to some positive definite matrix* $\boldsymbol{\Sigma}_{\Xi}$.[7] *For the random variables* $\boldsymbol{X}_T^{(U)}$ *and* $\boldsymbol{\Delta}_T^{(U)}$ *it trivially follows that* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{X}_T^{(U)}\boldsymbol{X}'^{(U)}_T = \boldsymbol{\Sigma}_{XX}^{(UU)}$ *and* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T = \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}$, *respectively. Also note that assumptions (2.59) and (2.60) imply that* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta}'^{(U)}_T = \boldsymbol{0}$ *and* $\underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T \neq \boldsymbol{0}$.

Then the covariance matrix $\boldsymbol{\Sigma}_{\Xi}$ *(when* $\boldsymbol{\Xi}_T$ *is random), or equivalently, the probability limit of the sum of squares* $\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T$ *(when* $\boldsymbol{\Xi}_T$ *is fixed) can be expressed as*

$$\boldsymbol{\Sigma}_{\Xi} = \boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)} \tag{2.62}$$

Proof From (2.58), using assumptions (2.59) and (2.60), we have

$$\begin{aligned} E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X}'^{(U)}_T\right] &= E\left[\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)'\right] \\ &= E\left[\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T\right] + E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right] \\ &= \boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{aligned} \tag{2.63}$$

and for the fixed case, using Assumption (2.61), equivalently

$$\begin{aligned} \underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{X}_T^{(U)}\boldsymbol{X}'^{(U)}_T &= \underset{T\to\infty}{p\lim}\frac{1}{T}\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)' \\ &= \underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T + \underset{T\to\infty}{p\lim}\frac{1}{T}\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T \\ &= \boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)} \end{aligned} \tag{2.64}$$

hence $\boldsymbol{\Sigma}_{\Xi} = \boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}$, as required.

Q.E.D.

A simple corollary of Lemma (2.3.4.1), i.e., Assumption (2.60), is that

$$\begin{aligned} E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right] &= E\left[\left(\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T^{(U)}\right)\boldsymbol{\Delta}'^{(U)}_T\right] \\ &= E\left[\boldsymbol{\Xi}_T\boldsymbol{\Delta}'^{(U)}_T\right] + E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right] \\ &= \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}, \end{aligned} \tag{2.65}$$

[7] Clearly, the probability limit becomes the simple limit for non-random $\boldsymbol{\Xi}_T$, thus by using the probability limit we cover both cases.

and similarly that $E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta}_T'^{(\Lambda)}\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}$, which would not be the case if (2.58) was a Berkson measurement model. In a Berkson model we would have $E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta}_T'^{(U)}\right] = \boldsymbol{0}$.

We have seen that by Proposition 2.3.1.1 $\boldsymbol{\Xi}_T$ can be treated as weakly exogenous, but we also needed to integrate it out of the likelihood because we could not observe it. On the other hand the OF model, by decomposing $\boldsymbol{\Xi}_T$ into an observable part and the measurement error, potentially makes the conditional model feasible, hence it would be of particular interest to investigate under which conditions is such conditioning valid.

To this end, we firstly define an OF counterpart to the structural form model considered previously. The relationship between the SLF model and a structural observed form (SOF) model can be seen as a linear transform of the form $\boldsymbol{L}_T^{(OF)} = \boldsymbol{D}_{OF}\boldsymbol{L}_T$ for some zero-one transformation matrix $\boldsymbol{D}_{OF}$. It can be verified that

$$\underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T^{OF}} = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{D}_{OF}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T}. \tag{2.66}$$

We can now write the OF-transformed DSEM model as a linear form in $\boldsymbol{L}_T^{(OF)}$ as

$$\underbrace{\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{F}_T^{(OF)}} = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & -\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & -\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{K}_{OF}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T^{(OF)}}. \tag{2.67}$$

Thus we have defined the OF vector as a transformation of the independent latent components vector. Also we defined the transformation that gives the OF-transformed DSEM model. Now note that since $\boldsymbol{F}_T^{(OF)} = \boldsymbol{K}_{OF}\boldsymbol{L}_T^{(OF)}$ we have

$$\boldsymbol{F}_T^{(OF)} = \boldsymbol{K}_{OF}\boldsymbol{D}_{OF}\boldsymbol{L}_T = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{K}_{OF}\boldsymbol{D}_{OF}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{L}_T}, \tag{2.68}$$

which has the effect of trivially decomposing $\boldsymbol{\Xi}_T$ into the observable and unobservable part.

We can obtain the covariance matrix of the OF model as follows. firstly observe we can re-arrange $\boldsymbol{L}_T$ by using a zero-one shifting matrix

$$\boldsymbol{D}_S \equiv \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}, \tag{2.69}$$

and hence obtain

$$\underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{0} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{M}} \underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{L}_T} = \underbrace{\begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Xi}_T + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T \\ \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{M}\boldsymbol{D}_S\boldsymbol{L}_T} = \underbrace{\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{F}_T^{(OF)}}. \tag{2.70}$$

Before proceeding further we will need to make an additional assumption about the measurement errors.

Assumption 2.3.4.2 (Block-diagonal $\boldsymbol{\Theta}_\delta$) *The measurement errors in $\boldsymbol{X}_T^{(U)}$ are uncorrelated with the measurement errors in $\boldsymbol{X}_T^{(\Lambda)}$, hence $\boldsymbol{\Theta}_\delta$ is block-diagonal with $\boldsymbol{\Theta}_{\delta\delta}^{U\Lambda} = \boldsymbol{0}$.*

Now, by making use of the shifting matrix (2.69) and invoking the Assumption 2.3.4.2, we obtain a re-arranged density of $\boldsymbol{L}_T$,

$$\underbrace{\begin{pmatrix} \boldsymbol{E}_T \\ \boldsymbol{\Delta}_T^{(\Lambda)} \\ \boldsymbol{\Xi}_T \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{L}_T} \sim N_{(n+k+g+m)T}\left[\boldsymbol{0}, \underbrace{\begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_{\Xi} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}}_{\boldsymbol{D}_S\boldsymbol{\Sigma}_L\boldsymbol{D}_S'}\right] \tag{2.71}$$

Therefore it follows that

$$\boldsymbol{D}_S\boldsymbol{F}_T^{(OF)} \sim N_{(n+k+g+m)T}\left(\boldsymbol{0}, \boldsymbol{M}\boldsymbol{D}_S\boldsymbol{\Sigma}_L\boldsymbol{D}_S'\boldsymbol{M}'\right). \tag{2.72}$$

Note that without the Assumption 2.3.4.2 we wold have

$$\boldsymbol{D}_S\boldsymbol{\Sigma}_L\boldsymbol{D}'_S = \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}. \quad (2.73)$$

We will see that the block-diagonality Assumption 2.3.4.2 has no effect on the marginal covariance structure (reduced OF model) of $\boldsymbol{Y}_T$, $\boldsymbol{X}_T^{(\Lambda)}$, and $\boldsymbol{X}_T^{(U)}$, but it does have an effect on the conditional distribution of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$. Moreover, the following proposition establishes the validity of the conditional OF model given Assumption 2.3.4.2 holds.

Proposition 2.3.4.3 (OF likelihood decomposition) *Suppose the Assumption 2.3.4.2 holds. Let $\boldsymbol{D}_S\boldsymbol{F}_T^{(OF)} \equiv \boldsymbol{F}_T^*$, hence from* (2.72) *it follows that the log-likelihood of $\boldsymbol{F}_T^*$ is*

$$\ell_{OF}^{(S)}\left(\boldsymbol{F}_T^*;\boldsymbol{\theta}\right) = \alpha - \frac{1}{2}\ln\left|\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}'\right| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{F}_T^*\left(\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}'\right)^{-1}\boldsymbol{F}'^*_T, \quad (2.74)$$

where $\boldsymbol{\Sigma}_L^* \equiv \boldsymbol{D}_S\boldsymbol{\Sigma}_L\boldsymbol{D}'_S$. Let $\ell_{Y,X^\Lambda|X^U,\Delta^U,Z}\left(\boldsymbol{Y}_T,\boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)},\boldsymbol{\Delta}_T^{(U)},\boldsymbol{Z}_T;\boldsymbol{\theta}_1^*\right)$ denote the conditional log-likelihood of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$.

Similarly, let $\ell_{X^U-\Delta^U}\left(\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)};\boldsymbol{\theta}_2^*\right)$ denote the marginal log-likelihood of $\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}$ and denote the marginal log-likelihood of $\boldsymbol{Z}_T$ by $\ell_{\Delta^U}\left(\boldsymbol{\Delta}_T^{(U)};\boldsymbol{\theta}_3^*\right)+\ell_Z\left(\boldsymbol{Z}_T;\boldsymbol{\theta}_4^*\right)$. Then the OF log-likelihood (2.74) can be factorised as

$$\begin{aligned} \ell_{OF}^{(S)}\left(\boldsymbol{F}_T^{(OF)};\boldsymbol{\theta}\right) &= \ell_{Y,X^\Lambda|X^U,\Delta^U,Z}\left(\boldsymbol{Y}_T,\boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)},\boldsymbol{\Delta}_T^{(U)},\boldsymbol{Z}_T;\boldsymbol{\theta}_1^*\right) \quad (2.75)\\ &+ \ell_{X^U-\Delta^U}\left(\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)};\boldsymbol{\theta}_2^*\right)+\ell_{\Delta^U}\left(\boldsymbol{\Delta}_T^{(U)};\boldsymbol{\theta}_3^*\right)+\ell_Z\left(\boldsymbol{Z}_T;\boldsymbol{\theta}_4^*\right), \end{aligned}$$

where $\boldsymbol{\theta}_1^* \equiv \left(\boldsymbol{\theta}'^{(B_i)}:\boldsymbol{\theta}'^{(\Gamma_j)}:\boldsymbol{\theta}'^{(\Lambda_y)}:\boldsymbol{\theta}'^{(\Lambda_x)}:\boldsymbol{\theta}'^{(\Theta_\varepsilon)}:\boldsymbol{\theta}'^{(\Theta_{\delta\delta}^{(\Lambda\Lambda)})}\right)'$, $\boldsymbol{\theta}_2^* \equiv \boldsymbol{\theta}^{(\Phi_j)}$, $\boldsymbol{\theta}_3^* \equiv \boldsymbol{\theta}'^{(\Theta_{\delta\delta}^{(UU)})}$, and $\boldsymbol{\theta}_4^* \equiv \boldsymbol{\theta}^{(\Psi)}$. Thus, $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$ are weakly exogenous for $\boldsymbol{\theta}_1^*$.

Proof See Appendix §2C.

A potentially useful implication of Proposition 2.3.4.3 is the validity of the conditional model for $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$. Unlike the conditional model in latent form considered in Proposition 2.3.1.1, the conditional OF model is feasible since it was formulated by decomposing $\boldsymbol{\Xi}_T$ into an observable and unobservable part. The observable part, $\boldsymbol{X}_T^{(U)}$ can be taken as given, while the unobservable part, $\boldsymbol{\Delta}_T$ needs to be summarised in terms of its second moment matrix.

However, while the likelihood decomposition stated in Proposition 2.3.4.3 enables separation of the conditional model, it does not include a separate expression for the marginal likelihood of $\boldsymbol{X}_T$. Instead, (2.75) includes marginal likelihood of the decomposed $\boldsymbol{\Xi}_T$ into the observable and unobservable parts, i.e., the marginal likelihood of $\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}$. It thus follows that conditioning on $\boldsymbol{X}_T$ in the OF model would be valid in the sense of Definition 2.3.0.5 if the Assumption 2.3.4.2 holds, and if $\boldsymbol{\Delta}_T^{(U)}$ is known or observable (the same goes for $\boldsymbol{Z}_T$, which is always unobservable but can be taken as zero). Not knowing $\boldsymbol{\Delta}_T^{(U)}$ necessitates estimation of its covariance matrix as an additional matrix of parameters $\boldsymbol{\theta}_{\delta\delta}^{(UU)}$. For random $\boldsymbol{X}_T^{(U)}$ this leads us back to the reduced-type of a model and we next show the OF model in the reduced form has the same likelihood (in expectation or in probability limit) as the RSLF model.

Reduced observed form (ROF)

Consider the OF model (2.51)–(2.53). If all variables in the OF model are random with zero mean, it follows that

$$
\begin{aligned}
E\left[\boldsymbol{Y}_T\right] &= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(E\left[\boldsymbol{X}_T^{(U)}\right] - E\left[\boldsymbol{\Delta}_T^{(U)}\right]\right) + \boldsymbol{A}_{\Xi}^{(1)}E\left[\boldsymbol{Z}_T\right] + E\left[\boldsymbol{E}_T\right] = \boldsymbol{0}\\
E\left[\boldsymbol{X}_T^{(\Lambda)}\right] &= \boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\left(E\left[\boldsymbol{X}_T^{(U)}\right] - E\left[\boldsymbol{\Delta}_T^{(U)}\right]\right) + E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\right] = \boldsymbol{0}\\
E\left[\boldsymbol{X}_T^{(U)}\right] &= \boldsymbol{0}.
\end{aligned}
$$

Being a linear combination of normally distributed quantities,

$$
\tilde{\boldsymbol{F}}_T^{(R)} \equiv \left(\boldsymbol{Y}_T' : \boldsymbol{X}_T'^{(\Lambda)} : \boldsymbol{X}_T'^{(U)}\right)'
$$

will have $(n+k)T$-variate multinormal distribution

$$
\tilde{\boldsymbol{F}}_T^{(R)} \sim N_{(n+k)T}\left(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}\right), \tag{2.76}
$$

where $\tilde{\boldsymbol{\Sigma}}$ is defined as

$$
\tilde{\boldsymbol{\Sigma}} \equiv \begin{pmatrix} \tilde{\boldsymbol{\Sigma}}_{YY} & \tilde{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{YX}^{(U)} \\ \tilde{\boldsymbol{\Sigma}}_{XY}^{(\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)} \\ \tilde{\boldsymbol{\Sigma}}_{XY}^{(U)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(U\Lambda)} & \tilde{\boldsymbol{\Sigma}}_{XX}^{(UU)} \end{pmatrix}. \tag{2.77}
$$

Therefore, the likelihood of the OF model (2.51)–(2.53) and the likelihood of the UL-restricted RSLF model will differ only in their covariance matrices $\tilde{\boldsymbol{\Sigma}}$ and $\bar{\boldsymbol{\Sigma}}$.

The following proposition establishes the equivalence of these two matrices either in expectation or in probability limit for the random and fixed cases, respectively.

Proposition 2.3.4.4 (OF equivalence) *Let $\boldsymbol{X}_T^{(U)} = \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T^{(U)}$, where $\boldsymbol{\Xi}_T$ can be either random or fixed. Let $\bar{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\Sigma}}$ be defined by* (2.77) *and* (2.50)*, respectively. For random $\boldsymbol{\Xi}_T$ suppose $\boldsymbol{\Xi}_T$ has a multivariate probability distribution $\boldsymbol{\Xi}_T \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_\Xi)$. Then $E[\tilde{\boldsymbol{\Sigma}}] = E[\bar{\boldsymbol{\Sigma}}]$. In the case when $\boldsymbol{\Xi}_T$ is fixed (non-random) we treat it as a vector of fixed but possibly unobservable constants (incidental parameters), in which case $\operatorname*{plim}_{T\to\infty} \frac{1}{T}\tilde{\boldsymbol{\Sigma}} = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\bar{\boldsymbol{\Sigma}}$.*

Proof See Appendix §2D.

Note that that Proposition 2.3.4.4 did not require the Assumption 2.3.4.2. Therefore, the OF transform of the model with all variables random does not offer any obvious advantage over the RSLF model. The advantage of the OF formulation becomes apparent in the fixed case. Before moving to such model, we briefly make few additional remarks about the random OF model.

The marginal distribution of $\left(\boldsymbol{Y'}_T : \boldsymbol{X'}_T^{(\Lambda)}\right)'$ is $T(n+k-g)$-dimensional Gaussian

$$\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \end{pmatrix} \sim N_{T(n+k-g)}\left[\boldsymbol{0}, \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX}^{(\Lambda)} \\ \boldsymbol{\Sigma}_{XY}^{(\Lambda)} & \boldsymbol{\Sigma}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}\right]$$

with $\boldsymbol{X}_T^{(U)}$ integrated out. The conditional expectation is

$$E\left[\begin{array}{c|c} \boldsymbol{Y}_T & \\ \boldsymbol{X}_T^{(\Lambda)} & \boldsymbol{X}_T^{(U)} \end{array}\right] = \begin{pmatrix} \boldsymbol{\Sigma}_{YX}^{(U)} \\ \boldsymbol{\Sigma}_{XX}^{(\Lambda,UL)} \end{pmatrix} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{X}_T^{(U)} \tag{2.78}$$

$$= \begin{pmatrix} \boldsymbol{\Sigma}_{YX}^{(U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \\ \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \end{pmatrix} \boldsymbol{X}_T^{(U)}, \tag{2.79}$$

and the conditional variance is

$$\operatorname{Var}\left(\begin{array}{c|c} \boldsymbol{Y}_T & \\ \boldsymbol{X}_T^{(\Lambda)} & \boldsymbol{X}_T^{(U)} \end{array}\right) = \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX}^{(\Lambda)} \\ \boldsymbol{\Sigma}_{XY}^{(\Lambda)} & \boldsymbol{\Sigma}_{XX}^{(\Lambda\Lambda)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Sigma}_{YX}^{(U)} \\ \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \end{pmatrix} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \left(\boldsymbol{\Sigma}_{XY}^{(U)} : \boldsymbol{\Sigma}_{XX}^{(U\Lambda)}\right)$$

$$= \begin{pmatrix} \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}^{(U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XY}^{(U)} & \boldsymbol{\Sigma}_{YX}^{(\Lambda)} - \boldsymbol{\Sigma}_{YX}^{(U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XX}^{(U\Lambda)} \\ \boldsymbol{\Sigma}_{XY}^{(\Lambda)} - \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XY}^{(U)} & \boldsymbol{\Sigma}_{XX}^{(\Lambda\Lambda)} - \boldsymbol{\Sigma}_{XX}^{(\Lambda U)} \left(\boldsymbol{\Sigma}_{XX}^{(UU)}\right)^{-1} \boldsymbol{\Sigma}_{XX}^{(U\Lambda)} \end{pmatrix}.$$

Thus it is obvious that conditioning on $\boldsymbol{X}_T^{(U)}$ will be the same as conditioning on $\boldsymbol{\Xi}_T$ in the special case with no measurement error ($\boldsymbol{\Delta}_T^{(U)} = \boldsymbol{0}$).

We now turn to the model with fixed $\boldsymbol{\Xi}_T$. Firstly, consider the standard " functional" model (Wansbeek and Meijer 2000, Cheng and Van Ness 1999), given by

$$\boldsymbol{Y}_T = \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)} \underbrace{\boldsymbol{\Xi}_T}_{\text{fixed part}} + \underbrace{\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T}_{\text{residual}} \tag{2.80}$$

$$\boldsymbol{X}_T = \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{\boldsymbol{\Xi}_T}_{\text{fixed part}} + \underbrace{\boldsymbol{\Delta}_T}_{\text{residual}}, \tag{2.81}$$

which has residual covariance matrix

$$\boldsymbol{\Omega}_F = \begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A}_{\Xi}^{(1)} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta \end{pmatrix}, \tag{2.82}$$

and hence the density function

$$\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T \end{pmatrix} \sim N_{(n+k-g)T}\left[\begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Xi}_T \\ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Xi}_T \end{pmatrix}, \begin{pmatrix} \boldsymbol{A}_{\Xi}^{(1)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A}_{\Xi}^{(1)} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta \end{pmatrix}\right].$$

The log-likelihood of the functional model is then

$$\ell_{Y,X|\Xi}\left(\boldsymbol{Y}_T, \boldsymbol{X}_T|\boldsymbol{\Xi}_T;\boldsymbol{\theta}\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\boldsymbol{\Omega}_F\right|$$
$$-\frac{1}{2}\mathrm{tr}\begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Xi}_T \\ \boldsymbol{X}_T - \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\boldsymbol{\Xi}_T \end{pmatrix}' \boldsymbol{\Omega}_F^{-1} \begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Xi}_T \\ \boldsymbol{X}_T - \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\boldsymbol{\Xi}_T \end{pmatrix}.$$

Note that the log-likelihood (2.83) includes $\boldsymbol{\Xi}_T$, which is unobservable.

Next, consider the OF-transformed model

$$\begin{aligned}
\boldsymbol{Y}_T &= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right) + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)} + \underbrace{\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{E}_T\right)}_{\boldsymbol{U}_T^{(Y)}} &\quad (2.83)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{X}_T^{(\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{X}_T^{(U)} - \boldsymbol{\Delta}_T^{(U)}\right) + \boldsymbol{\Delta}_T^{(\Lambda)} \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} + \underbrace{\left(\boldsymbol{\Delta}_T^{(\Lambda)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Delta}_T^{(U)}\right)}_{\boldsymbol{U}_T^{(X)}} &\quad (2.84)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{X}_T^{(UL)} &= \boldsymbol{X}_T^{(UL)} - \boldsymbol{\Delta}_T^{(UL)} + \boldsymbol{\Delta}_T^{(UL)} \\
&= \boldsymbol{X}_T^{(UL)}, &\quad (2.85)
\end{aligned}$$

and denote the covariance matrix of the residuals $U_T^{(Y)}$ and $U_T^{(X)}$ in (2.83) and (2.84) by

$$\begin{pmatrix} \boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)} \\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}. \tag{2.86}$$

Therefore, the distribution of the OF functional model is given by

$$\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \end{pmatrix} \sim N_{(n+k-g)T}\left[\begin{pmatrix} \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{X}_T^{(U)} \\ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)} \\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}\right], \tag{2.87}$$

and the log-likelihood

$$\ell_{Y,X^\Lambda|X^U}\left(\boldsymbol{Y}_T, \boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)};\boldsymbol{\theta}\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\begin{pmatrix} \boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)} \\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}\right|$$
$$-\frac{1}{2}\mathrm{tr}\begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{X}_T^{(U)} \\ \boldsymbol{X}_T^{(\Lambda)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)} \\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{X}_T^{(U)} \\ \boldsymbol{X}_T^{(\Lambda)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} \end{pmatrix} \tag{2.88}$$

The structure of (2.86) for the special case with $\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \boldsymbol{0}$ (i.e. under Assumption 2.3.4.2) is given by the following proposition.

Proposition 2.3.4.5 *Assume $\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \boldsymbol{0}$. Then the block-elements of (2.86) are given by*

$$\begin{aligned} \boldsymbol{\Omega}_{YY} &= \boldsymbol{A}_\Xi^{(1)}\left(\boldsymbol{A}_\Xi^{(2)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\boldsymbol{A}'^{(2)}_\Xi + \boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)\boldsymbol{A}'^{(1)}_\Xi + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & (2.89)\\ \boldsymbol{\Omega}_{YX}^{(\Lambda)} &= \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) & (2.90)\\ \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} &= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}, & (2.91) \end{aligned}$$

and $\boldsymbol{\Omega}_{XY}^{(\Lambda)} = \boldsymbol{\Omega}'^{(\Lambda)}_{YX}$. Furthermore, it follows that

$$\begin{pmatrix} \boldsymbol{\Omega}_{YY} & \boldsymbol{\Omega}_{YX}^{(\Lambda)} \\ \boldsymbol{\Omega}_{XY}^{(\Lambda)} & \boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} \end{pmatrix} = \boldsymbol{K}_R\bar{\boldsymbol{D}}_S\boldsymbol{\Sigma}_L\bar{\boldsymbol{D}}'_S\boldsymbol{K}'_R, \tag{2.92}$$

where

$$\bar{\boldsymbol{D}}_S \equiv \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}. \tag{2.93}$$

Proof See Appendix §2E.

Using the above result, we can thus simplify the log-likelihood of the functional OF model as

$$\ell_{Y,X^{\lambda}|X^{U}}\left(\boldsymbol{Y}_T, \boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)};\boldsymbol{\theta}\right) = -\tfrac{(n+k-g)T}{2}\ln(2\pi) - \tfrac{1}{2}\ln\left|\boldsymbol{K}_R\bar{\boldsymbol{D}}_S\boldsymbol{\Sigma}_L\bar{\boldsymbol{D}}_S'\boldsymbol{K}_R'\right|$$
$$-\tfrac{1}{2}\mathrm{tr}\begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)} \\ \boldsymbol{X}_T^{(\Lambda)} - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} \end{pmatrix}'\left(\boldsymbol{K}_R\bar{\boldsymbol{D}}_S\boldsymbol{\Sigma}_L\bar{\boldsymbol{D}}_S'\boldsymbol{K}_R'\right)^{-1}\begin{pmatrix} \boldsymbol{Y}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)} \\ \boldsymbol{X}_T^{(\Lambda)} - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{X}_T^{(U)} \end{pmatrix}. \tag{2.94}$$

Therefore, the log-likelihood (2.94) of the functional OF model has gT unknowns less then the log-likelihood of the functional model in latent form as a consequence of not having to estimate $\boldsymbol{\Xi}_T$.

2.3.5 State-space form (SSF)

Various special cases of the general DSEM model have been analysed in the " state-space" form including dynamic factor model and DYMIMIC model (Engle and Watson 1981, Watson and Engle 1983) and the shock-error model (Aigner et al. 1984, Ghosh 1989, Terceiro Lomba 1990). The motivation behind casting particular dynamic models in state-space form is primarily in the possibility of using the Kalman filter algorithm (Kalman 1960) for estimation of the unknown parameters.[8]

The state-space model can be specified in its basic form as

$$\boldsymbol{\vartheta}_t = \boldsymbol{H}\boldsymbol{\vartheta}_{t-1} + \boldsymbol{w}_t, \tag{2.95}$$
$$\boldsymbol{W}_t = \boldsymbol{F}\boldsymbol{\vartheta}_t + \boldsymbol{u}_t, \tag{2.96}$$

where (2.95) is the state equation, (2.96) is the measurement equation, $\boldsymbol{\vartheta}_t$ is the possibly unobservable state vector, and $\boldsymbol{H}$ is the transition matrix (Harvey 1989, Durbin and Koopman 2001).[9] The specification (2.95)–(2.96) is particularly appealing for dynamic models involving unobservable variables since the state equation can contain dynamic unobservable variables and the measurement equation can link them with the observable indicators. These attractive properties of the Kalman filter resulted in numerous empirical papers in the applied statistics and econometric

[8]The Kalman filter was developed by Rudolph E. Kalman as a solution to discrete data linear filtering problem in control engineering. The filter is based on a set of recursive equations, which allow efficient estimation of the state of the process by minimising the mean of the squared error. The Kalman filter recursive algorithm proved to be considerably simpler then the previously available (non-recursive) filters such as the Winer filter, see Brown (1992) for a review.

[9]A simple generalisation of the measurement equation is to include a vector of observable regressors.

literature. Harvey (1989, p. 100), for example, calls the state-space form " an enormously powerful tool which opens the way to handling a wide range of time series models".

To enable estimation of a statistical model by Kalman filter, it is necessary to formulate it in the state-space. We will show that a state-space representation of the general DSEM model (2.1)–(2.3) and hence of all its special cases listed in Table 2.1 exists. In addition, it can be verified that for the transition matrix $\boldsymbol{H}$ to be non-singular we will need to make the following assumption.[10]

Assumption 2.3.5.1 *Let $\boldsymbol{\xi}_t$ follow a VAR(q) process with $q \geq 1$*

$$\boldsymbol{\xi}_t = \sum_{j=1}^{q} \boldsymbol{R}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{v}_t, \tag{2.97}$$

with the roots of $|\boldsymbol{I} - \lambda \boldsymbol{R}_1 - \lambda^2 \boldsymbol{R}_2 - \cdots - \lambda^q \boldsymbol{R}_q| = 0$ greater then one in absolute value and $\boldsymbol{v}_t$ is a Gaussian zero-mean homoscedastic white noise process with $E[\boldsymbol{v}_t \boldsymbol{v}_t'] = \boldsymbol{\Sigma}_v$.

Definition 2.3.5.2 *Let $\boldsymbol{\Pi}_j \equiv (\boldsymbol{I} - \boldsymbol{B}_0)^{-1}\boldsymbol{B}_j$, $\boldsymbol{G}_j \equiv (\boldsymbol{I} - \boldsymbol{B}_0)^{-1}(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_0 \boldsymbol{R}_j)$, and $\boldsymbol{K}_t \equiv (\boldsymbol{I} - \boldsymbol{B}_0)^{-1}(\boldsymbol{\zeta}_t + \boldsymbol{\Gamma}_0 \boldsymbol{v}_t)$, where $\boldsymbol{B}_j$, $\boldsymbol{\Gamma}_j$, and $\boldsymbol{\zeta}_t$ are defined as in* (2.1)–(2.3).

The following result establishes the existence of the state-space form of the general DSEM model given Assumption 2.97.

Proposition 2.3.5.3 *Let $\boldsymbol{\xi}_t$ be generated by a VAR (q) process as in* (2.97). *Then the general DSEM model* (2.1)–(2.3) *can be written in the state-space form* (2.95)–(2.96) *as*

$$\begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \\ \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \vdots \\ \boldsymbol{\eta}_{t-r+1} \\ \boldsymbol{\xi}_{t-r+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{G}_1 & \cdots & \boldsymbol{\Pi}_{r-1} & \boldsymbol{G}_{r-1} & \boldsymbol{\Pi}_r & \boldsymbol{G}_r \\ \boldsymbol{0} & \boldsymbol{R}_1 & \cdots & \boldsymbol{0} & \boldsymbol{R}_{r-1} & \boldsymbol{0} & \boldsymbol{R}_r \\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\eta}_{t-2} \\ \boldsymbol{\xi}_{t-2} \\ \vdots \\ \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{K}_t \\ \boldsymbol{v}_t \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \tag{2.98}$$

and

[10]Since the state-space representation is achieved by dynamically linking the current state with the past-period state via a first-order Markov process, the first equation (for time t) is the actual model, while the rest of the stacked elements (for time $t-1, t-2, \ldots, t-q$) of $\boldsymbol{\vartheta}_t$ are set trivially equal to themselves as they appear in both $\boldsymbol{\vartheta}_t$ and $\boldsymbol{\vartheta}_{t-1}$. Hence, if any of the elements of $\boldsymbol{\vartheta}_t$ cannot be related to an element of $\boldsymbol{\vartheta}_{t-1}$ (such as in the case of white unobservable regressors) the transition matrix $\boldsymbol{H}$ will contain a row of zeros and thus it will be singular.

$$\begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{x}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_y & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_x & \cdots & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \\ \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \vdots \\ \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\delta}_t \end{pmatrix}, \qquad (2.99)$$

where $r = \max(p, q)$*, with notation defined in 2.3.5.2.*

Proof See Appendix §2F.

While Proposition 2.3.5.3 gives the state-space form of the general DSEM model, it is not immediately clear how the state-space form compares with the forms considered earlier. Namely, the SSF model (2.98) is in a recursive form required for the Kalman filter, hence it is specified in t-notation. On the other hand the T-notation (Table 2.2) we used to analyse the statistical properties of other DSEM forms leads to a closed-form rather then a recursive form of the model. Nevertheless, we can write the SSF model (2.98) for the process ($t = 1, 2 \ldots, T$) and compare its likelihood with those of the other forms of the model. In the context of the RSLF model, for example, this would call for additional modelling of the VAR(q) process for $\boldsymbol{\Xi}_T$, thereby increasing the dimensionality of the multivariate density function from $(n + k)T$ to $(n + k + g)T$. However, we will show that such extended model can still be reduced to the $(n + k)T$-dimensional model.

Given the VAR(q) process for $\boldsymbol{\Xi}_T$ (Assumption 2.97), the SLF model will have to include an additional equation for $\boldsymbol{\Xi}_T$. The structural equation remains as before and it can be reduce as

$$\begin{aligned} \boldsymbol{H}_T &= \left(\sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right) \boldsymbol{H}_T + \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_T + \mathbf{Z}_T \\ &= \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_T + \mathbf{Z}_T\right]. \end{aligned} \qquad (2.100)$$

A T-notation equivalent of the VAR(q) model (2.97) can be written as

$$\boldsymbol{\Xi}_T = \left(\sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right) \boldsymbol{\Xi}_T + \boldsymbol{\Upsilon}_T = \underbrace{\left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1}}_{\boldsymbol{A}_\Sigma^{(3)}} \boldsymbol{\Upsilon}_T. \qquad (2.101)$$

Finally, the measurement equations as as before

$$\begin{aligned} \boldsymbol{Y}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\, \boldsymbol{H}_T + \boldsymbol{E}_T && (2.102) \\ \boldsymbol{X}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)\, \boldsymbol{\Xi}_T + \boldsymbol{\Delta}_T. && (2.103) \end{aligned}$$

Substituting (2.100) and (2.101) in (2.102) and (2.103), respectively, we obtan the reduced SSF model

$$\begin{aligned} \begin{pmatrix} \boldsymbol{H}_T \\ \boldsymbol{\Xi}_T \end{pmatrix} &= \begin{pmatrix} \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j & \sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j \\ \boldsymbol{0} & \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j \end{pmatrix} \begin{pmatrix} \boldsymbol{H}_T \\ \boldsymbol{\Xi}_T \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_T \\ \boldsymbol{\Upsilon}_T \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j & -\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j \\ \boldsymbol{0} & \boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z}_T \\ \boldsymbol{\Upsilon}_T \end{pmatrix}, \quad (2.104) \end{aligned}$$

where the inverse of the matrix of parameters in (2.104) is given by[11]

$$\begin{pmatrix} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} & \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \\ \boldsymbol{0} & \left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \end{pmatrix}, \quad (2.105)$$

therefore the reduced SSF model becomes

$$\begin{aligned} \boldsymbol{Y}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \\ &\times \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \boldsymbol{\Upsilon}_T + \mathbf{Z}_T\right] + \boldsymbol{E}_T \quad (2.106) \\ \boldsymbol{X}_T &= (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x) \left(\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j\right)^{-1} \boldsymbol{\Upsilon}_T + \boldsymbol{\Delta}_T. \quad (2.107) \end{aligned}$$

Using the simplifying notation from Definition 2.3.0.4 we can write (2.106) and (2.107) as

[11]We make use of the result

$$\begin{pmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \boldsymbol{0} & \boldsymbol{D}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{D}_{11}^{-1} & -\boldsymbol{D}_{11}^{-1}\boldsymbol{D}_{12}\boldsymbol{D}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{D}_{22}^{-1} \end{pmatrix}.$$

$$\boldsymbol{Y}_T = \boldsymbol{A}_\Sigma^{(1)}\boldsymbol{A}_\Sigma^{(2)}\underbrace{\boldsymbol{A}_\Sigma^{(3)}\boldsymbol{\Upsilon}_T}_{\boldsymbol{\Xi}_T}+\boldsymbol{A}_\Sigma^{(1)}\mathbf{Z}_T+\boldsymbol{E}_T \tag{2.108}$$

$$\boldsymbol{X}_T = (\boldsymbol{I}_T\otimes\boldsymbol{A}_x)\underbrace{\boldsymbol{A}_\Sigma^{(3)}\boldsymbol{\Upsilon}_T}_{\boldsymbol{\Xi}_T}+\boldsymbol{\Delta}_T \tag{2.109}$$

Next, we consider the covariance structure of $\boldsymbol{\Xi}_T$, which can be easily obtained from the reduced form T-notation expression (2.101). The following lemma gives the required expression.

Lemma 2.3.5.4 *Consider the VAR process* (2.101). *By Assumption 2.3.5.1,* $E[\boldsymbol{v}_t\boldsymbol{v}_t'] = \boldsymbol{\Sigma}_v \Rightarrow E[\boldsymbol{\Upsilon}_T\boldsymbol{\Upsilon}'_T] \equiv \boldsymbol{I}_T\otimes\boldsymbol{\Sigma}_v$. *Then* $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)}(\boldsymbol{I}_T\otimes\boldsymbol{\Sigma}_v)\boldsymbol{A}'^{(3)}_\Sigma$, *where* $\boldsymbol{A}_\Sigma^{(3)} \equiv (\boldsymbol{I}_{gT}-\sum_{j=1}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{R}_j)^{-1}$.

Proof Since $\boldsymbol{\Xi}_T = \boldsymbol{A}_\Sigma^{(3)}\boldsymbol{\Upsilon}_T$ and $E[\boldsymbol{\Xi}_T\boldsymbol{\Xi}'_T] \equiv \boldsymbol{\Sigma}_\Xi$, we have $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)}E[\boldsymbol{\Upsilon}_T\boldsymbol{\Upsilon}'_T]\boldsymbol{A}'^{(3)}_\Sigma = \boldsymbol{A}_\Sigma^{(3)}E[\boldsymbol{\Upsilon}_T\boldsymbol{\Upsilon}'_T]\boldsymbol{A}'^{(3)}_\Sigma = \boldsymbol{A}_\Sigma^{(3)}(\boldsymbol{I}_T\otimes\boldsymbol{\Sigma}_v)\boldsymbol{A}'^{(3)}_\Sigma$, as required.

Q.E.D

To examine the likelihood of the SSF model firstly note that the reduced SSF model (2.104) is $(n+k)T$-dimensional, thus the SSF likelihood will be $(n+k)T$-variate Gaussian, thus of the same form and dimension as the likelihood of the RSLF model given by (2.33). Recall that by Assumption (2.97) as shown in (2.101), the $\boldsymbol{\Xi}_T$ process can be expressed as a linear function of the residual vector $\boldsymbol{\Upsilon}_T$. Defining $\boldsymbol{L}_T^{SSF} \equiv (\boldsymbol{E}'_T : \boldsymbol{\Delta}'_T : \boldsymbol{\Upsilon}'_T : \mathbf{Z}'_T)'$, assuming $\boldsymbol{\Upsilon}_T$ is Gaussian and independent of other latent components it follows that

$$\underbrace{\begin{pmatrix}\boldsymbol{E}_T\\ \boldsymbol{\Delta}_T\\ \boldsymbol{\Upsilon}_T\\ \mathbf{Z}_T\end{pmatrix}}_{\boldsymbol{L}_T^{SSF}} \sim N_{(n+k+g+m)T}\left[\boldsymbol{0}, \underbrace{\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Sigma}_v & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Psi}\end{pmatrix}}_{\boldsymbol{\Sigma}_{SSF}}\right]. \tag{2.110}$$

Now, by letting

$$\boldsymbol{K}_{SSF} = \begin{pmatrix}\boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{A}_\Sigma^{(1)}\boldsymbol{A}_\Sigma^{(2)}\boldsymbol{A}_\Sigma^{(3)} & \boldsymbol{A}_\Sigma^{(1)}\\ \boldsymbol{0} & \boldsymbol{I}_{kT} & (\boldsymbol{I}_T\otimes\boldsymbol{A}_x)\boldsymbol{A}_\Sigma^{(3)} & \boldsymbol{0}\end{pmatrix}, \tag{2.111}$$

the reduced SSF model (2.106)–(2.107) can be written as a linear form in $\boldsymbol{L}_T^{SSF}$, i.e., as $\boldsymbol{K}_{SSF}\boldsymbol{L}_T^{SSF}$. Therefore, by Proposition 2.3.0.3 it follows that

$$\boldsymbol{L}_T^{SSF} \sim N_{(n+k+g+m)T}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{SSF}\right) \Rightarrow \boldsymbol{K}_{SSF}\boldsymbol{L}_T^{SSF} \sim N_{(n+k)T}\left(\boldsymbol{0}, \boldsymbol{K}_{SSF}\boldsymbol{\Sigma}_{SSF}\boldsymbol{K}'_{SSF}\right).$$

Finally, since $\boldsymbol{\Xi}_T$ is a VAR(q) process by Assumption 2.3.5.1 whose covariance structure, by Lemma 2.3.5.4, is $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v\right)\boldsymbol{A}'^{(3)}_\Sigma$, we can parametrise $\boldsymbol{\Sigma}_L$ as

$$\boldsymbol{\Sigma}_L = \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \underbrace{\boldsymbol{A}_\Sigma^{(3)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v\right)\boldsymbol{A}'^{(3)}_\Sigma}_{\boldsymbol{\Sigma}_\Xi} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}. \tag{2.112}$$

Therefore, modelling the $\boldsymbol{\Xi}_T$ process as a VAR(q) imposes the parametrisation $\boldsymbol{\Sigma}_\Xi = \boldsymbol{A}_\Sigma^{(3)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v\right)\boldsymbol{A}'^{(3)}_\Sigma$ on $\boldsymbol{\Sigma}_L$. Hence with such structure imposed on $\boldsymbol{\Sigma}_L$ it can be easily verified that $\boldsymbol{K}_{SSF}\boldsymbol{\Sigma}_{SSF}\boldsymbol{K}'_{SSF} = \boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R$, thus the likelihood of the reduced SSF model (2.106)–(2.107) is equal to the likelihood of the RSLF model (2.26) with the covariance matrix of $\boldsymbol{\Xi}_T$ parametrised as $\boldsymbol{A}_\Sigma^{(3)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v\right)\boldsymbol{A}'^{(3)}_\Sigma$.

2.4 Comparison of different forms

There are several possible criteria on which to compare different forms of the general DSEM model. We have seen that different forms of the general model discussed in this section are not identical re-arrangements of the same model in the statistical sense. Rather different assumptions about the modelled variables had to be made as well as some specific parametrisations needed to be considered. In this respect, while substantively we are dealing with the same model, its different forms might favour certain estimation methods and applications over the others. In particular, we would be interested in the criteria such as 1) choice of estimation method, 2) identification of the parameters, and 3) statistical assumptions about modelled variables. We will look into some of these criteria, in turn, by focusing on particular forms of the general model.

RSLF model. The RSLF model (section §2.3.2) has appealing implications when repeated observations on the time series process $\boldsymbol{F}_{iT}^{(R)}$ are available. Consider N independent realizations of $\boldsymbol{F}_{iT}^{(R)}$ are being observed. Then the log-likelihood (2.33) can be written for a single realization as

$$\ell_R\left(\boldsymbol{F}_{iT}^{(R)};\boldsymbol{\theta}\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{F}'^{(R)}_{iT}\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\boldsymbol{F}_{iT}^{(R)}, \tag{2.113}$$

thus for N independent realizations, the log-likelihood becomes

$$\ell_R\left(\boldsymbol{F}_{NT}^{(R)};\boldsymbol{\theta}\right) = -\frac{(n+k)NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{F}'^{(R)}_{NT}\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\boldsymbol{F}_{NT}^{(R)}, \tag{2.114}$$

where $\boldsymbol{F}_{NT}^{(R)} \equiv (\boldsymbol{F}_{1T}^{(R)}, \ldots, \boldsymbol{F}_{NT}^{(R)})$. Now, ignoring the constant term and rearranging the matrices under the trace, and multiplying by $-2/N$ yields

$$\ln|\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R| + \frac{1}{2}\mathrm{tr}\,\left(\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)^{-1}\frac{1}{N}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}, \tag{2.115}$$

which can be minimised to obtain the maximum likelihood estimates of the model parameters. Inspecting (2.115), we can observe that that $N^{-1}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$ is the empirical covariance matrix of the observable data, hence making (2.115) is a closed form likelihood. The log-likelihood (2.115) is asymptotically equivalent to the Wishart log-likelihood of $(N-1)^{-1}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$, the empirical covariance matrix of the observable data.

Alternatively, an assumption that the observable variables are multivariate Gaussian along with the independence of the N realizations (hence independence of the columns of $\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$ would imply a Wishart distribution of $(N-1)^{-1}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT}$, hence a log-likelihood different from (2.115) only in a scaling constant.

The availability of the closed-form covariance structure 2.3.2.1 implied by the RSLF model also motivates generalised methods of moments or weighted least squares type of estimators. Consider a quadratic form in a positive definite matrix $\boldsymbol{W}$,

$$\left(\mathrm{vech}\,\frac{1}{N}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT} - \mathrm{vech}\,\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right)\boldsymbol{W}^{-1}\left(\mathrm{vech}\,\frac{1}{N}\boldsymbol{F}_{NT}^{(R)}\boldsymbol{F}'^{(R)}_{NT} - \mathrm{vech}\,\boldsymbol{K}_R\boldsymbol{\Sigma}_L\boldsymbol{K}'_R\right). \tag{2.116}$$

Clearly, (2.116) is a fairly general fitting function not depending on any distributional assumptions. Various different choices of $\boldsymbol{W}$ might be considered.

The RSLF model summarises the information about the latent variables in terms of their population moments and hence does not require estimation of the unobservable vectors $\boldsymbol{\Xi}_T$ and $\boldsymbol{H}_T$. Table 2.3 lists the matrices of parameters in the RSLF model.

OF model. In section §2.3.4 the reduced structural OF model was shown to be equivalent to the UL-restricted random-case RSLF model, obviously the number of parameters to be estimated will be the same hence we do not need to consider the random-case OF model. It is thus more interesting to compare the fixed OF model with the RSLF model. The OF model has an immediate advantage of encompassing

Table 2.3: Matrices of parameters in different model forms

Vector/matrix	Dimension	Number of parameters	RSLF	OF	SSF
$\boldsymbol{H}_T$	$mT \times 1$	mT	–	–	✓
$\boldsymbol{\Xi}_T$	$gT \times 1$	gT	–	–	✓
$\boldsymbol{\Psi}$	$m \times m$	$m(m+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}_\varepsilon$	$n \times n$	$n(n+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}_\delta$	$k \times k$	$k(k+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}$	$(k-g) \times (k-g)$	$(k-g)k-g+1)/2$	✓	✓	✓
$\boldsymbol{\Theta}_{\delta\delta}^{(UU)}$	$g \times g$	$g(g-1)/2$	✓	✓	✓
$\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}$	$(k-g) \times g$	$g(k-g)$	✓	–	✓
$\boldsymbol{A}_{\Xi}^{(1)}$	$nT \times nT$	$m(n+pm-1)$	✓	✓	✓
$\boldsymbol{A}_{\Xi}^{(2)}$	$mT \times gT$	qmg	✓	✓	✓
$\boldsymbol{A}_{\Xi}^{(3)}$	$gT \times gT$	qg^2	–	–	✓
$\boldsymbol{\Sigma}_{\Xi}$	$gT \times gT$	$g^2(q+1/2+1/g)$	✓	–	–
$\boldsymbol{\Sigma}_{XX}^{(UU)}$	$gT \times gT$	$g^2(q+1/2+1/g)$	–	✓	–
$\boldsymbol{\Sigma}_v$	$g \times g$	$\frac{1}{2}g(g+1)$	–	–	✓
$\boldsymbol{A}_{\Sigma}^{(3)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v\right)\boldsymbol{A}_{\Sigma}'^{(3)}$	$gT \times gT$	$g^2(q+1/2+1/g)$	–	–	✓

the cases with fixed observables or perfectly observable indicators of fixed latent variables. The OF model can be estimated with maximum likelihood, but it also facilitates instrumental variables estimators, hence the assumption of multivariate normality can be relaxed easily in the context of the OF model.

Another interesting feature of the OF model is its suitability for estimation of DSEM models with pure time series. We have pointed out to a straightforward estimation method for the RSLF model when a cross-section time series data is available. While maximisation of (2.33) for a single realization of $\boldsymbol{F}_{iT}^{(R)}$ might be considered, the OF model suggests a more feasible approach. Namely, the log-likelihood (2.88) is of a standard multivariate Gaussian form but with parametrised residual covariance matrix, hence it would be straightforward to maximise it in respect to the model parameters.

SSF model. In section §2.3.5 we have shown that a state-space form of the general DSEM model requires modelling $\boldsymbol{\Xi}_T$ as a VAR(q) process. This imposes a parametric structure on the covariance matrix of $\boldsymbol{\Xi}_T$ given by Lemma 2.3.5.1. Specifically, estimating coefficients of a VAR(q) process for $\boldsymbol{\Xi}_T$ has the effect of imposing parametric structure $\boldsymbol{A}_{\Sigma}^{(3)}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v\right)\boldsymbol{A}_{\Sigma}'^{(3)}$ on $\boldsymbol{\Sigma}_{\Xi}$. Without modelling the $\boldsymbol{\Xi}_T$ as a VAR(q) we defined $\boldsymbol{\Sigma}_{\Xi}$ unconstrained with bound-Toeplitz structure owing to covariance stationarity of $\boldsymbol{\Xi}_T$ (see Appendix B). By Proposition 2.3.2.1 we had $\boldsymbol{\Sigma}_{\Xi} = \boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S}_T'^j \otimes \boldsymbol{\Phi}_j'\right)$, where $\boldsymbol{\Phi}_0$ is symmetric $g \times g$ matrix with $g(g+1)/2$ distinct elements. Similarly, for $j = 1, 2, \ldots, q$, $\boldsymbol{\Phi}_j$ is

$g \times g$ with g^2 distinct elements. Thus $\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right)$ has $qg^2 + g(g+1)/2 = g^2(q+1/2+1/g)$ distinct elements. On the other hand, imposing a VAR(q) structure on $\boldsymbol{\Sigma}_\Xi$ results in parametrisation of the covariance matrix of $\boldsymbol{\Xi}_T$ given by $\boldsymbol{A}_\Sigma^{(3)}(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_v)\boldsymbol{A'}_\Sigma^{(3)}$, where $\boldsymbol{A}_\Sigma^{(3)} \equiv (\boldsymbol{I}_{gT} - \sum_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{R}_j)^{-1}$. Hence we have q $g \times g$ matrices $\boldsymbol{R}_j$, each having g^2 elements, and a symmetric $g \times g$ matrix $\boldsymbol{\Sigma}_v$ with $g(g+1)/2$ distinct elements. Thus, the VAR(q) parametrisation of $\boldsymbol{\Sigma}_\Xi$ results in the same number of distinct elements of $\boldsymbol{\Sigma}_\Xi$, namely $g^2(q+1/2+1/g)$.

However, when the aim of the SSF specification is the application of the Kalman filter, then the model needs to be in its recursive form (i.e., t-notation) given by (2.98), therefore, the state vector that includes $\boldsymbol{\Xi}_T$ and $\boldsymbol{H}_T$ will be treated as a vector of missing values, thus requiring estimation of additional $(n+k)T$ parameters. Recall this was not the case in the RSLF model which used the summary information about these vectors in the form of their second moment matrices.

Chapter 3

Maximum likelihood estimation with panel data

3.1 Introduction

The methods for estimating static simultaneous equation models (SEM) containing unobservable (latent) variables or variables measured with error are widely available and frequently used in the applied literature. Bartholomew and Knott (1999) and Wansbeek and Meijer (2000) provide a comprehensive review of these methods. Panel data methods for models with latent variables or with errors-in-variables have been considered in the literature in the context of the instrumental variables (IV) and the generalised method of moments (GMM) estimation (Arellano and Bover 1995, Wansbeek 2001, Arellano 2003, Hsiao 2003). Moreover, static panel random effects models with latent variables can be estimated in the standard SEM modelling framework using the covariance structure analysis methods of Jöreskog (1981) and Jöreskog and Sörbom (1996a); see e.g. Aasness et al. (1993) and Aasness et al. (1995) for empirical applications.

On the other hand, dynamic panel models with latent variables have not been extensively analysed and there is a lack of suitable estimation methods for dynamic simultaneous equation models with latent variables or with all variables measured with error. Single equation and systems IV estimators were suggested by Cziráky (2004d) for time series and random effects panel models.

We consider estimation of dynamic simultaneous equation panel models with latent variables and fixed effects. Such models include unobservable variables that are measurable by multiple observable indicators. We consider full information maximum likelihood estimation, which has the potential advantages over the non-parametric IV and GMM methods in respect to modelling and testing the implied (latent) structure rather then merely providing consistent estimates of the structural parameters. This is an important aspect in the economic applications where the substantive theory is formulated in terms of the latent variables where the measurement of these variables as well as the structural relationships are tested.

3.2 Dynamic panel structural equation model

In this section we consider a dynamic panel simultaneous equation model with latent variables and fixed effects (DPSEM(p,q)). A DPSEM(p,q) model for the individual $i = 1,\ldots,N$ at time $t = 1,\ldots,T$ can be written for the generic individual at any time period t using the " t-notation" as

$$\begin{aligned} \boldsymbol{\eta}_{it} &= \sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{it-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{it-j} + \boldsymbol{\zeta}_{it} && (3.1) \\ \boldsymbol{y}_{it} &= \boldsymbol{\Lambda}_y \boldsymbol{\eta}_{it} + \boldsymbol{\mu}_{yi} + \boldsymbol{\varepsilon}_{it} && (3.2) \\ \boldsymbol{x}_{it} &= \boldsymbol{\Lambda}_x \boldsymbol{\xi}_{it} + \boldsymbol{\mu}_{xi} + \boldsymbol{\delta}_{it} && (3.3) \end{aligned}$$

where $\boldsymbol{\eta}_{it} = \left(\eta_{it}^{(1)}, \eta_{it}^{(2)}, \ldots, \eta_{it}^{(m)}\right)'$ and $\boldsymbol{\xi}_{it} = \left(\xi_{it}^{(1)}, \xi_{it}^{(2)}, \ldots, \xi_{it}^{(g)}\right)'$ are vectors of latent variables, $\boldsymbol{y}_{it} = \left((y_{it}^{(1)}, y_{it}^{(2)}, \ldots, y_{it}^{(n)}\right)'$ and $\boldsymbol{x}_{it} = \left(x_{it}^{(1)}, x_{it}^{(2)}, \ldots, x_{it}^{(k)}\right)'$ are vectors of observable variables, and $\boldsymbol{B}_j$ $(m \times m)$, $\boldsymbol{\Gamma}_j$ $(m \times g)$, $\boldsymbol{\Lambda}_x$ $(k \times g)$, and $\boldsymbol{\Lambda}_y$ $(n \times m)$ are coefficient matrices. The contemporaneous and simultaneous coefficients are in $\boldsymbol{B}_0$, and $\boldsymbol{\Gamma}_0$, while $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, ..., $\boldsymbol{B}_p$, and $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, ..., $\boldsymbol{\Gamma}_q$ contain coefficients of the lagged endogenous and exogenous latent variables. Finally, $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ are the $n \times 1$ and $k \times 1$ vectors of individual means, respectively. We treat $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ as vectors of coincidental (fixed) parameters, which makes the DPSEM model (3.1)-(3.3) a " fixed-effects" panel model. The statistical assumptions about the variables in (3.1)–(3.3) are as follows.

Assumption 3.2.0.1 *The vectors of measurement errors $\boldsymbol{\varepsilon}_{it}$ and $\boldsymbol{\delta}_{it}$ are homoscedastic Gaussian white noise stochastic processes, uncorrelated with the errors in the structural model ($\boldsymbol{\zeta}_{it}$). We require for $l = \ldots,-1,0,1,\ldots$ and $s = \ldots,-1,0,1,\ldots$ that*

$$E\left[\boldsymbol{\zeta}_{il}\boldsymbol{\zeta}_{js}'\right] = \begin{cases} \boldsymbol{\Psi}, & l = s, i = j \\ \boldsymbol{0}, & l \neq s, i \neq j \end{cases}, \quad E\left[\boldsymbol{\varepsilon}_{il}\boldsymbol{\varepsilon}_{js}'\right] = \begin{cases} \boldsymbol{\Theta}_{\varepsilon}, & l = s, i = j \\ \boldsymbol{0}, & l \neq s, i \neq j \end{cases},$$
$$E\left[\boldsymbol{\delta}_{il}\boldsymbol{\delta}_{js}'\right] = \begin{cases} \boldsymbol{\Theta}_{\delta}, & l = s, i = j \\ \boldsymbol{0}, & l \neq s, i \neq j \end{cases},$$

where $\boldsymbol{\Psi}$ $(m \times m)$, $\boldsymbol{\Theta}_{\varepsilon}$ $(n \times n)$, and $\boldsymbol{\Theta}_{\delta}$ $(k \times k)$ are symmetric positive definite matrices. We also require that $E\left[\boldsymbol{\zeta}_{it}\boldsymbol{\xi}_{jt-s}'\right] = E\left[\boldsymbol{\zeta}_{it}\boldsymbol{\varepsilon}_{jt-s}'\right] = E\left[\boldsymbol{\zeta}_{it}\boldsymbol{\delta}_{jt-s}'\right] = E\left[\boldsymbol{\delta}_{it}\boldsymbol{\varepsilon}_{jt-s}'\right] = \boldsymbol{0}$, $\forall s$.

Finally, using the notation from Table 2.2, the DPSEM model DPSEM model (3.1)-(3.3) can now be written for the individual i as

$$
\begin{aligned}
\boldsymbol{H}_{iT} &= \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_{iT} + \boldsymbol{Z}_{iT}\right) && (3.4)\\
\boldsymbol{Y}_{iT} &= \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{H}_{iT} + \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n\right) \boldsymbol{\mu}_{yi} + \boldsymbol{E}_{iT} && (3.5)\\
\boldsymbol{X}_{iT} &= \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{\Xi}_{iT} + \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k\right) \boldsymbol{\mu}_{xi} + \boldsymbol{\Delta}_{iT}, && (3.6)
\end{aligned}
$$

using the notation defined in Table 2.2.

It follows that (3.4) can be substituted into (3.5) to obtain a system of equations with observable variables on the left-hand side

$$
\begin{aligned}
\boldsymbol{Y}_{iT} &= \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{\Xi}_{iT} + \boldsymbol{Z}_{iT}\right] \\
&+ \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n\right) \boldsymbol{\mu}_{yi} + \boldsymbol{E}_{iT} && (3.7)\\
\boldsymbol{X}_{iT} &= \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{\Xi}_{iT} + \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k\right) \boldsymbol{\mu}_{xi} + \boldsymbol{\Delta}_{iT}. && (3.8)
\end{aligned}
$$

By Assumption 3.2.0.1 the unobservable variables in (3.1)–(3.3) have expectation zero, thus it is easy to see that $E\left[\boldsymbol{Y}_{iT}\right] = \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n\right) \boldsymbol{\mu}_{yi}$ and $E\left[\boldsymbol{X}_{iT}\right] = \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_g\right) \boldsymbol{\mu}_{xi}$. Therefore, the expectations of the observable variables are the individual fixed-effects so we can define

$$
\boldsymbol{V}_{iT} \equiv \begin{pmatrix} \boldsymbol{Y}_{iT} - E\left[\boldsymbol{Y}_{iT}\right] \\ \boldsymbol{X}_{iT} - E\left[\boldsymbol{X}_{iT}\right] \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{iT} - \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n\right) \boldsymbol{\mu}_{yi} \\ \boldsymbol{X}_{iT} - \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k\right) \boldsymbol{\mu}_{xi} \end{pmatrix}. \tag{3.9}
$$

Since $E\left[\boldsymbol{V}_{iT}\right] = \boldsymbol{0}$ we have $\mathrm{Var}\left(\boldsymbol{V}_{iT}\right) = E \Big) = \Big({}_{T}]$.

3.2.1 Maximum likelihood estimation of the parameters

The maximum likelihood estimation proceeds in two steps. Firstly, since we treat the vectors of fixed effects $\boldsymbol{\mu}_{yi}$ and $\boldsymbol{\mu}_{xi}$ as incidental parameters of no substantive interest, we concentrate them out of the log-likelihood. Secondly, we maximise the concentrated log-likelihood to obtain the estimates of the parameter vector $\boldsymbol{\theta}$. We will assume that sufficient restrictions (e.g. zero restrictions) are placed on the model parameters so that the model is identified. The following assumption outlines the basic regularity conditions.

Assumption 3.2.1.1 *Let $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ be a function of the parameters* $\mathrm{vec}\boldsymbol{B}_i'$, $\mathrm{vec}\boldsymbol{\Gamma}_j'$, $\mathrm{vec}\boldsymbol{\Lambda}_y$, $\mathrm{vec}\boldsymbol{\Lambda}_x$, $\mathrm{vech}\,\boldsymbol{\Phi}_j'$, $\mathrm{vech}\,\boldsymbol{\Psi}'$, $\mathrm{vech}\,\boldsymbol{\Theta}_\delta'$, *and* $\mathrm{vech}\,\boldsymbol{\Theta}_\varepsilon'$; $i = 0, \ldots, p$, $j = 0, \ldots, q$, *where $\boldsymbol{\theta}$ is an open set in the parameter space $\boldsymbol{\Upsilon}$. We assume that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is positive definite and continuous in $\boldsymbol{\theta}$ at every point in $\boldsymbol{\Upsilon}$. We also require that $\partial \boldsymbol{\Sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'$ and $\partial^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ are continuous in the neighborhood of $\boldsymbol{\theta}_0$, and that $\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}'$ has full column rank at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Finally, $\forall \varepsilon > 0, \exists \delta > 0 : ||\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)|| < \delta \Rightarrow ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \varepsilon$.*

We firstly consider estimation of the fixed effects parameters $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$. Let

$$\boldsymbol{M}_i \equiv \begin{pmatrix} \boldsymbol{\mu}_{yi} \\ \boldsymbol{\mu}_{xi} \end{pmatrix}, \quad \boldsymbol{F} \equiv \begin{pmatrix} \iota_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \iota_T \otimes \boldsymbol{I}_k \end{pmatrix}, \tag{3.10}$$

so we can write

$$E\left[\begin{pmatrix} \boldsymbol{Y}_{iT} \\ \boldsymbol{X}_{iT} \end{pmatrix}\right] = \begin{pmatrix} \iota_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \iota_T \otimes \boldsymbol{I}_k \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_{yi} \\ \boldsymbol{\mu}_{xi} \end{pmatrix} = \boldsymbol{F}\boldsymbol{M}_i.$$

Therefore, by letting $\boldsymbol{W}_{iT} \equiv (\boldsymbol{Y}'_{iT} : \boldsymbol{X}'_{iT})'$, the multivariate Gaussian likelihood of the DPSEM model for the individual i is given by

$$L(\boldsymbol{W}_{iT}, \boldsymbol{M}_i) = (2\pi)^{T/2} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{W}_{iT} - \boldsymbol{F}\boldsymbol{M}_i)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) (\boldsymbol{W}_{iT} - \boldsymbol{F}\boldsymbol{M}_i)\right),$$

and thus the log-likelihood is

$$\begin{aligned} \ln L(\boldsymbol{W}_{iT}, \boldsymbol{M}_i) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| \\ &- \frac{1}{2}(\boldsymbol{W}_{iT} - \boldsymbol{F}\boldsymbol{M}_i)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) (\boldsymbol{W}_{iT} - \boldsymbol{F}\boldsymbol{M}_i). \end{aligned} \tag{3.11}$$

The maximum likelihood estimate of $\boldsymbol{M}_i$ can be obtained by solving the first-order condition

$$\frac{\partial \ln L(\boldsymbol{W}_{iT}, \boldsymbol{M}_i)}{\partial \boldsymbol{M}_i} = \boldsymbol{F}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{W}_{iT} - \boldsymbol{F}\boldsymbol{M}_i) = \boldsymbol{0} \tag{3.12}$$

which gives the ML solution

$$\hat{\boldsymbol{M}}_i = (\boldsymbol{F}'\boldsymbol{F})^{-1} \boldsymbol{F}' \boldsymbol{W}_{iT}. \tag{3.13}$$

Substituting (3.13) into (3.11) yields the concentrated log-likelihood of the form

$$\begin{aligned} \ln L\left(\boldsymbol{W}_{iT}, \hat{\boldsymbol{M}}_i\right) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| \\ &- \frac{1}{2}\left[\left(\boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{iT}\right]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left[\left(\boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{iT}\right] \end{aligned}$$

which, by letting $\tilde{\boldsymbol{W}}_{iT} \equiv \left(\boldsymbol{I} - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{iT}$, simplifies to

$$-\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right| - \frac{1}{2}\tilde{\boldsymbol{W}}'_{iT}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\tilde{\boldsymbol{W}}_{iT}. \tag{3.14}$$

The concentrated log-likelihood (3.14) is the log-likelihood for the within-group (WG) transformed data. To see this, note that $\left(\boldsymbol{I} - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)$ is the WG transformation matrix, i.e.,

$$\left(\boldsymbol{I} - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right) = \boldsymbol{I}_{(n+k)T} - \frac{1}{T}\begin{pmatrix} \boldsymbol{\iota}_T\boldsymbol{\iota}'_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota}_T\boldsymbol{\iota}'_T \otimes \boldsymbol{I}_k \end{pmatrix}, \tag{3.15}$$

which follows from the fact that

$$\begin{aligned} \boldsymbol{F}'\boldsymbol{F} &= \begin{pmatrix} \boldsymbol{\iota}'_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota}'_T \otimes \boldsymbol{I}_k \end{pmatrix}\begin{pmatrix} \boldsymbol{\iota}_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota}_T \otimes \boldsymbol{I}_k \end{pmatrix} \\ &= \begin{pmatrix} \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n\right)'\left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_n\right) & \boldsymbol{0} \\ \boldsymbol{0} & \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k\right)'\left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_k\right) \end{pmatrix} \\ &= T\begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_k \end{pmatrix}, \end{aligned}$$

and thus $\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1} = T^{-1}\boldsymbol{I}_{(n+k)}$. Therefore,

$$\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}' = \frac{1}{T}\begin{pmatrix} \boldsymbol{\iota}_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota}_T \otimes \boldsymbol{I}_k \end{pmatrix}\begin{pmatrix} \boldsymbol{\iota}'_T \otimes \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota}'_T \otimes \boldsymbol{I}_k \end{pmatrix},$$

which yields (3.15). It now follows that the Gaussian log-likelihood for the sample of N mutually independent time series process $\boldsymbol{W}_{iT} \equiv \left(\boldsymbol{Y}'_{iT} : \boldsymbol{X}'_{iT}\right)'$ is the concentrated likelihood given by

$$\begin{aligned} \sum_{i=1}^{N}\ln L\left(\boldsymbol{W}_{iT}, \hat{\boldsymbol{M}}_i\right) &= -\frac{TN}{2}\ln(2\pi) - \frac{N}{2}\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right| - \frac{1}{2}\sum_{i=1}^{N}\tilde{\boldsymbol{W}}'_{iT}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\tilde{\boldsymbol{W}}_{iT} \\ &= -\frac{NT}{2}\ln\left(2\pi\right) - \frac{N}{2}\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT} \end{aligned} \tag{3.16}$$

where $\boldsymbol{W}_{NT} = \left(\boldsymbol{W}_{1T}, \ldots \boldsymbol{W}_{NT}\right)$ and $\tilde{\boldsymbol{W}}_{NT} \equiv \left(\boldsymbol{I} - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)\boldsymbol{W}_{NT}$ is the within-group transformed data matrix. It thus follows that the maximum likelihood estimator of $\boldsymbol{\theta}$ solves

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\theta}\left(\sum_{i=1}^{N}\ln L\left(\boldsymbol{W}_{iT}, \hat{\boldsymbol{M}}_i\right)\right), \tag{3.17}$$

Equivalently, the maximisation problem (3.17) can be turned into an equivalent minimisation problem

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\min_{\boldsymbol{\theta}} \left(-2 \sum_{i=1}^{N} \left(\boldsymbol{W}_{iT}, \hat{\boldsymbol{M}}_i \right) \right), \tag{3.18}$$

ignoring the constant term. Optimisation of (3.17) or (3.18) requires numerical methods such as the method of scoring or the Newton-Raphson algorithm. We will derive the closed form expressions for the analytical first and second derivatives in §3.2, which facilitates both methods. As we will show, the expectation of the Hessian matrix (or its probability limit) turns out to be notably simpler then the Hessian itself. Therefore, the method of scoring, which requires only the expectation of the Hessian matrix, is simpler to implement. The parameters' estimates can hence be obtained by iterating

$$\hat{\boldsymbol{\theta}}_f = \hat{\boldsymbol{\theta}}_{f-1} + \Im^{-1}(\boldsymbol{\theta}_{f-1}) \left. \frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_{f-1}}, \tag{3.19}$$

which can be implemented by using the closed form analytical expressions for the score vector and the information matrix provided in §3.2 and §3.3. The method of scoring generally requires good starting values, which can be provided using the IV methods suggested by ?.

At this point construction of the empirical covariance matrix merits few remarks. The $1/N$ times $\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}$ is the empirical covariance matrix of the within-group transformed data on N individual time series vectors $\tilde{\boldsymbol{W}}_{iT}$. To show this, we point out that the within-group transformed data for the individual i for T time periods can be stacked into the $(n+k)T \times 1$ vector

$$\tilde{\boldsymbol{W}}_{iT} = \begin{pmatrix} \tilde{\boldsymbol{Y}}_{iT} \\ \tilde{\boldsymbol{X}}_{iT} \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{y}}_{i1} \\ \vdots \\ \tilde{\boldsymbol{y}}_{iT} \\ \tilde{\boldsymbol{x}}_{i1} \\ \vdots \\ \tilde{\boldsymbol{x}}_{iT} \end{pmatrix}, \tag{3.20}$$

where

$$
\tilde{\boldsymbol{Y}}_i = \begin{pmatrix} y_{i1}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} y_{ij}^{(1)} \\ \vdots \\ y_{i1}^{(n)} - \frac{1}{T}\sum_{j=1}^{T} y_{ij}^{(l)} \\ \vdots \\ y_{iT}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} y_{iT}^{(1)} \\ \vdots \\ y_{iT}^{(n)} - \frac{1}{T}\sum_{j=1}^{T} y_{iT}^{(l)} \end{pmatrix}, \quad \text{and} \quad \tilde{\boldsymbol{X}}_i = \begin{pmatrix} x_{i1}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(1)} \\ \vdots \\ x_{i1}^{(k)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(k)} \\ \vdots \\ x_{iT}^{(1)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(1)} \\ \vdots \\ x_{iT}^{(k)} - \frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(k)} \end{pmatrix} \tag{3.21}
$$

are $nT \times 1$ and $kT \times 1$ vectors, respectively. We now define an $(n+k)T \times N$ matrix whose columns are data vectors on N individuals as

$$
\tilde{\boldsymbol{W}}_{NT} \equiv \begin{pmatrix} \tilde{\boldsymbol{Y}}_1 & \tilde{\boldsymbol{Y}}_2 & \cdots & \tilde{\boldsymbol{Y}}_N \\ \tilde{\boldsymbol{X}}_1 & \tilde{\boldsymbol{X}}_2 & \cdots & \tilde{\boldsymbol{X}}_N \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{y}}_{11} & \tilde{\boldsymbol{y}}_{21} & \cdots & \tilde{\boldsymbol{y}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{y}}_{1T} & \tilde{\boldsymbol{y}}_{2T} & \cdots & \tilde{\boldsymbol{y}}_{NT} \\ \tilde{\boldsymbol{x}}_{11} & \tilde{\boldsymbol{x}}_{21} & \cdots & \tilde{\boldsymbol{x}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{x}}_{1T} & \tilde{\boldsymbol{x}}_{2T} & \cdots & \tilde{\boldsymbol{x}}_{NT} \end{pmatrix} \tag{3.22}
$$

hence $\tilde{\boldsymbol{W}}_{NT}$ is the empirical data matrix for the entire sample (panel) of N individuals observed over T time periods. The $(n+k)NT \times (n+k)NT$ empirical covariance matrix can be computed by noting that

$$
\begin{aligned}
\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT} &= \begin{pmatrix} \tilde{\boldsymbol{y}}_{11} & \tilde{\boldsymbol{y}}_{21} & \cdots & \tilde{\boldsymbol{y}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{y}}_{1T} & \tilde{\boldsymbol{y}}_{2T} & \cdots & \tilde{\boldsymbol{y}}_{NT} \\ \tilde{\boldsymbol{x}}_{11} & \tilde{\boldsymbol{x}}_{21} & \cdots & \tilde{\boldsymbol{x}}_{N1} \\ \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{x}}_{1T} & \tilde{\boldsymbol{x}}_{2T} & \cdots & \tilde{\boldsymbol{x}}_{NT} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{y}}'_{11} & \cdots & \tilde{\boldsymbol{y}}'_{1T} & \tilde{\boldsymbol{x}}'_{11} & \cdots & \tilde{\boldsymbol{x}}'_{1T} \\ \tilde{\boldsymbol{y}}'_{21} & \cdots & \tilde{\boldsymbol{y}}'_{2T} & \tilde{\boldsymbol{x}}'_{21} & \cdots & \tilde{\boldsymbol{x}}'_{2T} \\ \vdots & & \vdots & \vdots & & \vdots \\ \tilde{\boldsymbol{y}}'_{N1} & \cdots & \tilde{\boldsymbol{y}}'_{NT} & \tilde{\boldsymbol{x}}'_{N1} & \cdots & \tilde{\boldsymbol{x}}'_{NT} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1}\tilde{\boldsymbol{x}}'_{iT} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{i1T}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{iT}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{iT}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{y}}_{iT}\tilde{\boldsymbol{x}}'_{iT} \\ \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i1}\tilde{\boldsymbol{x}}'_{iT} \\ \vdots & & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{y}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{y}}'_{iT} & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{x}}'_{i1} & \cdots & \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{iT}\tilde{\boldsymbol{x}}'_{iT} \end{pmatrix}
\end{aligned}
$$

which can be written more concisely as

$$\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT} = \begin{pmatrix} \sum_{i=1}^{N} \tilde{\boldsymbol{Y}}_i \tilde{\boldsymbol{Y}}'_i & \sum_{i=1}^{N} \tilde{\boldsymbol{Y}}_i \tilde{\boldsymbol{X}}'_i \\ \sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{Y}}'_i & \sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}'_i \end{pmatrix} \tag{3.23}$$

Letting $\bar{y}_i^{(*)} \equiv T^{-1}\sum_{j=1}^{T} y_{ij}^{(*)}$ and $\bar{x}_i^{(*)} \equiv T^{-1}\sum_{j=1}^{T} x_{ij}^{(*)}$ it follows that the typical elements of $\sum_{i=1}^{N} \tilde{\boldsymbol{Y}}_i \tilde{\boldsymbol{Y}}'_i$, $\sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{Y}}'_i$, and $\sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}'_i$ are of the form

$$\sum_{i=1}^{N} \boldsymbol{y}_{ij}\boldsymbol{y}'_{if} = \begin{pmatrix} \sum_{i=1}^{N}\left(y_{ij}^{(1)} - \bar{y}_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(y_{i1}^{(1)} - \bar{y}_i^{(1)}\right)\left(y_{i1}^{(l)} - \bar{y}_i^{(l)}\right) \\ \vdots & & \vdots \\ \sum_{i=1}^{N}\left(y_{ij}^{(l)} - \bar{y}_i^{(l)}\right)\left(y_{if}^{(1)} - \bar{y}_i^{(1)}\right) & \cdots & \sum_{i=1}^{N}\left(y_{i1}^{(l)} - \bar{y}_i^{(l)}\right)^2 \end{pmatrix},$$

$$\sum_{i=1}^{N} \boldsymbol{x}_{ij}\boldsymbol{y}'_{if} = \begin{pmatrix} \sum_{i=1}^{N}\left(x_{ij}^{(1)} - \bar{x}_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(1)} - \bar{x}_i^{(1)}\right)\left(x_{i1}^{(k)} - \bar{x}_i^{(k)}\right) \\ \vdots & & \vdots \\ \sum_{i=1}^{N}\left(x_{ij}^{(k)} - \bar{x}_i^{(k)}\right)\left(x_{if}^{(1)} - \bar{x}_i^{(1)}\right) & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(k)} - \bar{x}_i^{(k)}\right)^2 \end{pmatrix},$$

and

$$\sum_{i=1}^{N} \boldsymbol{x}_{ij}\boldsymbol{x}'_{if} = \begin{pmatrix} \sum_{i=1}^{N}\left(x_{ij}^{(1)} - \bar{x}_i^{(1)}\right)^2 & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(1)} - \bar{x}_i^{(1)}\right)\left(x_{i1}^{(k)} - \bar{x}_i^{(k)}\right) \\ \vdots & & \vdots \\ \sum_{i=1}^{N}\left(x_{ij}^{(k)} - \bar{x}_i^{(k)}\right)\left(x_{if}^{(1)} - \bar{x}_i^{(k)}\right) & \cdots & \sum_{i=1}^{N}\left(x_{i1}^{(k)} - \bar{x}_i^{(k)}\right)^2 \end{pmatrix},$$

respectively. By assumption (2.2.0.2) the time means converge in probability to the population individual means

$$\underset{T\to\infty}{p\lim}\left(\frac{1}{T}\sum_{j=1}^{T} y_{ij}^{(k)}\right) = \mu_{yi}^{(k)} \quad \text{and} \quad \underset{T\to\infty}{p\lim}\left(\frac{1}{T}\sum_{j=1}^{T} x_{ij}^{(k)}\right) = \mu_{xi}^{(k)}$$

which implies that

$$\underset{T\to\infty}{p\lim}\, \tilde{\boldsymbol{W}}_i = \boldsymbol{W}_i - \boldsymbol{M}_i. \tag{3.24}$$

Therefore, the covariances of the within-group transformed data converge in probability limit to

$$\underset{T\to\infty}{p\lim}\sum_{i=1}^{N}\left(y_{is}^{(1)} - \bar{y}_i^{(1)}\right)\left(y_{is}^{(k)} - \bar{y}_i^{(k)}\right) = \sum_{i=1}^{N}\left(y_{is}^{(1)} - \mu_{yi}^{(l)}\right)\left(y_{is}^{(k)} - \mu_{yi}^{(k)}\right)$$

Hence, the within group estimator requires that $T \to \infty$. Sequentially, if we let $N \to \infty$, we obtain the convergence in probability of the the empirical covariance matrix as

$$\underset{T,N\to\infty}{p\lim} \frac{1}{N} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0). \tag{3.25}$$

3.2.2 Analytical derivatives and the score vector

We derive the closed form analytical expressions for the first and second derivatives of the DPSEM model, thus enabling the construction of the score vector and the information matrix.

Derivation of the analytical derivatives and components of the information matrix is a difficult problem for complex multivariate models, nevertheless, the modern matrix calculus methods (e.g. Magnus and Neudecker (1988), Turkington (2002)) make possible to obtain these results. However, detailed derivations of the score vector and the information matrix for multivariate models is not frequently undertaken and the theoretical literature is rather scarce in this area. Turkington (1998), for example, derives the score vector and the information matrix in the closed analytical form for the simultaneous equation model with vector autoregressive errors, which is so far the most complex linear model for which full analytical results were obtained.

While the main motivation behind the studies such as Turkington (1998) was to obtain the basic analytical results needed for the classical statistical inference and derivation of the Cramer-Rao lower bound, which can in turn be used for benchmarking the efficiency of various estimators, the motivation here is additionally in providing analytical inputs for implementation of efficient estimation algorithms. The computational efficiency is a major issue with complex multivariate models, specially dynamic models with unobservable variables, hence the availability of the analytical results might greatly facilitate practical implementation of the various special cases of the general model considered here. In particular, we give fully vectorised expressions for analytical derivatives and the Hessian matrix, which can be easily programmed in modern object oriented languages such as S or R.

The maximum likelihood estimator (3.17) can be interpreted as a covariance estimator, where all the unknown parameters are contained in the model-implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. To obtain the closed-form analytical derivatives of the log-likelihood (3.16) it is necessary to obtain the derivatives of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in respect to particular elements of the parameter vector $\boldsymbol{\theta}$ given in (2.3.0.6). We achieve this by firstly expressing the $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ as a linear function of its block elements $\boldsymbol{\Sigma}_{ij}$, and then trivially by expressing its derivatives as linear functions of the derivatives of the $\boldsymbol{\Sigma}_{ij}$

blocks.

Lemma 3.2.2.1 *Let $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ have the partition into $(n+k)T$ columns as*

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_1^{(11)} & \cdots & \boldsymbol{m}_{nT}^{(11)} & \boldsymbol{m}_1^{(12)} & \cdots & \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{m}_1^{(21)} & \cdots & \boldsymbol{m}_{nT}^{(21)} & \boldsymbol{m}_1^{(22)} & \cdots & \boldsymbol{m}_{kT}^{(22)} \end{pmatrix}, \tag{3.26}$$

thus each block is partitioned into columns as $\boldsymbol{\Sigma}_{ij} = \left(\boldsymbol{m}_1^{(ij)} \cdots \boldsymbol{m}_{nT}^{(ij)}\right)$, so that $\operatorname{vec} \boldsymbol{\Sigma}_{ij} = \left(\boldsymbol{m}'^{(ij)}_1, \cdots, \boldsymbol{m}'^{(ij)}_{nT}\right)'$. *Then* $\operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})$ *can be expressed as a linear combination of its vectorised columns as*

$$\operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{H}_{11} \operatorname{vec} \boldsymbol{\Sigma}_{11} + \boldsymbol{H}_{21} \operatorname{vec} \boldsymbol{\Sigma}_{21} + \boldsymbol{H}_{12} \operatorname{vec} \boldsymbol{\Sigma}_{12} + \boldsymbol{H}_{22} \operatorname{vec} \boldsymbol{\Sigma}_{22}, \tag{3.27}$$

where the $T^2(n+k)^2 \times nT$ zero-one matrices $\boldsymbol{H}_{i1}$, and the $T^2(n+k)^2 \times nkT$ zero-one matrices $\boldsymbol{H}_{i2}$, $i = 1, 2$ are specified as

$$\boldsymbol{H}_{11} \equiv \begin{pmatrix} \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} a \end{array} \qquad \boldsymbol{H}_{21} \equiv \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} b \end{array}$$

and

$$
\boldsymbol{H}_{12} \equiv \begin{pmatrix}
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0}
\end{pmatrix}
\begin{matrix}
\left.\begin{matrix} \\ \\ \end{matrix}\right\} b \\
 \\ \\ \\ \\ \\ \\ \\ \\
\end{matrix}
\qquad
\boldsymbol{H}_{22} \equiv \begin{pmatrix}
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0} \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT}
\end{pmatrix}
\begin{matrix}
\left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} c \\
 \\ \\ \\ \\ \\ \\ \\
\end{matrix}
$$

where $a = T^2k(n+k) - kT$, $b = T^2k(n+k)$, *and* $c = T^2k(n+k) - nT$.

Proof See Appendix §3 A.

Corollary 3.2.2.2 *The first derivative of the vec of a* 2×2 *block matrix* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *is a linear function of the derivatives of its vectorised block elements of the form*

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial \boldsymbol{H}_{11} \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{H}_{21} \operatorname{vec} \boldsymbol{\Sigma}_{21}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{H}_{12} \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{H}_{22} \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \boldsymbol{\theta}} \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{H}'_{ij}. \qquad (3.28)
\end{aligned}
$$

Proof By the chain rule for matrix calculus (see Magnus and Neudecker (1988, pg. 96) and Turkington (2002, pg. 71)) we have

$$
\frac{\partial \boldsymbol{H}_{ij} \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \boldsymbol{H}_{ij} \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}} \right) = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{H}'_{ij}.
$$

Therefore,

$$
\sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{\partial \boldsymbol{H}_{ij} \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) = \sum_{i=1}^{2} \sum_{j=1}^{2} \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{H}'_{ij},
$$

as required.

Q.E.D.

The following proposition gives the general expression for the analytical derivatives of the log-likelihood, $\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)/\partial \boldsymbol{\theta}$.

Proposition 3.2.2.3 *The score vector* $\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)/\partial \boldsymbol{\theta}$ *of the log likelihood (3.16) has the jth component of the form*

$$\frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]. \tag{3.29}$$

Proof See Appendix §3 B.

To obtain analytical expressions for the partial derivatives $\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}_j^{(*)}$ in respect to particular elements $\boldsymbol{\theta}_j^{(*)}$ of the parameter vector $\boldsymbol{\theta}$, we firstly introduce some new notation. We will make use of two special types of zero-one matrices, $\boldsymbol{K}_{ab}$ and $\boldsymbol{D}_a$. We define the commutation matrix $\boldsymbol{K}_{ab}$ as an orthogonal $ab \times ab$ zero-one permutation matrix

$$\boldsymbol{K}_{ab} \equiv \left(\boldsymbol{I}_a \otimes \boldsymbol{e}_1^b : \boldsymbol{I}_a \otimes \boldsymbol{e}_2^b : \cdots : \boldsymbol{I}_a \otimes \boldsymbol{e}_b^b\right) \tag{3.30}$$

such that $\boldsymbol{K}_{ab} \operatorname{vec} \boldsymbol{X} = \operatorname{vec} \boldsymbol{X}'$, where $\boldsymbol{e}_j^b$ is the jth column of a $b \times b$ identity matrix, i.e., $\boldsymbol{I}_b = \left(\boldsymbol{e}_1^b : \boldsymbol{e}_2^b : \cdots : \boldsymbol{e}_b^b\right)$. Additionally, let

$$\boldsymbol{K}_{ab}^* \equiv \mathrm{devec_b} \boldsymbol{K}_{ab} = \left[\boldsymbol{I}_b \otimes (\boldsymbol{e}_1^a)' : \boldsymbol{I}_b \otimes (\boldsymbol{e}_2^a)' : \cdots : \boldsymbol{I}_b \otimes (\boldsymbol{e}_a^a)'\right]. \tag{3.31}$$

The $a^2 \times a(a+1)/2$ duplication matrix $\boldsymbol{D}_a$ is defined as a zero-one matrix such that for an $a \times a$ matrix $\boldsymbol{X}$, $\boldsymbol{D}_a \operatorname{vech} \boldsymbol{X} = \operatorname{vec} \boldsymbol{X}$. To further simplify the exposition, we define some abbreviating notation as follows.

$$\begin{aligned}
\boldsymbol{X} &\equiv \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \\
&\times \left(\left[\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S}_T'^j \otimes \boldsymbol{\Phi}_j'\right)\right] + (\boldsymbol{I}_T \otimes \boldsymbol{\Psi})\right) \\
&\times \left(\sum_{j=0}^{q} \boldsymbol{S}_T'^j \otimes \boldsymbol{\Gamma}_j'\right)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T'^j \otimes \boldsymbol{B}_j'\right)^{-1}, \\
\boldsymbol{Y} &\equiv \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right)\left(\left[\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S}_T'^j \otimes \boldsymbol{\Phi}_j'\right)\right] + (\boldsymbol{I}_T \otimes \boldsymbol{\Psi})\right)\left(\sum_{j=0}^{q} \boldsymbol{S}_T'^j \otimes \boldsymbol{\Gamma}_j'\right),
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{A} &\equiv (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}, \\
\boldsymbol{F} &\equiv \left[\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S}'^j_T \otimes \boldsymbol{\Phi}'_j\right)\right], \\
\boldsymbol{Z} &\equiv (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right), \\
\boldsymbol{D} &\equiv (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right), \\
\boldsymbol{Q} &\equiv \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right), \\
\boldsymbol{F} &\equiv \left[\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S}'^j_T \otimes \boldsymbol{\Phi}'_j\right)\right].
\end{aligned}
$$

Proposition 3.2.2.4 *The the partial derivatives of $\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}\,(\boldsymbol{\theta})/\partial\,\boldsymbol{\theta}_j^{(*)}$ in respect to the elements of the parameter vector $\boldsymbol{\theta}$ are of the form*

$$
\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}_{ij}}{\partial\,\boldsymbol{\theta}_j^{(*)}} \mathbf{H}_{ij},
$$

where the analytical expressions for the matrices $\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}_{ij}/\partial\,\boldsymbol{\theta}_j^{(*)}$ are as follows. The derivatives of the block elements of $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, and $\boldsymbol{\Sigma}_{22}$ in respect to $\boldsymbol{\theta}^{(B_i)}$ for any $i = 0, \ldots, p$ are[1]

$$
\begin{aligned}
\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}_{11}}{\partial\,\mathrm{vec}\,\boldsymbol{B}_i} &= \left[\boldsymbol{K}^*_{T,m}\left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S}'^i_T\right) \otimes \boldsymbol{I}_m\right]\left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}'^j_T \otimes \boldsymbol{B}'_j\right)^{-1}\right] \\
&\times \left(\left[\boldsymbol{Y}\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}'^j_T \otimes \boldsymbol{B}'_j\right)^{-1} \otimes \boldsymbol{I}_{mT}\right]\right. \\
&+ \left.\left[\boldsymbol{Y}'\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}'^j_T \otimes \boldsymbol{B}'_j\right)^{-1} \otimes \boldsymbol{I}_{mT}\right]\boldsymbol{K}_{mT,mT}\right) \times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)
\end{aligned}
$$

[1]Since $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$ we do not need to give a separate expression for $\boldsymbol{\Sigma}_{21}$.

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \left[\boldsymbol{K}_{T,m}^{*}\left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S}'^{i}_{T}\right) \otimes \boldsymbol{I}_m\right] \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}^{j}_{T} \otimes \boldsymbol{B}_j\right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}'^{j}_{T} \otimes \boldsymbol{B}'_j\right)^{-1}\right] \\
&\times \left(\left[\left(\sum_{j=0}^{q} \boldsymbol{S}^{j}_{T} \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right)\right] \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \boldsymbol{0}.
\end{aligned}
$$

In respect to $\boldsymbol{\theta}^{(\Gamma_i)}$, for any $i = 0, \ldots, q$, the derivatives of the individual blocks are

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \left[\boldsymbol{K}_{T,g}^{*}\left(\boldsymbol{I}_{Tg} \otimes \boldsymbol{S}'^{i}_{T}\right) \otimes \boldsymbol{I}_m\right] \left[\boldsymbol{Y}\left(\boldsymbol{S}^{i}_{T} \otimes \boldsymbol{\Gamma}_i\right)' \otimes \boldsymbol{I}_{mT} + \boldsymbol{Y}'\left(\boldsymbol{S}^{i}_{T} \otimes \boldsymbol{\Gamma}_i\right)' \otimes \boldsymbol{I}_{mT})\right] \\
&\times \boldsymbol{K}_{mT,mT}\left(\boldsymbol{A}' \otimes \boldsymbol{A}'\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \left[\boldsymbol{K}_{T,g}^{*}\left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S}'^{i}_{T}\right) \otimes \boldsymbol{I}_m\right] \\
&\times \left(\left[\boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right)\right] \otimes \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}'^{j}_{T} \otimes \boldsymbol{B}'_j\right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)\right]\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \boldsymbol{0}.
\end{aligned}
$$

In respect to $\boldsymbol{\theta}^{(\Lambda_y)}$, the derivatives are

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \left(\boldsymbol{K}_{T,m}^{*} \otimes \boldsymbol{I}_n\right)\left(\left[\boldsymbol{X}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) \otimes \boldsymbol{I}_{nT}\right] + \left[\boldsymbol{X}'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right) \otimes \boldsymbol{I}_{nT}\right] \boldsymbol{K}_{nT,nT}\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right]\left(\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T\right)\left(\left[\boldsymbol{Q}\boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right)\right] \otimes \boldsymbol{I}_{nT}\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \boldsymbol{0}.
\end{aligned}
$$

In respect to $\boldsymbol{\theta}^{(\Lambda_x)}$, the derivatives are

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right]\left(\boldsymbol{K}_{k,T} \otimes \boldsymbol{I}_T\right) \boldsymbol{K}_{k,T}\left(\boldsymbol{I}_{gT} \otimes \boldsymbol{F}\boldsymbol{Q}'\left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_y\right)\right]\right) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \left(\boldsymbol{K}_{T,g}^{*} \otimes \boldsymbol{I}_k\right)\left(\left[\boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right) \otimes \boldsymbol{I}_{kT}\right] + \left[\boldsymbol{F}'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right) \otimes \boldsymbol{I}_{kT}\right] \boldsymbol{K}_{k,T}\right).
\end{aligned}
$$

The contemporaneous covariance matrix $\boldsymbol{\Phi}_0$ of the exogenous latent variables appears on the diagonal of the block Toeplitz matrix (8.12), while for any other $j \neq 0$, both $\boldsymbol{\Phi}_j$ and $\boldsymbol{\Phi}'_j$ appear off-diagonally. Hence we differentiate each $\boldsymbol{\Sigma}_{ij}$ separately for $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Phi}_j$ $(j \neq 0)$ in respect to $\boldsymbol{\theta}^{(\Phi)}$, which yields

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T) (\boldsymbol{Z}' \otimes \boldsymbol{Z}') \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^{*} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S}'^{i}_{T}\right) \otimes \boldsymbol{I}_g\right] (\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}) (\boldsymbol{Z}' \otimes \boldsymbol{Z}') \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Phi}_0} &= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes \boldsymbol{Q}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right] \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^{*} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S}'^{i}_{T}\right) \otimes \boldsymbol{I}_g\right] (\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes \boldsymbol{Q}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right] \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Phi}_0} &= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x')\right] \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^{*} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S}'^{i}_{T}\right) \otimes \boldsymbol{I}_g\right] (\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}) \left[(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x') \otimes (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x')\right].
\end{aligned}
$$

Finally, the derivatives in respect to the error covariance matrices are as follows. For $\boldsymbol{\theta}^{(\Psi)}$ we have

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Psi}} &= \boldsymbol{D}_m' \left[\boldsymbol{I}_m \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{m,T} \otimes \boldsymbol{I}_T) (\boldsymbol{D}' \otimes \boldsymbol{D}') \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Psi}} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Psi}} &= \boldsymbol{0}.
\end{aligned}
$$

For $\boldsymbol{\theta}^{(\Theta_\varepsilon)}$ we have

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} &= \boldsymbol{D}_n' \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T) \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} &= \boldsymbol{0},
\end{aligned}
$$

and for $\boldsymbol{\theta}^{(\Theta_\delta)}$,

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} &= \boldsymbol{0} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} &= \boldsymbol{D}_k' \left(\boldsymbol{I}_k \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right) (\boldsymbol{K}_{kT} \otimes \boldsymbol{I}_T).
\end{aligned}
$$

Proof See Appendix §3 C.

The score vector can now be constructed by substituting the partial derivatives given in proposition 3.2.2.4 into the general expression for the components of the score vector given by the expression (3.29).

3.2.3 Asymptotic inference

The basic inferential properties of the multivariate Gaussian models whose likelihood can be written by separating the unknown parameters from the observable variables, e.g. the likelihood of the DPSEM model (3.16), are asymptotically equivalent to the properties of the Wishart estimators analysed by Anderson and Amemiya (1988), Anderson (1989), and Amemiya and Anderson (1990). In addition to these known results, we give the analytical expressions in the closed form of the Hessian and information matrices.

We make the standard assumption that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is twice continuously differentiable in a neighborhood of $\boldsymbol{\theta}_0$, and that $\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_j^{(*)}$ has full column rank at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proposition 3.2.3.1 *Let $\boldsymbol{\theta}^{(*)}$ denote any component of the parameter vector $\boldsymbol{\theta}$, as defined in (2.3.0.6). Then the Hessian matrix is of the form*

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{\partial \ln L \;\; \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(B_0)} \partial \boldsymbol{\theta}'^{(B_0)}} & \cdots & \dfrac{\partial \ln L \;\; \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(B_0)} \partial \boldsymbol{\theta}'^{(\Theta_\delta)}} \\ \vdots & & \vdots \\ \dfrac{\partial \ln L \;\; \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(\Theta_\delta)} \partial \boldsymbol{\theta}'^{(B_0)}} & \cdots & \dfrac{\partial \ln L \;\; \tilde{\boldsymbol{W}}_{NT}}{\partial \boldsymbol{\theta}^{(\Theta_\delta)} \partial \boldsymbol{\theta}'^{(\Theta_\delta)}} \end{pmatrix} \tag{3.32}$$

where the typical element is given by

$$\begin{aligned} \frac{\partial^2 \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}_i^{(*)\prime}} &= \frac{1}{2}\left(\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}'^{(*)}_i}\right)\left(\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \otimes \boldsymbol{I}_{p_i}\right) \\ &- \frac{1}{2}\left[\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right. \\ &\times \left(\left[\tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{mT}\right] - \left[\boldsymbol{I}_{mT} \otimes \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right) \\ &- \left. N\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right]\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)'. \end{aligned} \tag{3.33}$$

Proof See Appendix §3 D.

Proposition 3.2.3.2 *The information matrix is of the form $\Im(\boldsymbol{\theta}_0) = -\boldsymbol{H}(\boldsymbol{\theta}_0)$ with typical block elements given by*

$$\operatorname*{plim}_{T,N\to\infty} \left.\frac{\partial^2 \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)} \partial \boldsymbol{\theta}'^{(*)}_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\right]\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_j^{(*)}}\right)', \tag{3.34}$$

where $\boldsymbol{\theta}_0$ is the population value of $\boldsymbol{\theta}$.

Proof See Appendix §3 E.

The information matrix (3.34) can be constructed by using the analytical expressions given in the proposition 3.2.2.4 for the partial derivatives of the log-likelihood in respect to the particular elements of the parameter vector $\boldsymbol{\theta}$. Note that the asymptotics in the temporal dimension (i.e., $T \rightarrow \infty$) are required only for the consistent estimation of the time-means (fixed effects).

The asymptotic normality of the maximum likelihood estimator of $\boldsymbol{\theta}$ can be established in the standard way by using the Taylor series expansion of the log-likelihood

$$\left.\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}} = \left.\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \left.\frac{\partial^2 \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\right) = \boldsymbol{0},$$

which implies

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0 &= \frac{1}{2}\left(\frac{\partial^2 \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}\right)^{-1} \frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_0} \\ &= \frac{1}{2}\boldsymbol{H}^{-1}\left(\boldsymbol{\theta}_0\right)\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\theta}_0}\right) \\ &\times \left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right) \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right)\right] + o_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \tag{3.35}$$

From (8.45) now have that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left[\boldsymbol{0}, 2\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\right]. \tag{3.36}$$

Subsequently, hypotheses of the goodness of fit of the form $H_0 : E\left[\tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}_{NT}'\right] = \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)$ can be tested using the statistic $T = N \ln L \tilde{\boldsymbol{W}}_{NT}\left(\hat{\boldsymbol{\theta}}_{ML}\right)$, which is asymptotically χ^2 distributed with degrees of freedom d (for the proof see Anderson (1989)'s, theorem 2.3; see also Browne (1984)). The degrees of freedom parameter d is the difference between the number of distinct elements in the data covariance matrix $(1/N) \tilde{\boldsymbol{W}} \tilde{\boldsymbol{W}}'$ and the number of elements in $\boldsymbol{\theta}$, i.e., the number of parameters to be estimated. This χ^2-distributed fit statistic can be used for testing the null hypothesis corresponding to a particular model-implied covariance structure against the alternative of a completely unconstrained covariance matrix.

In practice, the reliance on this statistic must be taken with caution as it is known to be sensitive to departures from normality. While here we have assumed normality, Amemiya and Anderson (1990) have shown that this statistic will be still asymptotically valid for the non-normal data as well as for certain classes of dependent data, though the model they considered is somewhat less general then the one we are analysing in here.[2]

3.3 Conclusion

We considered maximum likelihood estimation of dynamic panel structural equation models with latent variables and fixed effects (DPSEM). The methods considered in this chapter derive from the structural equation modelling tradition where latent variables are measured by multiple observable indicators and where the structural equations are estimated jointly with the measurement model. Here, these methods are generalised to dynamic panel models with fixed effects. The DPSEM model encompasses virtually any dynamic or static linear model, and it can be trivially shown that classical dynamic simultaneous equation models, vector autoregressive moving average models, seemingly unrelated regression models with autoregressive disturbances, as well as factor analysis models and static structural equation models can all be specified by imposing zero restrictions on the parameter matrices of the general DPSEM model.

We derived analytical expressions for the covariance structure of the DPSEM model as well as the score vector and the Hessian matrix, in a closed form, and suggested a scoring method approach to the estimation of the unknown parameters. The closed form covariance structure allowed us to write the likelihood function of the DPSEM model by separating the observable covariance matrix from the model-implied covariance matrix in the likelihood function, which enabled application of the existing asymptotic results for the general class of Wishart estimators.

Further research should consider small-sample properties of these estimators as well as their properties when the observable variables are not normally distributed. Another extension of the present research framework would be to obtain an analytical expression for the Cramer-Rao lower bound, which would provide a general lower bound for virtually any linear model and thus enable benchmarking of asymptotic efficiency of alternative estimators. This would require analytical inversion of a the information matrix derived here.

[2]The asymptotic results of Amemiya and Anderson (1990) strictly apply to models without the stochastic error term in the structural equation; the extension of these results to the non-zero error case is not straightforward and it requires a more general framework.

Chapter 4

Maximum likelihood estimation with pure time series data

4.1 Introduction

In this chapter we consider estimation of DSEM models with pure time series data, i.e., the case with $N = 1$. There appears to be a growing interest in SEM models for such data in the literature (Oud 2001, Oud 2002, Hamaker et al. 2002, Hamaker et al. 2003, Oud 2004). With pure time series data it is not possible to compute a sufficient statistic in the form of an empirical estimate of the covariance matrix, $\boldsymbol{S}$, which was needed for the methods considered in chapter §3. Namely, with $N = 1$ manipulation of the log-likelihood (3.16) that allowed us to replace the raw data with the empirical covariance matrix is not possible. Thus in such case we need to consider estimation using raw data. To see this, note that the log-likelihood for a sample of N independent Gaussian observations $\mathbf{z}_i \sim i.i.d.N(\mu, \boldsymbol{\Sigma})$,

$$L(\boldsymbol{\theta}) = (2\pi)^{-\frac{N(l+k)}{2}} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-\frac{N}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} \mathbf{z}'_i \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{z}_i\right) \tag{4.1}$$

can be rewritten as

$$\ln L(\boldsymbol{\theta}) = \frac{-N(l+k)}{2}\ln(2\pi) - \frac{N}{2}\left[\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}\left(\frac{N-1}{N}\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)\right], \tag{4.2}$$

by noting that we rewrite the term in the trace as

$$\begin{aligned}
\frac{1}{2}\sum_{i=1}^{N} \mathbf{z}_i' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{z}_i &= -\frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\left(\mathbf{z}_i'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{z}_i\right) \\
&= \frac{N}{2}\sum_{i=1}^{N} \mathrm{tr}\left(N^{-1}\mathbf{z}_i\mathbf{z}_i'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right) \\
&= \frac{N}{2}\mathrm{tr}\left(\frac{N-1}{N}\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right),
\end{aligned} \tag{4.3}$$

where S is the empirical covariance matrix that can be calculated separately and used as a "sufficient statistic" in the maximisation/minimisation of the (log)likelihood function. Jöreskog (1981) and Anderson (1989) pointed out that maximising (4.2) is equivalent to minimising the Wishart log-likelihood $\ln L(\boldsymbol{\theta})^* = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}\ \mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \ln|\mathbf{S}| - l - k$, therefore, the use of S in the numerical optimisation of the likelihood function has firm theoretical grounds and leads to the "standard" SEM estimation approach based on covariance structure analysis.[1]

This problem lead some authors to consider computing an artificial covariance matrix where lagged values of the time series vectors are treated as additional variables. This allows computation of a $k \times k$ matrix of variances and lagged covariances, where k denotes the number of observable variables in the model including lagged ones (Hershberger et al. 1994, Hershberger et al. 1994, Molenaar 1985, Molenaar 1999, Van Buuren 1997). This approach, however, does not yield maximum likelihood estimates, rather it results in the method of moments estimates.

In chapter §2 we showed that the general DSEM model can be written in the state space form. Hence, we might consider using the Kalman filter to recursively evaluate the likelihood of the model, where at each step new parameters estimates are obtained until convergence Terceiro Lomba (1990).

This approach was shown to be feasible for simple univariate dynamic models (Engle and Watson 1981, Engle et al. 1985, Ghosh 1989).

Engle et al. (1985) and Ghosh (1989) estimated similar univariate models using an expectation-maximisation (EM) algorithm to update the likelihood, which was calculated using Kalman-filtered estimates of the unobservable state variables.

However, these methods have not been widely applied to estimation of multivariate dynamic models such as DSEM, although some simple special cases of the general DSEM model (i.e. univariate models) can estimated in the state-space form using the `SsfPack` in Ox (Koopman et al. 1999) or in the S+FinMetrics module of S-Plus (Zivot et al. 2002).[2] Maximum likelihood estimation of the unknown parameters for certain univariate models is straightforward and `SsfPack` provides built-in functions for placing these models in the state-space form. Kalman filtering and smoothing can be applied efficiently for such models, however, estimation is difficult if gradient and good starting values are not provided.

One reason is the difficulty in providing analytical derivatives for such models, without which minimisation using e.g. quasi-Newton methods in combination with

[1]Note that by ignoring the constant term $(-N(l+k)/2)\ln(2\pi)$ in (4.2), and noting that since $\lim_{N\to\infty} \frac{1}{N-1}\sum_{i=1}^{N} \mathbf{z}_i\mathbf{z}_i' = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i\mathbf{z}_i'$, the unbiased covariance matrix $\frac{N-1}{N}\mathbf{S}$ can be replaced by the asymptotically equivalent $\mathbf{S}$. Multiplying (4.2) throughout by $-2/N$ and replacing the constant term by $-\ln|\mathbf{S}| - l - k$ yields the Wishart log-likelihood.

[2]The `SsfPack` is a suit of routines for state space modelling, which implements the Kalman filter and smoother.

the Kalman filter is likely to fail in multivariate models.[3] Another reason is the recursive nature of such estimation, which contributes to slow convergence problems with large models due to the necessity of using the Kalman filter recursions at each evaluation step.

An alternative to fitting the theoretical covariance structure to an empirical covariance matrix, or to using recursive estimation, is the raw-data maximum likelihood (RD-ML), where the likelihood function is written for the entire sample. Such an approach was considered by Neale et al. (2004) who implemented an actveset QP optimisation algorithm NPSOL for the estimation of standard SEM models in the software package Mx (Gill et al. 1998). However, the use of purely numerical methods to optimise the likelihood proved to be feasible only for low-dimensional problems and small data sets. In a recent paper, Hamaker et al. (2003) report difficulties in estimating even simple univariate ARMA(1,1) models with $T > 50$ using Mx. An additional problem with the full-sample ML implementation in Mx is the necessity to specify the time-series relationship among variables for each time point and subsequently impose a series of equality constrains to obtain the required stationary dynamic structure.

Note that in the context of dynamic SEM models, the derivation of the theoretical covariance structure in chapter §2 did not assume that $N > T$ nor did require computation of a sufficient statistic (i.e. empirical covariance matrix). Thus, the closed-form covariance structure derived in §2 remains valid for the case when $N < T$ including the case with $N = 1$ (pure time series data). What differs is the log-likelihood, which can no longer be simplified by replacing the data with the sufficient statistic $\boldsymbol{S}$, and the analytical derivatives are consequently different.

In the next section we provide a suitable parametrisation for the full-sample likelihood of the DSEM model and obtain analytical expressions for the gradient using the matrix calculus methods based on zero-one matrices (Turkington 1998, Turkington 2002). This aims to make a contribution to the literature by providing general and programmable closed-form expressions for both the likelihood and the score. We aim to obtain expressions that can be written in terms of parameter matrices/vectors and zero-one matrices, and are hence easily implementable in modern object-oriented programming languages such as S or R (Venables and Ripley 2000, Venables and Ripley 2002).

In addition to providing the analytical results needed for full-sample maximum likelihood estimation of DSEM models, we suggest a simple approach to programming the estimation algorithms.

In the next section we specify the DSEM model in the form that enables us

[3]The gradient (analytic derivatives) are not provided in the S-Plus version of `SsfPack`, while the Ox version provides the gradient only for certain parameters of the model.

to obtain a closed form likelihood and covariance structure for pure time series data. Third section derives the analytical derivatives and develops functions that implement the likelihood and score formulae in the S language. The fourth section outlines an approach to estimation of the parameters by using the suggested functions in combination with the existing optimisation routines in S-Plus and R environments.

4.2 The likelihood function

Consider the DSEM(p, q) model suggested in section §2. The model can be written in the recursive form (i.e. for a typical time t) as

$$\boldsymbol{\eta}_t = \sum_{j=0}^{p} \boldsymbol{B}_j \boldsymbol{\eta}_{t-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{\zeta}_t \tag{4.4}$$

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_y \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t \tag{4.5}$$

$$\boldsymbol{x}_t = \boldsymbol{\Lambda}_x \boldsymbol{\xi}_t + \boldsymbol{\delta}_t \tag{4.6}$$

where $\boldsymbol{\eta}_t = (\eta_t^{(1)}, \eta_t^{(2)}, \ldots, \eta_t^{(f)})'$ and $\boldsymbol{\xi}_t = (\xi_t^{(1)}, \xi_t^{(2)}, \ldots, \xi_t^{(g)})'$ are vectors of possibly unobserved (latent) variables, $\boldsymbol{y}_t = (y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(n)})'$ and $\boldsymbol{x}_t = (x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(k)})'$ are vectors of observable variables, and $\boldsymbol{B}_j$ $(f \times f)$, $\boldsymbol{\Gamma}_j$ $(f \times g)$, $\boldsymbol{\Lambda}_y$ $(n \times f)$, and $\boldsymbol{\Lambda}_x$ $(k \times g)$ are coefficient matrices. The contemporaneous and simultaneous coefficients are in $\boldsymbol{B}_0$, and $\boldsymbol{\Gamma}_0$, while $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, ..., $\boldsymbol{B}_p$, and $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, ..., $\boldsymbol{\Gamma}_q$ contain coefficients of the lagged variables.

The structural part of the model (4.4) is a standard dynamic simultaneous equation system, but with measurement error in all variables or with all variables unobservable (latent). Therefore, the unobservable variables need to be measured by observable indicators. This is achieved by the equations (4.5) and (4.6), which are specified as factor analysis models.

By making an additional assumption that $\boldsymbol{\xi}_t$ follows a vector autoregressive process of order s, i.e., VAR(s) process with $s = \max(p, q)$ we can append the model (4.4)–(4.6) with

$$\boldsymbol{\xi}_t = \sum_{j=1}^{s} \boldsymbol{R}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{v}_t, \tag{4.7}$$

which allows us to write (4.4)–(4.7) in a simplified form as

$$\underbrace{\begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix}}_{h_t\ (m\times 1)} = \sum_{j=0}^{s} \underbrace{\begin{pmatrix} B_j & \Gamma_j \\ 0 & R_j \end{pmatrix}}_{C_j\ (m\times m)} \underbrace{\begin{pmatrix} \eta_{t-j} \\ \xi_{t-j} \end{pmatrix}}_{h_t\ (m\times 1)} + \underbrace{\begin{pmatrix} \zeta_t \\ v_t \end{pmatrix}}_{z_t\ (m\times 1)}$$

$$\underbrace{\begin{pmatrix} y_t \\ x_t \end{pmatrix}}_{w_t\ (r\times 1)} = \underbrace{\begin{pmatrix} \Lambda_y & 0 \\ 0 & \Lambda_x \end{pmatrix}}_{\Lambda\ (r\times m)} \underbrace{\begin{pmatrix} \eta_t \\ \xi_t \end{pmatrix}}_{h_t\ (m\times 1)} + \underbrace{\begin{pmatrix} \varepsilon_t \\ \delta_t \end{pmatrix}}_{e_t\ (r\times 1)},$$

where $R_0 \equiv 0$, $m \equiv f+g$, and $r \equiv n+k$. We will hereafter work with the simplified notation writing the DSEM model (4.4)–(4.7) as

$$h_t = \sum_{j=0}^{s} C_j h_t + z_t \tag{4.8}$$

$$w_t = \Lambda h_t + e_t. \tag{4.9}$$

Note that (4.8)–(4.9) can be interpreted as a state-space form of the model (4.4)–(4.7).

The formulation (4.8)–(4.9), however, contains too many unknowns and we need to factor out the latent variables, for which purpose we will revert to the vectorised version of the model. Hence, by vectorising (4.8)–(4.9), we obtain

$$\underbrace{\mathrm{vec}\{h_t\}_1^T}_{H_T\ (mT\times 1)} = \left(\sum_{j=0}^{s} S_T^j \otimes C_j\right) \underbrace{\mathrm{vec}\{h_t\}_1^T}_{H_T (mT\times 1)} + \underbrace{\mathrm{vec}\{z_t\}_1^T}_{Z_T\ (mT\times 1)} \tag{4.10}$$

$$\underbrace{\mathrm{vec}\{w_t\}_1^T}_{W_T\ (rT\times 1)} = (I_T \otimes \Lambda) \underbrace{\mathrm{vec}\{h_t\}_1^T}_{H_T\ (mT\times 1)} + \underbrace{\mathrm{vec}\{e_t\}_1^T}_{E_T\ (rT\times 1)}, \tag{4.11}$$

which now enables us to re-write the model with only observable variables on the left-hand side and latent vector H_T factored out

$$W_T = (I_T \otimes \Lambda)\left(I_{mT} - \sum_{j=0}^{s} S_T^j \otimes C_j\right)^{-1} Z_T + E_T. \tag{4.12}$$

Let

$$\Psi \equiv E\begin{bmatrix} \zeta_t\zeta'_t & \zeta_t v'_t \\ v_t\zeta'_t & v_t v'_t \end{bmatrix}, \quad \Theta \equiv E\begin{bmatrix} \varepsilon_t\varepsilon'_t & \varepsilon_t\delta'_t \\ \delta_t\varepsilon'_t & \delta_t\delta'_t \end{bmatrix}.$$

It follows that the model-implied covariance matrix is of the form $\Sigma(\theta) = E[W_T W'_T]$, which can be evaluated as

$$\begin{aligned}\Sigma(\boldsymbol{\theta}) &= E\left[\left((\boldsymbol{I}_T \otimes \Lambda)\boldsymbol{X}^{-1}\boldsymbol{Z}_T + \boldsymbol{E}_T\right)\left((\boldsymbol{I}_T \otimes \Lambda)\boldsymbol{X}^{-1}\boldsymbol{Z}_T + \boldsymbol{E}_T\right)'\right] \\ &= (\mathbf{I}_T \otimes \Lambda)\boldsymbol{X}^{-1}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{Z}_T'\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Psi}}\boldsymbol{X}'^{-1}(\boldsymbol{I}_T \otimes \Lambda') + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{E}_T'\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}} \\ &= (\boldsymbol{I}_T \otimes \Lambda)\boldsymbol{X}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi})\boldsymbol{X}'^{-1}(\boldsymbol{I}_T \otimes \Lambda') + \boldsymbol{I}_T \otimes \boldsymbol{\Theta} \qquad (4.13)\end{aligned}$$

where $\boldsymbol{X} \equiv \boldsymbol{I}_{mT} - \sum_{j=0}^{s} \boldsymbol{S}_T^j \otimes \boldsymbol{C}_j$ and $\boldsymbol{\theta}' \equiv [(\mathrm{vec}\,\boldsymbol{C}_j)', (\mathrm{vec}\,\Lambda)', (\mathrm{vech}\,\boldsymbol{\Psi})', (\mathrm{vech}\,\boldsymbol{\Theta})']$, where for simplicity, we write only the j-th (generic) lag-coefficient matrix $\boldsymbol{C}_j$. By introducing further simplifying notation $\boldsymbol{A} \equiv \boldsymbol{I}_T \otimes \Lambda$, and $\boldsymbol{B} \equiv \boldsymbol{I}_T \otimes \boldsymbol{\Psi}$ we obtain a more compact expression for $\Sigma(\boldsymbol{\theta})$,

$$\Sigma(\boldsymbol{\theta}) = \boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}' + \boldsymbol{I}_T \otimes \boldsymbol{\Theta} \qquad (4.14)$$

The $\Sigma(\boldsymbol{\theta})$ matrix (4.14) can be easily programmed using object-constructor functions as follows. Firstly, we define the S function that constructs $\boldsymbol{X} \equiv \boldsymbol{I}_{mT} - \sum_{j=0}^{s} \boldsymbol{S}_T^j \otimes \boldsymbol{C}_j$ as

```
"Xsum" <- function(x,m,ss,tt)
{
    Cm=array(dim=c(m,m,(ss+1)))
        for(i in 0:ss)
        {
            Cm[,,i+1]=matrix(x[(i*(m^2)+1):((i+1)*m^2)],m,m)
        }
    Xm=kronecker(power.shift(tt,0),Cm[,,1])
    if(ss>=1)
        {
        for(i in 1:ss)
            {
            Xm=Xm+kronecker(power.shift(tt,i),Cm[,,(i+1)])
            }
        return(solve(diag(m*tt)-Xm))
        }
    else
        {
        return(solve(diag(m*tt)-Xm))
        }
}
```

Similarly, the constructor function for $\boldsymbol{A} \equiv \boldsymbol{I}_T \otimes \Lambda$ can be written as

```
"makeA" <- function(x,m,r,ss,tt)
{
    Ad=(ss+1)*m^2
    kronecker(diag(tt),matrix(x[(Ad+1):(Ad+r*m)],r,m))
}
```

Next, the $\boldsymbol{B} \equiv \boldsymbol{I}_T \otimes \boldsymbol{\Psi}$ matrix can be built with the following function,

```
"makeB" <- function(x,m,r,ss,tt)
{
    Pd=(ss+1)*m^2+r*m
    kronecker(diag(tt),SymMat(m,x[(Pd+1):(Pd+m*(m+1)/2)]))
}
```

Finally, the last component we need is $\boldsymbol{I}_T \otimes \boldsymbol{\Theta}$, which can be constructed simply as

```
"makeT" <- function(x,m,r,ss,tt)
{
    Td=(ss+1)*m^2+r*m+m*(m+1)/2
    kronecker(diag(tt),SymMat(r,x[(Td+1):(Td+r*(r+1)/2)]))
}
```

where we made use of a function that constructs a symmetric matrix, SymMat,[4] by taking a vector argument, i.e.,

```
"SymMat" <- function(k,x)
{
    M = as.matrix(bdsmatrix(k, x))
    return(M)
}
```

With these functions at hand, we can programme an expression for DSEM covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, namely,

```
"mSigma" <- function(x,m,r,ss,tt)
{
    Xm=Xsum(x,m,ss,tt)
    Am=makeA(x,m,r,ss,tt)
    Bm=makeB(x,m,r,ss,tt)
    Tm=makeT(x,m,r,ss,tt)
    SM <- Am%*%Xm%*%Bm%*%t(Xm)%*%t(Am)%*%Tm
}
```

Note that (4.14) gives the expression for the theoretical covariance matrix of the DSEM model in a closed form as a function of model parameters only. Therefore, the closed-form log-likelihood of the DSEM model is

$$\ln L\left(\boldsymbol{W}_T\right) = -\frac{T}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right| - \frac{1}{2}\mathrm{tr}\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T. \tag{4.15}$$

To obtain the maximum likelihood estimates of $\boldsymbol{\theta}$ note that maximising the log-likelihood (4.15) is equivalent to minimising[5]

[4]Note that our SymMat function makes use of the bdsmatrix, which is included in S-Plus, but not in R.

[5]Note that $\mathrm{tr}\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T = \boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T$ since $\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T$ is a scalar.

$$\max\left(\ln L\left(\boldsymbol{W}_T\right)\right) = \arg\min_{\boldsymbol{\theta}} \left(\ln\left|\Sigma\left(\boldsymbol{\theta}\right)\right| + \boldsymbol{W}_T'\Sigma^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T\right),$$

thus our objective function is

$$F_{ML} = \ln\left|\Sigma\left(\boldsymbol{\theta}\right)\right| + \boldsymbol{W}_T'\Sigma^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T. \tag{4.16}$$

The log-likelihood (4.16) is a function of the $\Sigma(\boldsymbol{\theta})$ covariance matrix, hence it can be easily programmed using the mSigma function, thus we can construct F_{ML} as

```
"makeFML" <- function(x,data,m,r,ss,tt)
{
    Xm=mSigma(x,m,r,ss,tt)
    vecWm=vec(as.matrix(data))
    LnS=log(abs(det(solve(Xm))))
    trWm=as.numeric(crossprod(vecWm,Xm)%*%vecWm)
    return(LnS+trWm)
}
```

Table 4.1: S+ functions for likelihood evaluation

Matrix/vector	Dimension	S+ function	Arguments
$(\boldsymbol{I}_{mT} - \sum_{j=0}^{s} \boldsymbol{S}_T^j \otimes \boldsymbol{C}_j)^{-1}$	$mT \times mT$	Xsum	makeC, f, g, s, T
$\boldsymbol{A} \equiv \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}$	$rT \times mT$	makeA	$\boldsymbol{\Lambda}_y$, $\boldsymbol{\Lambda}_x$
$\boldsymbol{B} \equiv \boldsymbol{I}_T \otimes \Psi$	$rT \times rT$	makeB	$\boldsymbol{\Psi}$, T
$\boldsymbol{I}_T \otimes \Theta$	$rT \times rT$	makeT	$\boldsymbol{\Theta}$, T
$\Sigma^{-1}(\boldsymbol{\theta})$	$rT \times rT$	mSigma	makeA, makeB, Xsum, makeT
$\ln\left|\Sigma(\boldsymbol{\theta})\right| + \boldsymbol{W}_T'\Sigma^{-1}(\boldsymbol{\theta})\boldsymbol{W}_T$	1×1	makeFML	mSigma, $m \times m$ data matrix $\boldsymbol{X}$
–	–	DSEMobj	$\boldsymbol{B}_1, \ldots, \boldsymbol{B}_s$, $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_s$, $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_s$, $\boldsymbol{\Psi}$, $\boldsymbol{\Lambda}_y$, $\boldsymbol{\Lambda}_x$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}$

In the next section we will derive the analytical expressions for the derivatives of (4.16) and thus obtain the score vector.

4.3 The score

We make use of several useful zero-one matrices, $\boldsymbol{S}^j$, $\boldsymbol{K}_{ab}$ and $\boldsymbol{D}_a$. The $\boldsymbol{S}^j$ is the general (integer) power of the shifting matrix $\boldsymbol{S}$, defined in (2.11), which can be constructed as

```
"power.shift" <- function(x,p)
{
         if(p==0)  {diag(x)}
    else if(p==1)  {shift.mat(x)}
    else     {
             s1=diag(nrow(shift.mat(x)))
             s2=shift.mat(x)
             while(p>0)
                {
                   if(p%%2)
                   s1=s1%*%s2
                    p=p%/%2
                   s2=s2%*%s2
                }
    return(s1)
             }
}
```

where the shifting matrix $\boldsymbol{S}$ can be obtained with a simple function

```
"shift.mat" <- function(x)
{
    S1 <- diag(x - 1)
    S2 <- cbind(S1, rep(0, x - 1))
    S3 <- rbind(rep(0, x), S2)
    S3
}
```

We define the commutation matrix $\boldsymbol{K}_{ab}$ as an orthogonal $ab \times ab$ zero-one permutation matrix

$$\boldsymbol{K}_{ab} \equiv \left(\boldsymbol{I}_a \otimes \boldsymbol{e}_1^b : \boldsymbol{I}_a \otimes \boldsymbol{e}_2^b : \cdots : \boldsymbol{I}_a \otimes \boldsymbol{e}_b^b\right) \tag{4.17}$$

such that $\boldsymbol{K}_{ab} \operatorname{vec} \boldsymbol{X} = \operatorname{vec} \boldsymbol{X}'$, where $\boldsymbol{e}_j^b$ is the jth column of a $b \times b$ identity matrix, i.e., $\boldsymbol{I}_b = \left(\boldsymbol{e}_1^b : \boldsymbol{e}_2^b : \cdots : \boldsymbol{e}_b^b\right)$. The S+ function that constructs $\boldsymbol{K}_{ab}$ matrix is given by

```
"K.mn" <- function(x,y)
{
    K = kronecker(diag(x), e.vec(y,1))
    for(i in 2:y)
    {
        K = cbind(K, kronecker(diag(x),e.vec(y,i)))
        K
    }
    return(K)
}
```

Additionally, let

$$K_{ab}^{*} \equiv \mathrm{devec_b} K_{ab} = [I_b \otimes (e_1^a)' : I_b \otimes (e_2^a)' : \cdots : I_b \otimes (e_a^a)'] , \tag{4.18}$$

which can be constructed as a simple zero-one matrix using the following S+ function

```
"K.s.mn" <- function(y,x)
{
    K = kronecker(diag(x), t(e.vec(y,1)))
    for(i in 2:y)
    {
        K = cbind(K, kronecker(diag(x),t(e.vec(y,i))))
        K
    }
    return(K)
}
```

The $a^2 \times a(a+1)/2$ duplication matrix D_a is defined as a zero-one matrix such that for an $a \times a$ matrix X, $D_a \,\mathrm{vech}\, X = \mathrm{vec}\, X$. Computer implementation of the duplication matrix is somewhat more involved and it can be written as

```
"D.n" <- function(x)
{
"make.D" <- function(j,x)
        {
        r =(j-1)*x - sum(1:j-2)
        v = r + x-j
        m = x*(x+1)/2
        D = diag(x^2)[r:v, 1:m]
        return(D)
        }
"make.block.D" <- function(i,x)
        {
        m = x*(x+1)/2
        D.d = make.D(i,x)
        D.e = matrix(NA,i,m)
            for(j in 1:i)
            {
            D.e[j,] = e.vec(m,(i+1)+(j-1)*x - sum(0:(j-1)))
            }
        D = rbind(D.d,D.e)
        return(D)
        }
if(x==1) return(matrix(1,1,1))
    else if(x==2) return(matrix(c(1,0,0,
                                  0,1,0,
                                  0,1,0,
                                  0,0,1),4,3,byrow=T))
    else if(x>2)
    {
        D.mat = make.block.D(1,x)
```

```
    for(i in 2:(x-1))
        {
        D.mat = rbind(D.mat, make.block.D(i,x))
        }
    D = rbind(D.mat,make.D(x,x))
    return(D)
    }
}
```

Note that the generalised inverse of $\boldsymbol{D}_n$ can be obtained simply as

```
"D.rev" <- function(x)
{
    solve(t(D.n(x))%*%D.n(x))%*%t(D.n(x))
}
```

This matrix is important since for a symmetric $m \times m$ matrix $\boldsymbol{X}$

$$\frac{\partial \text{vec}\boldsymbol{X}}{\partial \text{vech}\boldsymbol{X}} = \boldsymbol{D}_m^*,$$

where $\boldsymbol{D}_m^*$ is the generalised inverse of $\boldsymbol{D}_m$, i.e., $\boldsymbol{D}_m^* = (\boldsymbol{D}_m'\boldsymbol{D}_m)^{-1}\boldsymbol{D}_m'$, hence vech$\boldsymbol{X} = \boldsymbol{D}_m^*vec\boldsymbol{X}$. Note that unlike $\boldsymbol{D}_m$, $\boldsymbol{D}_m^*$ is not a 'zero-one' matrix. Turkington (2002) (p. 112) gives an incorrect expression $\frac{\partial \text{vec}\boldsymbol{X}}{\partial \text{vech}\boldsymbol{X}} = \boldsymbol{D}_m'$. To see this consider a simple numerical example,

```
> cbind(vech(SymMat(3, 1:6)), D.rev(3) %*% vec(SymMat(3, 1:6)))
     [,1] [,2]
[1,]    1    1
[2,]    2    2
[3,]    3    3
[4,]    4    4
[5,]    5    5
[6,]    6    6
```

hence vech$\boldsymbol{X} = \boldsymbol{D}_m^*vec\boldsymbol{X}$. On the other hand,

```
> cbind(vech(SymMat(3, 1:6)), t(D.n(3)) %*% vec(SymMat(3, 1:6)))
     [,1] [,2]
[1,]    1    1
[2,]    2    4
[3,]    3    6
[4,]    4    4
[5,]    5   10
[6,]    6    6
```

thus vech$\boldsymbol{X} \neq \boldsymbol{D}_m'vec\boldsymbol{X}$.

In addition, we will need convenient constructor functions for vec and vech operators, we use the following:

```
"vec" <- function(x)
{
    if(!is.matrix(x))
    {
    stop("The argument must be a matrix")
    }
    n=dim(x)[1]
    m=dim(x)[2]
    d=n*m
    vecM=as.numeric(x)
    return(as.matrix(vecM,d))
}
```

and

```
"vech" <- function(x)
{
    if(!is.matrix(x) & dim(x)[1]==dim(x)[2])
        {
        stop("The argument must be a square matrix")
        }

    vechM = x[,1]
    for(i in 1:(dim(x)[1]-1))
    {
        vechM = c(vechM, x[-c(1:i),(i+1)])
        vechM
    }
    return(as.matrix(vechM))
}
```

Finally, we will make use of a simple function that builds columns of an identity matrix,

```
"e.vec" <- function(x,y)
{
    e = as.matrix(rep(0,x))
    e[y,1] <- 1
    return(e)
}
```

Preliminaries. Note that from (4.14) it follows that the derivative vector of $\mathbf{\Sigma}(\boldsymbol{\theta})$ in respect to any coefficient vector contained in $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}_i$ is of the form

$$\begin{aligned}\frac{\partial \text{vec}\mathbf{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} &= \frac{\partial \text{vec}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}' + \boldsymbol{I}_T \otimes \Theta\right)}{\partial \boldsymbol{\theta}_i} \\ &= \frac{\partial \text{vec}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\right)}{\partial \boldsymbol{\theta}_i} + \frac{\partial \text{vec}\left(\boldsymbol{I}_T \otimes \Theta\right)}{\partial \boldsymbol{\theta}_i}. \end{aligned} \tag{4.19}$$

Additionally, by the chain rule of matrix calculus we have

$$\frac{\partial ln\left|\Sigma\left(\boldsymbol{\theta}\right)\right|}{\partial\boldsymbol{\theta}_i}=\left(\frac{\partial \mathrm{vec}\Sigma\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}_i}\right)\left(\frac{\partial ln\left|\Sigma\left(\boldsymbol{\theta}\right)\right|}{\partial\,\mathrm{vec}\Sigma\left(\boldsymbol{\theta}\right)}\right)=\left(\frac{\partial \mathrm{vec}\Sigma\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}_i}\right)\mathrm{vec}\Sigma^{-1}\left(\boldsymbol{\theta}\right)\qquad(4.20)$$

We can now derive the components of the score vector for the likelihood function (4.16). We shall do so by obtaining the expressions for partial derivatives of (4.16) in respect to the components of the parameters vector $\boldsymbol{\theta}$, namely $(\mathrm{vec}\,\boldsymbol{C}_j)$, $(\mathrm{vec}\,\boldsymbol{\Lambda})$, $(\mathrm{vech}\,\boldsymbol{\Psi})$, and $(\mathrm{vech}\,\Theta)$. The score vector is then constructed from these individual vectors of partial derivatives (see e.g. Turkington (2002))

(i) Partial derivative in respect to $\boldsymbol{C}_i$, $\partial F_{ML}/\partial\,\mathrm{vec}\,\boldsymbol{C}_i$

The first component of the score vector is the derivative of log-likelihood (4.16) given by

$$\frac{\partial F_{ML}}{\partial \mathrm{vec}\boldsymbol{C}_i}=\frac{\partial \ln\left|\Sigma\left(\boldsymbol{\theta}\right)\right|}{\partial\,\mathrm{vec}\boldsymbol{C}_i}+\frac{\partial \mathrm{tr}\boldsymbol{W'}_T\Sigma^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T}{\partial \mathrm{vec}\boldsymbol{C}_i}.\qquad(4.21)$$

We will evaluate the two components of (4.21) in turn. Firstly, we have

$$\begin{aligned}
\frac{\partial \ln\left|\Sigma\left(\boldsymbol{\theta}\right)\right|}{\partial\,\mathrm{vec}\boldsymbol{C}_i} &= \left(\frac{\partial\,\mathrm{vec}\Sigma\left(\boldsymbol{\theta}\right)}{\partial\,\mathrm{vec}\boldsymbol{C}_i}\right)\mathrm{vec}\Sigma^{-1}\left(\boldsymbol{\theta}\right)\\
&= \frac{\partial \mathrm{vec}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'+\boldsymbol{I}_T\otimes\Theta\right)}{\partial \mathrm{vec}\boldsymbol{C}_i}=\frac{\partial \mathrm{vec}\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'}{\partial \mathrm{vec}\boldsymbol{C}_i}\\
&= \left(\frac{\partial\,\mathrm{vec}\boldsymbol{X}}{\partial\,\mathrm{vec}\boldsymbol{C}_i}\right)\left(\frac{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}}{\partial\,\mathrm{vec}\boldsymbol{X}}\right)\left(\frac{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}}{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}}\right)\left(\frac{\partial\,\mathrm{vec}\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'}{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}}\right)\\
&= \left[\boldsymbol{K}^*_{Tm}\left(\boldsymbol{I}_{mT}\otimes\boldsymbol{S}'^i_T\right)\otimes\boldsymbol{I}_m\right]\left(\boldsymbol{X}^{-1}\otimes\boldsymbol{X}'^{-1}\right)\\
&\times \left[\boldsymbol{B}\boldsymbol{X}'^{-1}\otimes\boldsymbol{I}_{mT}+\left(\boldsymbol{B}'\boldsymbol{X}'^{-1}\otimes\boldsymbol{I}_{mT}\right)\boldsymbol{K}_{mT,mT}\right]\left(\boldsymbol{A}'\otimes\boldsymbol{A}'\right)\mathrm{vec}\Sigma^{-1}\left(\boldsymbol{\theta}\right),\qquad(4.22)
\end{aligned}$$

where we made use of the following results

$$\begin{aligned}
\frac{\partial\,\mathrm{vec}\boldsymbol{X}}{\partial\,\mathrm{vec}\boldsymbol{C}_i} &= \frac{\partial \mathrm{vec}\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{s}\boldsymbol{S}^j_T\otimes\boldsymbol{C}_j\right)}{\partial \mathrm{vec}\boldsymbol{C}_i}\\
&= -\frac{\partial \mathrm{vec}\left(\boldsymbol{S}^i_T\otimes\boldsymbol{C}_i\right)}{\partial \mathrm{vec}\boldsymbol{C}_i}=-\boldsymbol{K}^*_{Tm}\left(\boldsymbol{I}_{mT}\otimes\boldsymbol{S}'^i_T\right)\otimes\boldsymbol{I}_m\qquad(4.23)\\
\frac{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}}{\partial\,\mathrm{vec}\boldsymbol{X}} &= -\boldsymbol{X}^{-1}\otimes\boldsymbol{X}'^{-1}\qquad(4.24)\\
\frac{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}}{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}} &= \boldsymbol{B}\boldsymbol{X}'^{-1}\otimes\boldsymbol{I}_{mT}+\left(\boldsymbol{B}'\boldsymbol{X}'^{-1}\otimes\boldsymbol{I}_{mT}\right)\boldsymbol{K}_{mT,mT}\qquad(4.25)\\
\frac{\partial\,\mathrm{vec}\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'}{\partial\,\mathrm{vec}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}} &= \boldsymbol{A}'\otimes\boldsymbol{A}'.\qquad(4.26)
\end{aligned}$$

The second part of (4.21) is

$$\begin{aligned}
\frac{\partial \mathrm{tr} \boldsymbol{W}'_T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{W}_T}{\partial \mathrm{vec} \boldsymbol{C}_i} &= \left(\frac{\partial \mathrm{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \mathrm{vec} \boldsymbol{C}_i}\right)\left(\frac{\partial \mathrm{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \mathrm{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}\right)\left(\frac{\partial \mathrm{tr} \boldsymbol{W}'_T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{W}_T}{\partial \mathrm{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}\right) \\
&= -[\boldsymbol{K}^*_{Tm} (\boldsymbol{I}_{mT} \otimes \boldsymbol{S}'^{i}_T) \otimes \boldsymbol{I}_m] (\boldsymbol{X}^{-1} \otimes \boldsymbol{X}'^{-1}) \\
&\times \left[\boldsymbol{B}\boldsymbol{X}'^{-1} \otimes \boldsymbol{I}_{mT} + (\boldsymbol{B}'\boldsymbol{X}'^{-1} \otimes \boldsymbol{I}_{mT}) \boldsymbol{K}_{mT,mT}\right] \\
&\times (\boldsymbol{A}' \otimes \boldsymbol{A}') \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \boldsymbol{W}_T \otimes \boldsymbol{W}_T, \qquad (4.27)
\end{aligned}$$

noting that

$$\begin{aligned}
\frac{\partial \mathrm{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \mathrm{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})} &= -\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \qquad (4.28) \\
\frac{\partial \mathrm{tr} \boldsymbol{W}'_T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{W}_T}{\partial \mathrm{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})} &= \frac{\partial \mathrm{vec} \boldsymbol{W}'_T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \boldsymbol{W}_T}{\partial \mathrm{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})} = \boldsymbol{W}_T \otimes \boldsymbol{W}_T. \qquad (4.29)
\end{aligned}$$

Therefore, it follows that the first component of the score vector is given by

$$\begin{aligned}
\frac{\partial F_{ML}}{\partial \mathrm{vec} \boldsymbol{C}_i} &= [\boldsymbol{K}^*_{Tm} (\boldsymbol{I}_{mT} \otimes \boldsymbol{S}'^{i}_T) \otimes \boldsymbol{I}_m] (\boldsymbol{X}^{-1} \otimes \boldsymbol{X}'^{-1}) \left[\boldsymbol{B}\boldsymbol{X}'^{-1} \otimes \boldsymbol{I}_{mT} + (\boldsymbol{B}'\boldsymbol{X}'^{-1} \otimes \boldsymbol{I}_{mT}) \boldsymbol{K}_{mT,mT}\right] \\
&\times (\boldsymbol{A}' \otimes \boldsymbol{A}') \left[\mathrm{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - (\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) \boldsymbol{W}_T \otimes \boldsymbol{W}_T\right]. \qquad (4.30)
\end{aligned}$$

The above expression can be easily programmed in an object-oriented language such as S. Namely, we can define functions that construct components of the score vector using zero-one matrices such as $\boldsymbol{K}^*_{Tm}$ and $\boldsymbol{S}'^{i}_T$. The general structure of these functions relies on a list object "x" constructed with the DSEMobj function,

```
"DSEMobj" <- function(x,m,r,ss,tt)
{
list(X.sum=Xsum(x,m,ss,tt),
     A.mat=makeA(x,m,r,ss,tt),
     B.mat=makeB(x,m,r,ss,tt),
     T.mat=makeT(x,m,r,ss,tt),
     S.mat=mSigma(x,m,r,ss,tt))
}
```

which computes and collects necessary matrices needed for constructing the derivative vectors. The following function constructs the partial derivative $\partial F_{ML}/\partial \mathrm{vec} \boldsymbol{C}_i$ as a single object by taking as arguments a DSEMobj list object (x), an $r \times T$ data matrix data where each row represents a single variable, and model parameters m, T, and j, denoted by m, tt, and j, respectively.[6] The following function constructs the required derivative equation for the j^{th} coefficient matrix $\boldsymbol{C}_j$,

[6]Note we need to avoid using characters such as T, which are reserved characters in the S.

```
"dFmlCj" <- function(x,data,m,tt,j)
{
    VC1=kronecker(K.s.mn(tt,m)%*%kronecker(diag(m*tt),t(power.shift(tt,j))),diag(m))
    VC2=kronecker(x$X.sum,t(x$X.sum))
    VC3=kronecker(x$B.mat%*%t(x$X.sum),diag(m*tt))
    VC4=kronecker(t(x$B.mat)%*%t(x$X.sum),diag(m*tt))%*%K.mn(m*tt,m*tt)
    VC5=kronecker(t(x$A.mat),t(x$A.mat))
    VC6=vec(x$S.mat)
    VC7=kronecker(x$S.mat,x$S.mat)%*%kronecker(vec(Wm),vec(Wm))
    return(VC1%*%VC2%*%(VC3+VC4)%*%VC5%*%(VC6-VC7))
}
```

Putting a simple loop around j it easy to build a vector that contains all s derivatives (where s is the lag length),[7] namely we can write

```
"dFmlC" <- function(x,data,m,ss,tt)
{
    if(ss>0)
    {
    DC <- dFmlCj(x,data,m,tt,j=0)
        for(k in 1:ss)
        {
            DC = rbind(DC, dFmlCj(x,data,m,tt,k))
        }
        return(DC)
    }
    else
        return(dFmlCj(x,data,m,tt,j=0))
}
```

(ii) Partial derivative in respect to Λ, $\partial F_{ML}/\partial\,\mathrm{vec}\,\Lambda$

The second component of the score vector is given by

$$\frac{\partial F_{ML}}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} = \frac{\partial \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} + \frac{\partial \mathrm{tr}\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\boldsymbol{W}_T}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}}, \tag{4.31}$$

where

$$\frac{\partial \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} = \left(\frac{\partial\,\mathrm{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}}\right)\mathrm{vec}\boldsymbol{\Sigma}^{-1}. \tag{4.32}$$

It follows that

$$\begin{aligned}
\frac{\partial\,\mathrm{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} &= \frac{\partial\,\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}' + \boldsymbol{I}_T\otimes\Theta\right)}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} \\
&= \frac{\partial\,\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\right)}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} \\
&= \left(\frac{\partial\,\mathrm{vec}\boldsymbol{A}}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}}\right)\left(\frac{\partial\,\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\right)}{\partial\,\mathrm{vec}\boldsymbol{A}}\right) \\
&= (\boldsymbol{K}_{Tm}^{*}\otimes\boldsymbol{I}_r)[(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}) + \left(\boldsymbol{X}'^{-1}\boldsymbol{B}\boldsymbol{X}^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)\boldsymbol{K}_{rT,rT}],
\end{aligned}$$

[7] We use ss rather then s for the lag arugment.

where we made use of the following results

$$\begin{aligned}
\frac{\partial\,\mathrm{vec}\boldsymbol{A}}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} &= \frac{\partial\,\mathrm{vec}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}\right)}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} = \left(\boldsymbol{K}^*_{Tm}\otimes\boldsymbol{I}_r\right)\\
\frac{\partial\,\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\right)}{\partial\,\mathrm{vec}\boldsymbol{A}} &= \left(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)+\left(\boldsymbol{X}'^{-1}\boldsymbol{B}\boldsymbol{X}^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)\boldsymbol{K}_{rT,rT}.
\end{aligned}$$

It now follows that

$$\frac{\partial\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} = \left(\boldsymbol{K}^*_{Tm}\otimes\boldsymbol{I}_{rT}\right)\left[\left(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)+\left(\boldsymbol{X}'^{-1}\boldsymbol{B}\boldsymbol{X}^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)\right]\boldsymbol{K}_{rT,rT}\mathrm{vec}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right),$$

and

$$\begin{aligned}
\frac{\partial\,\mathrm{tr}\boldsymbol{W}'_T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} &= \left(\frac{\partial\,\mathrm{vec}\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}}\right)\left(\frac{\partial\mathrm{vec}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}{\partial\mathrm{vec}\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)}\right)\left(\frac{\partial\mathrm{tr}\boldsymbol{W}'_T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T}{\partial\mathrm{vec}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)}\right)\\
&- \left(\boldsymbol{K}^*_{Tm}\otimes\boldsymbol{I}_r\right)\left[\left(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)+\left(\boldsymbol{X}'^{-1}\boldsymbol{B}\boldsymbol{X}^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)\right]\boldsymbol{K}_{rT,rT}\\
&\times \left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right]\boldsymbol{W}_T\otimes\boldsymbol{W}_T,
\end{aligned}$$

where we made use of the results (4.28) and (4.29). Finally, by combining the above derived terms we obtain

$$\begin{aligned}
\frac{\partial F_{ML}}{\partial\,\mathrm{vec}\boldsymbol{\Lambda}} &= \left(\boldsymbol{K}^*_{Tm}\otimes\boldsymbol{I}_r\right)\left[\left(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)+\left(\boldsymbol{X}'^{-1}\boldsymbol{B}\boldsymbol{X}^{-1}\boldsymbol{A}'\otimes\boldsymbol{I}_{rT}\right)\right]\boldsymbol{K}_{rT,rT}\\
&\times \left[\mathrm{vec}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)-\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\otimes\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\right)\boldsymbol{W}_T\otimes\boldsymbol{W}_T\right] \qquad (4.33)
\end{aligned}$$

An S function that constructs this derivative is given by

```
"dFmlL" <- function(x,data,m,r,tt)
{
    VC1=kronecker(K.s.mn(tt,m),diag(r))
    VC2=kronecker((x$X.sum%*%x$B.mat%*%t(x$X.sum)%*%t(x$A.mat)),diag(r*tt))
    VC3=kronecker((t(x$X.sum)%*%x$B.mat%*%x$X.sum%*%t(x$A.mat)),diag(r*tt))
    VC4=vec(x$S.mat)
    VC5=kronecker(x$S.mat,x$S.mat)%*%kronecker(vec(data),vec(data))
    return(VC1%*%(VC2+VC3)%*%K.mn(r*tt,r*tt)%*%(VC4-VC5))
}
```

(iii) Partial derivative in respect to Ψ, $\partial F_{ML}/\partial\mathrm{vec}\,\Psi$

The third component of the score vector is the partial derivative

$$\frac{\partial F_{ML}}{\partial\mathrm{vech}\boldsymbol{\Psi}} = \frac{\partial\ln\left|\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)\right|}{\partial\mathrm{vech}\boldsymbol{\Psi}} + \frac{\partial\mathrm{tr}\boldsymbol{W}'_T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}\right)\boldsymbol{W}_T}{\partial\mathrm{vech}\boldsymbol{\Psi}}, \qquad (4.34)$$

where the first part can be evaluated as

$$\begin{aligned}\frac{\partial \ln |\Sigma(\boldsymbol{\theta})|}{\partial \text{vech}\boldsymbol{\Psi}} &= \left(\frac{\partial \text{vec}\boldsymbol{\Psi}}{\partial \text{vech}\boldsymbol{\Psi}}\right)\left(\frac{\partial\, \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \text{vec}\boldsymbol{\Psi}}\right)\text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \\ &= \boldsymbol{D}_m^*(\boldsymbol{K}_{Tm}^*\otimes \boldsymbol{I}_m)\left(\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes \boldsymbol{X}'^{-1}\boldsymbol{A}'\right)\text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}), \quad (4.35)\end{aligned}$$

where we used the following results

$$\begin{aligned}\frac{\partial \text{vec}\boldsymbol{\Psi}}{\partial \text{vech}\boldsymbol{\Psi}} &= \boldsymbol{D}_m^* & (4.36)\\ \frac{\partial \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \text{vec}\boldsymbol{\Psi}} &= \frac{\partial \text{vec}\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'}{\partial \text{vec}\boldsymbol{\Psi}} \\ &= \left(\frac{\partial \text{vec}\boldsymbol{B}}{\partial \text{vec}\boldsymbol{\Psi}}\right)\left(\frac{\partial \text{vec}\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'}{\partial \text{vec}\boldsymbol{B}}\right) & (4.37)\\ \frac{\partial \text{vec}\boldsymbol{B}}{\partial \text{vec}\boldsymbol{\Psi}} &= \frac{\partial \text{vec}\left(\boldsymbol{I}_{\text{T}}\otimes \boldsymbol{\Psi}\right)}{\partial \text{vec}\boldsymbol{\Psi}} = (\boldsymbol{K}_{Tm}^*\otimes \boldsymbol{I}_m) & (4.38)\\ \frac{\partial \text{vec}\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}'^{-1}\boldsymbol{A}'}{\partial \text{vec}\boldsymbol{B}} &= \boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes \boldsymbol{X}'^{-1}\boldsymbol{A}'. & (4.39)\end{aligned}$$

The second part of (4.34) yields

$$\begin{aligned}\frac{\partial \text{tr}\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{W}_T}{\partial \text{vech}\boldsymbol{\Psi}} &= \left(\frac{\partial\, \text{vec}\boldsymbol{\Psi}}{\partial \text{vech}\boldsymbol{\Psi}}\right)\left(\frac{\partial\, \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \text{vec}\boldsymbol{\Psi}}\right)\left(\frac{\partial\, \text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial\, \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}\right)\left(\frac{\partial\, \text{tr}\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{W}_T}{\partial\, \text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}\right) \\ &= -\boldsymbol{D}_m^*(\boldsymbol{K}_{Tm}^*\otimes \boldsymbol{I}_m)\left(\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes \boldsymbol{X}'^{-1}\boldsymbol{A}'\right)\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)\boldsymbol{W}_T\otimes \boldsymbol{W}_T,\end{aligned}$$

using the results (4.28), (4.29), and (4.36)–(4.39). Finally, we can obtain the required result as

$$\frac{\partial F_{ML}}{\partial \text{vech}\boldsymbol{\Psi}} = \boldsymbol{D}_m^*(\boldsymbol{K}_{Tm}^*\otimes \boldsymbol{I}_m)\left(\boldsymbol{X}'^{-1}\boldsymbol{A}'\otimes \boldsymbol{X}'^{-1}\boldsymbol{A}'\right)\left[\text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)\boldsymbol{W}_T\otimes \boldsymbol{W}_T\right], \quad (4.40)$$

which can be implemented as an S function as follows,

```
"dFmlP" <- function(x,data,m,tt)
{
    VC1=D.rev(m)%*%kronecker(K.s.mn(tt,m),diag(m))
    VC2=kronecker(t(x$X.sum)%*%t(x$A.mat),t(x$X.sum)%*%t(x$A.mat))
    VC3=vec(x$S.mat)
    VC4=kronecker(x$S.mat,x$S.mat)%*%kronecker(vec(data),vec(data))
    return(VC1%*%VC2%*%(VC3-VC4))
}
```

(vi) Partial derivative in respect to Θ, $\partial F_{ML}/\partial \text{vec}\,\Theta$

The final, fourth, component of the score vector is

$$\frac{\partial F_{ML}}{\partial \text{vech}\Theta} = \frac{\partial \ln |\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial \text{vech}\Theta} + \frac{\partial \text{tr}\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{W}_T}{\partial \text{vech}\Theta}, \tag{4.41}$$

where

$$\begin{aligned}
\frac{\partial \ln |\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial \text{vech}\Theta} &= \left(\frac{\partial\, \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \text{vech}\Theta}\right) \text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \\
&= \boldsymbol{D}_r^*(\boldsymbol{K}_{Tr}^* \otimes \boldsymbol{I}_r)\text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}),
\end{aligned} \tag{4.42}$$

which is obtained by noting that

$$\frac{\partial\, \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \text{vech}\Theta} = \frac{\partial\, \text{vec}(\boldsymbol{I}_T \otimes \Theta)}{\partial \text{vech}\Theta} = \left(\frac{\partial\, \text{vec}\Theta}{\partial \text{vech}\Theta}\right)\left(\frac{\partial\, \text{vec}(\boldsymbol{I}_T \otimes \Theta)}{\partial \text{vec}\Theta}\right) \tag{4.43}$$

$$\frac{\partial\, \text{vec}\Theta}{\partial \text{vech}\Theta} = \boldsymbol{D}_r^* \tag{4.44}$$

$$\frac{\partial \text{vec}(\boldsymbol{I}_T \otimes \Theta)}{\partial \text{vec}\Theta} = (\boldsymbol{K}_{Tr}^* \otimes \boldsymbol{I}_r). \tag{4.45}$$

The second part of (4.41) yields

$$\begin{aligned}
\frac{\partial \text{tr}\mathbf{W}_T'\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{W}_T}{\partial \text{vech}\Theta} &= \frac{\partial \text{vec}\mathbf{W}_T'\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{W}_T}{\partial \text{vech}\Theta} \\
&= \left(\frac{\partial \text{vec}\Theta}{\partial \text{vech}\Theta}\right)\left(\frac{\partial \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \text{vec}\Theta}\right)\left(\frac{\partial \text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \text{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}\right)\left(\frac{\partial \text{vec}\boldsymbol{W}_T'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{W}_T}{\partial \text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}\right),
\end{aligned}$$

where we made use of (4.28), (4.44), (4.45). It follows that

$$\frac{\partial F_{ML}}{\partial \text{vech}\Theta} = \boldsymbol{D}_r^*(\boldsymbol{K}_{Tr}^* \otimes \boldsymbol{I}_r)\left[\text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right](\boldsymbol{W}_T \otimes \boldsymbol{W}_T)\right] \tag{4.46}$$

which completes the derivation of the score vector. The required S function can be written using similar approach as before, i.e.,

```
"dFmlT" <- function(x,data,r,tt)
{
    VC1=D.rev(r)%*%kronecker(K.s.mn(tt,r),diag(r))
    VC2=vec(x$S.mat)
    VC3=kronecker(x$S.mat,x$S.mat)%*%kronecker(vec(data),vec(data))
    return(VC1%*%(VC2-VC3))
}
```

Note that we might wish impose diagonal structure on the Θ matrix (as in the classical factor analysis model, for example). This can be accomplished by modifying the above derivative as follows. It is easy to see that

$$\frac{\partial \text{vec}\boldsymbol{\Theta}}{\partial \text{diag}\boldsymbol{\Theta}} = \begin{pmatrix} \boldsymbol{e}_1^r & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{e}_2^r & \cdots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{e}_r^r \end{pmatrix} \equiv \boldsymbol{D}_{diag},$$

which can be implemented simply as

```
"D.diag" <- function(x)
{
    D=matrix(0,x,x^2)
    D[1,1]=1
    for(i in 2:x) D[i,(i+(i-1)*x)]=1
    return(D)
}
```

which leads to the modified derivative expression

$$\frac{\partial F_{ML}}{\partial \text{vech}\boldsymbol{\Theta}} = \boldsymbol{D}_{diag} \left(\boldsymbol{K}_{Tr}^{*} \otimes \boldsymbol{I}_r\right) \left[\text{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \left(\boldsymbol{W}_T \otimes \boldsymbol{W}_T\right)\right] \tag{4.47}$$

Should we need to impose diagonal structure on $\boldsymbol{\Theta}$, matrix, we can use the following modified function,

```
"dFmlTdiag" <- function(x,data,r,tt)
{
    VC1=D.diag(r)%*%kronecker(K.s.mn(tt,r),diag(r))
    VC2=vec(x$S.mat)
    VC3=kronecker(x$S.mat,x$S.mat)%*%kronecker(vec(data),vec(data))
    return(VC1%*%(VC2-VC3))
}
```

Collecting the terms, the score vector is therefore given by

$$\boldsymbol{\Delta} F_{ML} = \begin{pmatrix} \frac{\partial F_{ML}}{\partial \text{vec}\boldsymbol{C}_1} \\ \vdots \\ \frac{\partial F_{ML}}{\partial \text{vec}\boldsymbol{C}_s} \\ \frac{\partial F_{ML}}{\partial \text{vec}\boldsymbol{\Lambda}} \\ \frac{\partial F_{ML}}{\partial \text{vech}\boldsymbol{\Psi}} \\ \frac{\partial F_{ML}}{\partial \text{vech}\boldsymbol{\Theta}} \end{pmatrix}, \tag{4.48}$$

which can be straightforwardly translated into an S function as

```
"ScoreFml" <- function(w,data,m,r,ss,tt)
{
    DSEM=DSEMobj(w,m,r,ss,tt)
    VC1=dFmlC(DSEM,data,m,ss,tt)
```

```
    VC2=dFmlL(DSEM,data,m,r,tt)
    VC3=dFmlP(DSEM,data,m,tt)
    VC4=dFmlT(DSEM,data,r,tt)
    return(rbind(VC1,VC2,VC3,VC4))
}
```

Therefore, we have a single function that constructs the entire score vector in a single call. Along with the likelihood evaluation function `makeFML`, the score (analytic derivative) function `ScoreFml` provide the necessary inputs for standard optimisation routines.

Table 4.2: S+ functions for score evaluation

Matrix/vector	Dimension/type	S+ function	Arguments
$\boldsymbol{S}_T$	$T \times T$ matrix	`shift.mat`	T
$\boldsymbol{S}_T^j$	$T \times T$ matrix	`power.shift`	T, j
$\text{vec}(\boldsymbol{X})$	$ab \times 1$ vector	`vec`	$\boldsymbol{X}$ $(a \times b)$
$\text{vech}(\boldsymbol{X})$	$a(a+1)/2 \times 1$ vector	`vech`	$\boldsymbol{X}$ $(a \times a)$
$\boldsymbol{e}_j^a$ (j^{th} column of $\boldsymbol{I}_a$)	$a \times 1$ vector	`e.vec`	a, j
$\boldsymbol{D}_n$	$n^2 \times n(n+1)/2$ matrix	`D.n`	n (dim.)
$\boldsymbol{D}_n^*$	$n(n+1)/2 \times n^2$ matrix	`D.rev`	n (dim.)
$\boldsymbol{D}_{diag}$	$n \times n^2/2$ matrix	`D.diag`	n (dim.)
$\boldsymbol{K}_{mn}$	$m \times n$ matrix	`K.mn`	n, m (dim.)
$\boldsymbol{K}_{mn}^*$	$m \times n$ matrix	`K.s.mn`	n, m (dim.)
$\frac{\partial F_{ML}}{\partial \text{vec}\boldsymbol{C}}$	$m^2 \times 1$ vector	`dFmlC`	`DSEMobj`, $\boldsymbol{W}_T$
$\frac{\partial F_{ML}}{\partial \text{vec}\boldsymbol{\Lambda}}$	$rm \times 1$ vector	`dFmlL`	`DSEMobj`, $\boldsymbol{W}_T$
$\frac{\partial F_{ML}}{\partial \text{vech}\boldsymbol{\Psi}}$	$\frac{m(m+1)}{2} \times 1$ vector	`dFmlP`	`DSEMobj`, $\boldsymbol{W}_T$
$\frac{\partial F_{ML}}{\partial \text{vech}\boldsymbol{\Theta}}$	$\frac{r(r+1)}{2} \times 1$ vector	`dFmlT`	`DSEMobj`, $\boldsymbol{W}_T$
ΔF_{ML}	$m(jm+r) + \frac{m(m+1)+r(r+1)}{2}$	`ScoreFml`	`DSEMobj`, $\boldsymbol{W}_T$

4.3.1 Estimation of parameters

In the previous sections we gave the S/R functions for the likelihood (objective) function and analytic derivatives of the DSEM model, which take data argument and model specification arguments and are the functions of the parameter vector $\boldsymbol{\theta}$ only. This enables us to use the native S or R optimisation routines such as `nlminb` and `optim`.[8] The `nlminb` function allows box constrains on the parameters, which facilitates convergence and allows restrictions on the parameters. The function call has the form

```
nlminb(start=w0,makeFML,data,m,r,ss,tt,
```

[8]See Venables and Ripley (2002) for discussion of the alternative optimisers in the S-Plus and R environments.

```
      gradient=ScoreFml,lower=-Inf,upper=Inf)
```

While `nlminb` does not compute the Hessian matrix, this can be done with the `vcov.nlminb` function provided by Venables and Ripley (2002) in the MASS library. An alternative that is available both in S-Plus and R is `optim`, which allows a choice of several optimisation algorithms and computes numerical Hessian matrix.[9] The `optim` has several different optimisation algorithms, out of which the "L-BFGS-B" algorithm allows box constraints. Hence `optim` with L-BFGS-B algorithm is similar to the `nlminb` algorithm. The advantage of the `optim` is in computing Hessian along with the parameter estimates in a single call. The application to DSEM models requires the following form

```
optim(par=w0,fn=makeFML,data,m,r,ss,tt,
      gr=ScoreTest,method = c("L-BFGS-B"), hessian=T,
      lower=-Inf,upper=Inf)
```

Note that both optimisers have infinite box constraints as default. It is simple to change them by specifying `lower` and `upper` bounds as numerical vectors of length equal to the number of elements in $\boldsymbol{\theta}$.

We can illustrate the specification of particular DSEM models using the above optimisation routines on the following example. Consider the a DSEM model for a time series data with $T = 50$ where the structural part given by

$$\begin{aligned}
\begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} &= \underbrace{\begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} \\ \beta_{21}^{(1)} & \beta_{22}^{(1)} \end{pmatrix}}_{\boldsymbol{B}_1} \begin{pmatrix} \eta_{t-1}^{(1)} \\ \eta_{t-1}^{(2)} \end{pmatrix} + \underbrace{\begin{pmatrix} \beta_{11}^{(2)} & \beta_{12}^{(2)} \\ \beta_{21}^{(2)} & \beta_{22}^{(2)} \end{pmatrix}}_{\boldsymbol{B}_2} \begin{pmatrix} \eta_{t-2}^{(1)} \\ \eta_{t-2}^{(2)} \end{pmatrix} \\
&+ \underbrace{\begin{pmatrix} \gamma_{11}^{(1)} \\ \gamma_{21}^{(1)} \end{pmatrix}}_{\boldsymbol{\Gamma}_1} \xi_{t-1} + \underbrace{\begin{pmatrix} \gamma_{11}^{(2)} \\ \gamma_{21}^{(2)} \end{pmatrix}}_{\boldsymbol{\Gamma}_2} \xi_{t-2} + \begin{pmatrix} \zeta_t^{(1)} \\ \zeta_t^{(1)} \end{pmatrix},
\end{aligned}$$

with

$$\xi_t = \underbrace{\rho_{11}^{(1)}}_{\boldsymbol{R}_1} \xi_{t-1} + \underbrace{\rho_{11}^{(2)}}_{\boldsymbol{R}_2} \xi_{t-2} + \upsilon_t^{(1)}.$$

Therefore the coefficient matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are

$$\boldsymbol{C}_1 = \begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} & \gamma_{11}^{(1)} \\ \beta_{21}^{(1)} & \beta_{22}^{(1)} & \gamma_{21}^{(1)} \\ 0 & 0 & \rho_{11}^{(1)} \end{pmatrix}, \quad \boldsymbol{C}_2 = \begin{pmatrix} \beta_{11}^{(2)} & \beta_{12}^{(2)} & \gamma_{11}^{(2)} \\ \beta_{21}^{(2)} & \beta_{22}^{(2)} & \gamma_{21}^{(2)} \\ 0 & 0 & \rho_{11}^{(2)} \end{pmatrix}.$$

[9]Note that `optim` is part of the MASS library.

Suppose the measurement model is specified as

$$\begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \\ y_t^{(6)} \\ y_t^{(7)} \\ y_t^{(8)} \\ x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \\ x_t^{(4)} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ \lambda_{21}^{(y)} & 0 & 0 \\ \lambda_{31}^{(y)} & 0 & 0 \\ \lambda_{41}^{(y)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{62}^{(y)} & 0 \\ 0 & \lambda_{72}^{(y)} & 0 \\ 0 & \lambda_{82}^{(y)} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{10,3}^{(x)} \\ 0 & 0 & \lambda_{11,3}^{(x)} \\ 0 & 0 & \lambda_{12,3}^{(x)} \end{pmatrix}}_{\boldsymbol{\Lambda}} \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \\ \xi_t^{(1)} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \varepsilon_t^{(3)} \\ \varepsilon_t^{(4)} \\ \varepsilon_t^{(5)} \\ \varepsilon_t^{(6)} \\ \varepsilon_t^{(7)} \\ \varepsilon_t^{(8)} \\ \delta_t^{(1)} \\ \delta_t^{(2)} \\ \delta_t^{(3)} \\ \delta_t^{(4)} \end{pmatrix},$$

with the residual covariance matrices

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{11}^{(\zeta\zeta)} & \psi_{12}^{(\zeta\zeta)} & \psi_{13}^{(\zeta\upsilon)} \\ \psi_{21}^{(\zeta\zeta)} & \psi_{22}^{(\zeta\zeta)} & \psi_{23}^{(\zeta\upsilon)} \\ \psi_{31}^{(\upsilon\zeta)} & \psi_{32}^{(\upsilon\zeta)} & \psi_{33}^{(\upsilon\upsilon)} \end{pmatrix}, \quad \boldsymbol{\Theta} = \begin{pmatrix} \theta_{11}^{(\varepsilon)} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_{88}^{(\varepsilon)} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \theta_{11}^{(\delta)} & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \theta_{44}^{(\delta)} \end{pmatrix}.$$

The above model has a 72-dimensional parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ where there are 26 free and 46 zero-restricted parameters. Thus we have

$$\begin{aligned}
\boldsymbol{\theta}_1 &= (\mathrm{vec}\boldsymbol{C}_1, \mathrm{vec}\boldsymbol{C}_2) = \Big(\beta_{11}^{(1)}, \beta_{21}^{(1)}, 0, \beta_{12}^{(1)}, \beta_{22}^{(1)}, 0, \gamma_{11}^{(1)}, \gamma_{21}^{(1)}, \rho_{11}^{(1)}, \\
&\qquad \beta_{11}^{(2)}, \beta_{21}^{(2)}, 0, \beta_{12}^{(2)}, \beta_{22}^{(2)}, 0, \gamma_{11}^{(2)}, \gamma_{21}^{(2)}, \rho_{11}^{(2)}\Big) \\
\boldsymbol{\theta}_2 &= (\mathrm{vec}\boldsymbol{\Lambda}) = \Big(1, \lambda_{21}^{(y)}, \lambda_{31}^{(y)}, \lambda_{31}^{(y)}, 0, 0, 0, 0, 0, 0, 0, 0, \\
&\qquad 0, 0, 0, 0, 1, \lambda_{62}^{(y)}, \lambda_{72}^{(y)}, \lambda_{82}^{(y)}, 0, 0, 0, 0, \\
&\qquad 0, 0, 0, 0, 0, 0, 0, 0, 1, \lambda_{10,3}^{(x)}, \lambda_{11,3}^{(x)}, \lambda_{12,3}^{(x)}\Big) \\
\boldsymbol{\theta}_4 &= (\mathrm{vec}\boldsymbol{\Psi}) = \Big(\psi_{11}^{(\zeta\zeta)}, \psi_{21}^{(\zeta\zeta)}, \psi_{31}^{(\upsilon\zeta)}, \psi_{22}^{(\zeta\zeta)}, \psi_{32}^{(\upsilon\zeta)}, \psi_{33}^{(\upsilon\upsilon)}\Big) \\
\boldsymbol{\theta}_5 &= (\mathrm{vec}\boldsymbol{\Theta}) = \Big(\theta_{11}^{(\varepsilon)}, \theta_{22}^{(\varepsilon)}, \theta_{33}^{(\varepsilon)}, \theta_{44}^{(\varepsilon)}, \theta_{55}^{(\varepsilon)}, \theta_{66}^{(\varepsilon)}, \theta_{77}^{(\varepsilon)}, \theta_{88}^{(\varepsilon)}, \\
&\qquad \theta_{11}^{(\delta)}, \theta_{22}^{(\delta)}, \theta_{33}^{(\delta)}, \theta_{44}^{(\delta)}\Big)
\end{aligned}$$

We can initialise the parameters by generating a random (e.g. uniform) 72-dimensional vector. The restricted parameters can be bound using equal upper and lower box constraints (e.g. 0 or 1). Suppose we wish to impose -1 and 2 as upper and lower constraints on the free parameters, respectively. Note we have 2 latent endogenous and 1 latent exogenous variables, hence we have m=3 and also we have r=12 since there are 8 + 4 observable variables in the model. Maximum lag is ss=2 and the length of time series is tt=50. Hence, the `nlminb` call to estimate such model would be

```
w0=runif(72)
nlminb(start=w0,makeFML,data=data,m=3,r=12,ss=2,tt=50,gradient=ScoreFml,
       lower=c(-1,-1, 0,-1,-1, 0,-1,-1,-1,-1,-1,0,-1,-1,0,-1,-1,-1,
               1,-1,-1,-1, 0, 0, 0, 0, 0, 0, 0,0,
               0,0,0,0, 1,-1,-1,-1, 0, 0, 0, 0,
               0,0,0,0, 0, 0, 0, 0, 1,-1,-1,-1,
               0,0,0,0, 0, 0,
               0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0 )
       upper=c( 2, 2, 0, 2, 2, 0, 2, 2, 2, 2, 2,0, 2, 2,0, 2, 2, 2,
               1, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0,0,
               0,0,0,0,1,2,2,2,0,0,0,0,
               0,0,0,0,0,0,0,0,1,2,2,2,
               0,0,0,0,0,0,
               0,0,0,0,0,0,0,0,0,0,0,0 )
```

Aside of using full-sample ML approach for estimation, the suggested approach is considerably simpler in specification of multivariate DSEM models then the alternative approaches. For instance, the use of `SsfPack` to specify the above DSEM model would require programming a particular function that will cast the model into the state-space form and derivation of the analytical derivatives (should we wish to use them) for each specific model. On the other hand, the above considered approach uses the general expressions for the likelihood and derivatives of the DSEM models, which allows simple specification of specific models by only defining the model dimensions (i.e., number of observable and unobservable variables and lag length) and specifying upper and lower bounds (including fixing constraints) on the coefficient vector.

4.3.2 A numerical example

As a simple numerical illustration, consider estimation of a bivariate DSEM asset price model where each of the two market factors is modelled with three different observable asset returns. Such model can be specified as

$$\begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} = \begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} \\ 0 & \beta_{22}^{(1)} \end{pmatrix} \begin{pmatrix} \eta_{t-1}^{(1)} \\ \eta_{t-1}^{(2)} \end{pmatrix} + \begin{pmatrix} \zeta_t^{(1)} \\ \zeta_t^{(1)} \end{pmatrix},$$

$$\begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \\ y_t^{(6)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_{21}^{(y)} & 0 \\ \lambda_{31}^{(y)} & 0 \\ 0 & 1 \\ 0 & \lambda_{52}^{(y)} \\ 0 & \lambda_{62}^{(y)} \end{pmatrix} \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \\ \varepsilon_t^{(3)} \\ \varepsilon_t^{(4)} \\ \varepsilon_t^{(5)} \\ \varepsilon_t^{(6)} \end{pmatrix},$$

$$\mathbf{\Psi} = \begin{pmatrix} \psi_{11}^{(\zeta\zeta)} & \psi_{12}^{(\zeta\zeta)} \\ \psi_{21}^{(\zeta\zeta)} & \psi_{22}^{(\zeta\zeta)} \end{pmatrix}, \quad \mathbf{\Theta} = \begin{pmatrix} \theta_{11}^{(\varepsilon)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_{66}^{(\varepsilon)} \end{pmatrix}.$$

Estimation of this model using daily data from 22 October 2001 to 30 November 2001 ($T = 30$) on the Dow Jones asset returns (in this example we use tickers MMM, AA, MO, AXP, AIG,and BA), took 43 iterations with the nlminb call

```
nlminb(start=rep(0,44),makeFML,data=Wm,m=2,r=6,ss=1,tt=30,
        gradient=ScoreFml,
        lower=c(0,0,0,0,-1,0,-1,-1, 1,-1,-1,0,0,0,0,0,0,1,-1,-1,
                rep(0,3),rep(0,21)),
        upper=c(0,0,0,0,2,0, 2, 2, 3, 2, 2,0,0,0,0,0,0,1, 2, 2,
                rep(5,3),5,rep(0,5),1,rep(0,4),5,rep(0,3),5,
                rep(0,2),5,0,5))
```

producing the following coefficient estimates

$$\begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} \\ 0 & \beta_{22}^{(1)} \end{pmatrix} = \begin{pmatrix} -0.70 & -0.21 \\ 0 & 0.03 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 \\ \lambda_{21}^{(y)} & 0 \\ \lambda_{31}^{(y)} & 0 \\ 0 & 1 \\ 0 & \lambda_{52}^{(y)} \\ 0 & \lambda_{62}^{(y)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.66 & 0 \\ 0.86 & 0 \\ 0 & 1 \\ 0 & 0.71 \\ 0 & 0.31 \end{pmatrix},$$

$$\mathbf{\Psi} = \begin{pmatrix} 1.64 & 0.61 \\ 0.61 & 1.27 \end{pmatrix}, \quad \mathbf{\Theta} = diag(0.53, 1.56, 0.40, 1.69, 1.82, 1.37).$$

The above estimation took approximately 5 minutes on a 2GHz machine, while a similar call to optim took over 20 minutes. A gain in efficiency could be achieved by re-writing the pure S code in C++ or Fortran and wrap it into a .Call or .Fortran interfaces as e.g. S.dll compiled code. In addition, the nlminb and optim routines are fairly modest in terms of optimisation capabilities and would need to be replaced by more powerful optimisers. A possible approach would be to use the S+NuOpt optimiser, which is a special optimisation module for S-Plus, though a similar industrial-strength routine is currently not available in the R environment.

With these modifications it might be possible to move closer to a commercial-capacity package for estimation of DSEM models. It is unlikely that further theoretical development, e.g., derivation of the Hessian matrix, would improve the efficiency of the software implementation, leaving the feasibility issues largely in the domain of using compiled vs. run-time code and more powerful optimisation algorithms.

4.3.3 Conclusion

We have considered a DSEM model for pure time series data and proposed estimation methods based on the closed form likelihood function for the entire sample and the analytical derivatives. Our approach was to obtain the required analytical formulae and demonstrate how these can be programmed in the S language. With S/R functions that compute the likelihood and analytical derivatives we can use the readily available optimisation routines in the S-Plus and R environments for estimation of the unknown coefficients.

We provided S/R functions written in pure S code that can be used for evaluation of the likelihood and the score vector. The likelihood evaluation can be done in a single step, i.e., non-recursively, which can be contrasted with the recursive evaluation using the Kalman filter. In additional we outline an approach giving a simple numerical example of how DSEM models can be estimated using such functions. However, our aims were to suggest an approach that can be used in programming estimation algorithms noting that development of a commercial-strength software package is well beyond the scope of the present work.

Chapter 5

Instrumental variables estimation

5.1 Introduction

In this chapter we propose non-parametric instrumental variables (IV) methods for estimation of DSEM models suitable for both pure time series and panel data. There is no requirement that $N > T$. We consider generalised instrumental variables (GIVE) and full information instrumental variable (FIVE) methods for estimation of DSEM models in the "observed form", i.e., as errors-in-variable models with composite error terms.

These methods are specific in terms of model specification and choice of instruments, which are here interdependent. Namely, we specify the latent variable model as a DSEM model in which measurement errors need to satisfy certain statistical criteria. These criteria are similar to those in the classical factor analysis and are based on the validity of observable indicators as measures of the unobservable (latent) variables. Valid measurement models should have uncorrelated measurement errors, which can be generalised to the time series context by further requiring zero lagged covariances of the measurement errors. We show that basic specification of the DSEM model implies lags of the observable indicators as potentially valid instruments. Empirical validity of such instruments can be tested using standard validity of instruments tests.

Instrumental variables methods have a well known advantage of not imposing any distributional assumptions on the data. The IV methods also provide non-iterative estimators that are very easy to compute using standard general purpose statistical software.

An additional important purpose of these methods is in obtaining good starting values for maximum likelihood estimation using standard SEM software such as LISREL.

5.2 Generalised instrumental variables (GIVE)

Consider the DSEM model (2.1)–(2.3) in t-notation, with added intercept terms,

$$\boldsymbol{\eta}_t = \boldsymbol{\alpha}_\eta + \sum_{j=0}^{p} \mathbf{B}_j \boldsymbol{\eta}_{t-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{\zeta}_t, \tag{5.1}$$

$$\mathbf{x}_t = \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_t + \boldsymbol{\delta}_t, \tag{5.2}$$

$$\mathbf{y}_t = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t. \tag{5.3}$$

Since instrumental variables methods can easily estimate intercepts we include them in the measurement and structural equations (5.1)–(5.3). Equivalently, we can consider zero-intercept models with variables measured in mean-deviation form.

We can re-write the measurement models for $\mathbf{x}_t$ and $\mathbf{y}_t$ as

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha}_2^{(x)} \end{pmatrix} + \begin{pmatrix} \mathbf{I} \\ \boldsymbol{\Lambda}_2^{(x)} \end{pmatrix} \boldsymbol{\xi}_t + \begin{pmatrix} \boldsymbol{\delta}_{1t} \\ \boldsymbol{\delta}_{2t} \end{pmatrix} \tag{5.4}$$

and

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha}_2^{(y)} \end{pmatrix} + \begin{pmatrix} \mathbf{I} \\ \boldsymbol{\Lambda}_2^{(y)} \end{pmatrix} \boldsymbol{\eta}_t + \begin{pmatrix} \boldsymbol{\varepsilon}_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{pmatrix} \tag{5.5}$$

Note that the observed indicators with unit loadings were placed in the top part of the vectors for $\mathbf{x}_t$ and $\mathbf{y}_t$ and thus the upper part of the lambda matrix is an identity matrix. Having divided $\mathbf{x}_t$ into $\mathbf{x}_{t1}$ and $\mathbf{x}_{t2}$, note that for $\mathbf{x}_{t1}$ it holds that

$$\mathbf{x}_{1t} = \boldsymbol{\xi}_t + \boldsymbol{\delta}_{1t} \Rightarrow \boldsymbol{\xi}_t = \mathbf{x}_{1t} - \boldsymbol{\delta}_{1t} \tag{5.6}$$

and, similarly, for $\mathbf{y}_{t1}$ we can replace the latent variable with its unit-loading indicators

$$\mathbf{y}_{1t} = \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_{1t} \Rightarrow \boldsymbol{\eta}_t = \mathbf{y}_{1t} - \boldsymbol{\varepsilon}_{1t} \tag{5.7}$$

It is now possible to use the relations in (5.6) and (5.7) to re-write the measurement model for $\mathbf{x}_t$ as

$$\begin{aligned} \mathbf{x}_{2t} &= \boldsymbol{\alpha}_2^{(x)} + \boldsymbol{\Lambda}_2^{(x)} (\mathbf{x}_{1t} - \boldsymbol{\delta}_{1t}) + \boldsymbol{\delta}_{2t} \\ &= \boldsymbol{\alpha}_2^{(x)} + \boldsymbol{\Lambda}_2^{(x)} \mathbf{x}_{1t} + \left(\boldsymbol{\delta}_{2t} - \boldsymbol{\Lambda}_2^{(x)} \boldsymbol{\delta}_{1t} \right) \end{aligned} \tag{5.8}$$

and for $\mathbf{y}_t$ as

$$\begin{aligned} \mathbf{y}_{2t} &= \boldsymbol{\alpha}_2^{(y)} + \boldsymbol{\Lambda}_2^{(y)} (\mathbf{y}_{1t} - \boldsymbol{\varepsilon}_{1t}) + \boldsymbol{\varepsilon}_{2t} \\ &= \boldsymbol{\alpha}_2^{(y)} + \boldsymbol{\Lambda}_2^{(y)} \mathbf{y}_{1t} + \left(\boldsymbol{\varepsilon}_{2t} - \boldsymbol{\Lambda}_2^{(y)} \boldsymbol{\varepsilon}_{1t} \right) \end{aligned} \tag{5.9}$$

Following the same principle it is possible to re-write the structural part of the model using definitions (5.6) and (5.7) as follows

$$\mathbf{y}_{1t} - \boldsymbol{\varepsilon}_{1t} = \boldsymbol{\alpha}_\eta + \sum_{j=0}^{p} \mathbf{B}_j(\mathbf{y}_{1t-j} - \boldsymbol{\varepsilon}_{1t-j}) + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j(\mathbf{x}_{1t-j} - \boldsymbol{\delta}_{1t-j}) + \boldsymbol{\zeta}_t. \tag{5.10}$$

Separating the observed part of the model from the latent errors we obtain

$$\mathbf{y}_{1t} = \boldsymbol{\alpha}_\eta + \sum_{j=0}^{p} \mathbf{B}_j\mathbf{y}_{1t-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j\mathbf{x}_{1t-j} + \left(\boldsymbol{\zeta}_t + \boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p} \mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q} \boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j} \right), \tag{5.11}$$

with the measurement model for the latent endogenous variables

$$\mathbf{y}_{2t} = \boldsymbol{\alpha}_2^{(y)} + \boldsymbol{\Lambda}_2^{(y)}\mathbf{y}_{1t} + \left(\boldsymbol{\varepsilon}_{2t} - \boldsymbol{\Lambda}_2^{(y)}\boldsymbol{\varepsilon}_{1t}\right), \tag{5.12}$$

and for the latent exogenous variables

$$\mathbf{x}_{2t} = \boldsymbol{\alpha}_2^{(x)} + \boldsymbol{\Lambda}_2^{(x)}\mathbf{x}_{1t} + \left(\boldsymbol{\delta}_{2t} - \boldsymbol{\Lambda}_2^{(x)}\boldsymbol{\delta}_{1t}\right). \tag{5.13}$$

Aside of the specific structure of the latent error terms, (5.11)–(5.13) present a classical structural equation system with observed variables. However, the OF form of the DSEM model differs from the standard econometric simultaneous equation system in respect to the exogeneity status of the OF variables, which are generally observable indicators of the latent variables.

It can be shown that estimation of the OF equations might be possible by the use of the instrumental variable (IV) methods. Furthermore, it can be shown that IV estimation might be based on model-implied instruments in the form of various lags of the OF variables.

We propose a limited information generalised IV (GIVE) technique for consistent estimation of the OF equations by using the model-implied instruments in the form of the lagged indicators of the latent variables.

5.2.1 Full-sample specification

Estimation of the OF equations aims at consistent and, possibly, efficient estimation of the structural and measurement-model parameters. However, the structural (latent) errors cannot be directly estimated. Therefore, ignoring the specific structure of the measurement error terms, let $\mathbf{u}_{1t} \equiv \boldsymbol{\zeta}_t + \boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p} \mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q} \boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j}$, $\mathbf{u}_{2t} \equiv \boldsymbol{\varepsilon}_{2t} - \boldsymbol{\Lambda}_2^{(y)}\boldsymbol{\varepsilon}_{1t}$, and $\mathbf{u}_{3t} \equiv \boldsymbol{\delta}_{2t} - \boldsymbol{\Lambda}_2^{(x)}\boldsymbol{\delta}_{1t}$ the structural OF equations can be written as

$$\mathbf{y}_{1t} = \boldsymbol{\alpha}_\eta + \sum_{j=0}^{p} \mathbf{B}_j\mathbf{y}_{1t-j} + \sum_{j=0}^{q} \boldsymbol{\Gamma}_j\mathbf{x}_{1t-j} + \mathbf{u}_{1t}, \tag{5.14}$$

with the measurement models

$$\mathbf{y}_{2t} = \boldsymbol{\alpha}_2^{(\mathrm{y})} + \boldsymbol{\Lambda}_2^{(\mathrm{y})}\mathbf{y}_{1t} + \mathbf{u}_{2t}, \tag{5.15}$$

and

$$\mathbf{x}_{2t} = \boldsymbol{\alpha}_2^{(x)} + \boldsymbol{\Lambda}_2^{(x)}\mathbf{x}_{1t} + \mathbf{u}_{3t}. \tag{5.16}$$

For notational convenience, we switch to full-sample notation, assuming that a $\max(p,q)$ pre-sample observations are available for estimation. Define $\mathbf{y}_{kj} \equiv \left(y_0^{(kj)}, y_1^{(kj)}, \ldots, y_T^{(kj)}\right)$, and $\mathbf{x}_{2j} \equiv \left(x_0^{(2j)}, x_1^{(2j)}, \ldots, x_T^{(2j)}\right)$, for $k = 1,2$ where the " j " subscript refers to the j^{th} equation where there are m individual $\mathbf{y}_1$ equations, n individual $\mathbf{y}_2$ equations, and h individual $\mathbf{x}_2$ equations. Further define $\mathbf{Y}_{1j} \equiv (\mathbf{Y}_{1jt}, \mathbf{Y}_{1jt-k})$, and $\mathbf{X}_{1j} \equiv (\mathbf{X}_{1jt}, \mathbf{X}_{1jt-k})$, where

$$\mathbf{Y}_{1jt} \equiv \begin{pmatrix} y_0^{(11)} & y_0^{(12)} & \cdots & y_0^{(1m)} \\ y_1^{(11)} & y_1^{(12)} & \cdots & y_1^{(1m)} \\ y_2^{(11)} & y_2^{(12)} & \cdots & y_2^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ y_T^{(11)} & y_T^{(12)} & \cdots & y_T^{(1m)} \end{pmatrix}, \quad \mathbf{X}_{1jt} \equiv \begin{pmatrix} x_0^{(11)} & x_0^{(12)} & \cdots & x_0^{(1m)} \\ x_1^{(11)} & x_1^{(12)} & \cdots & x_1^{(1m)} \\ x_2^{(11)} & x_2^{(12)} & \cdots & x_2^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_T^{(11)} & x_T^{(12)} & \cdots & x_T^{(1m)} \end{pmatrix},$$

and

$$\mathbf{Y}_{1jt-k} \equiv \begin{pmatrix} y_{-1}^{(11)} & y_{-1}^{(12)} & \cdots & y_{-1}^{(1m)} & \cdots & y_{-p}^{(11)} & y_{-p}^{(12)} & \cdots & y_{-p}^{(1m)} \\ y_0^{(11)} & y_0^{(12)} & \cdots & y_0^{(1m)} & \cdots & y_{1-p}^{(11)} & y_{1-p}^{(12)} & \cdots & y_{1-p}^{(1m)} \\ y_2^{(11)} & y_1^{(12)} & \cdots & y_1^{(1m)} & \cdots & y_{2-p}^{(11)} & y_{2-p}^{(12)} & \cdots & y_{2-p}^{(1m)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{T-1}^{(11)} & y_{T-1}^{(12)} & \cdots & y_{T-1}^{(1m)} & \cdots & y_{T-p}^{(11)} & y_{T-p}^{(12)} & \cdots & y_{T-p}^{(1m)} \end{pmatrix},$$

$$\mathbf{X}_{1jt-k} \equiv \begin{pmatrix} x_{-1}^{(11)} & x_{-1}^{(12)} & \cdots & x_{-1}^{(1g)} & \cdots & x_{-q}^{(11)} & x_{-q}^{(12)} & \cdots & x_{-q}^{(1g)} \\ x_0^{(11)} & x_0^{(12)} & \cdots & x_0^{(1g)} & \cdots & x_{1-q}^{(11)} & x_{1-q}^{(12)} & \cdots & x_{1-q}^{(1g)} \\ x_2^{(11)} & x_1^{(12)} & \cdots & x_1^{(1g)} & \cdots & x_{2-q}^{(11)} & x_{2-q}^{(12)} & \cdots & x_{2-q}^{(1g)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{T-1}^{(11)} & x_{T-1}^{(12)} & \cdots & x_{T-1}^{(1g)} & \cdots & x_{T-q}^{(11)} & x_{T-q}^{(12)} & \cdots & x_{T-q}^{(1g)} \end{pmatrix}.$$

In addition, we define the following notation for the parameter vectors

$$\boldsymbol{\lambda}_j^{(y)} \equiv \left(\lambda_{yj}^{(21)}, \lambda_{yj}^{(22)}, \ldots, \lambda_{yj}^{(2n)}\right)', \quad \lambda_j^{(x)} \equiv \left(\lambda_{xj}^{(21)}, \lambda_{xj}^{(22)}, \ldots, \lambda_{xj}^{(2h)}\right)',$$

$$\boldsymbol{\beta}_j \equiv \left(\beta_0^{(11)}, \beta_0^{(12)}, \ldots, \beta_0^{(1m)}, \beta_1^{(11)}, \beta_1^{(12)}, \ldots, \beta_1^{(1m)}, \ldots, \beta_p^{(11)}, \beta_p^{(12)}, \ldots, \beta_p^{(1m)}\right)',$$

and

$$\boldsymbol{\gamma}_j \equiv \left(\gamma_0^{(11)}, \gamma_0^{(12)}, \ldots, \gamma_0^{(1g)}, \gamma_1^{(11)}, \gamma_1^{(12)}, \ldots, \gamma_1^{(1g)}, \ldots, \gamma_q^{(11)}, \gamma_q^{(12)}, \ldots, \gamma_q^{(1g)}\right)'.$$

Using the above notation, we can now write the (5.14)–(5.16) as

$$\mathbf{y}_{1j} = \boldsymbol{\alpha}_{1j}^{(y)} + \mathbf{Y}_{1j}\boldsymbol{\beta}_j + \mathbf{X}_{1j}\boldsymbol{\gamma}_j + \mathbf{u}_{1j}, \tag{5.17}$$

$$\mathbf{y}_{2j} = \boldsymbol{\alpha}_{2j}^{(y)} + \mathbf{Y}_{1jt}\boldsymbol{\lambda}_j^{(y)} + \mathbf{u}_{2j}, \tag{5.18}$$

$$\mathbf{x}_{2j} = \boldsymbol{\alpha}_{2j}^{(x)} + \mathbf{X}_{1jt}\boldsymbol{\lambda}_j^{(x)} + \mathbf{u}_{3j}. \tag{5.19}$$

Note that the individual OF equations are specified as

$$y_{1j} = \alpha_{1j}^{(\mathrm{y})} + \sum_{k=1}^{m}\sum_{i=0}^{p} \beta_i^{(1k)} y_{t-i}^{(1k)} + \sum_{k=1}^{g}\sum_{i=0}^{q} \gamma_i^{(1k)} x_{t-i}^{(1k)} + u_{1jt},$$

for the structural part of the model, and as

$$y_{2j} = \alpha_{2j}^{(\mathrm{y})} + \sum_{k=1}^{m} \lambda_{2jk}^{(\mathrm{y})} y_t^{(1k)} + u_{2jt}, \quad x_{2j} = \alpha_{2i}^{(\mathrm{x})} + \sum_{k=1}^{g} \lambda_{2jk}^{(\mathrm{x})} x_{\mathrm{t}}^{(1k)} + u_{3jt},$$

for the measurement models. This completes the specification of the DSEM model.

It remains to show that the available instruments in the form of lags of the observed variables can enable consistent estimation. The issue of the choice of instruments is also discussed in Bollen (1996; 2001), however he does not discuss this issue in the context of dynamic models. The following discussion takes into account the specific structure of the OF system and the implications derived from the composition of the latent errors. This (known) composition of the latent error terms and their implied relation with the observed components of the model, as a consequence of the latent structure, presents the major difference between the DSEM OF equations and classical econometric models. Specifically, it is not possible to simply assume the availability of external instrumental variables that satisfy some general conditions such as being uncorrelated with the errors and correlated with the regressors. Rather, it will be necessary to show under which conditions the lagged modelled variables can serve as valid instruments in the estimation of the OF equations.

5.2.2 Consistency conditions and instrumental variables

The standard consistency conditions needed for the validity of instrumental variables (see e.g. Judge *et al.*, 1985) and Davidson and MacKinnon, 1993) can be stated in terms of the data matrix $\mathbf{X}$ defined as $\mathbf{X} \equiv (\boldsymbol{\iota}, \mathbf{Y}_j, \mathbf{X}_j)$ where $\mathbf{Y}_{1j} \equiv (\mathbf{Y}_{1jt}, \mathbf{Y}_{1jt-k})$ and $\mathbf{X}_{1j} \equiv (\mathbf{X}_{1jt}, \mathbf{X}_{1jt-k})$, as defined above. Let $\mathbf{Z}$ be a matrix of valid instruments defined as $\mathbf{Z} \equiv (\mathbf{Y}_1^*, \mathbf{Y}_2^*, \mathbf{X}_1^*, \mathbf{X}_2^*)$ where $\mathbf{Y}_1^* \equiv (\mathbf{Y}_{11}^*, \mathbf{Y}_{12}^*, \ldots, \mathbf{Y}_{1a}^*)$, $\mathbf{Y}_2^* \equiv (\mathbf{Y}_{21}^*, \mathbf{Y}_{22}^*, \ldots, \mathbf{Y}_{2b}^*)$, $\mathbf{X}_1^* \equiv (\mathbf{X}_{11}^*, \mathbf{X}_{12}^*, \ldots, \mathbf{X}_{1c}^*)$, $\mathbf{X}_2^* \equiv (\mathbf{X}_{21}^*, \mathbf{X}_{22}^*, \ldots, \mathbf{X}_{2d}^*)$, and

$$\mathbf{Y}_{1k}^* = \begin{pmatrix} y_{-p-k}^{(11)} & y_{-p-k}^{(12)} & \cdots & y_{-p-k}^{(1m)} \\ y_{1-p-k}^{(11)} & y_{1-p-k}^{(12)} & \cdots & y_{1-p-k}^{(1m)} \\ y_{2-p-k}^{(11)} & y_{2-p-k}^{(12)} & \cdots & y_{2-p-k}^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-p-k}^{(11)} & y_{T-p-k}^{(12)} & \cdots & y_{T-p-k}^{(1m)} \end{pmatrix}, \quad \mathbf{Y}_{2l}^* = \begin{pmatrix} y_{-l}^{(21)} & y_{-l}^{(22)} & \cdots & y_{-l}^{(2n)} \\ y_{-l+1}^{(21)} & y_{-l+1}^{(22)} & \cdots & y_{-l+1}^{(2n)} \\ y_{-l+2}^{(21)} & y_{-l+2}^{(22)} & \cdots & y_{-l+2}^{(2n)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-l}^{(21)} & y_{T-l}^{(22)} & \cdots & y_{T-l}^{(2n)} \end{pmatrix},$$

$$\mathbf{X}_{1i}^* = \begin{pmatrix} x_{-q-i}^{(11)} & x_{-q-i}^{(12)} & \cdots & x_{-q-i}^{(1m)} \\ x_{1-q-i}^{(11)} & x_{1-q-i}^{(12)} & \cdots & x_{1-q-i}^{(1m)} \\ x_{2-q-i}^{(11)} & x_{2-q-i}^{(12)} & \cdots & x_{2-q-i}^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-q-i}^{(11)} & x_{T-q-i}^{(12)} & \cdots & x_{T-q-i}^{(1m)} \end{pmatrix}, \quad \mathbf{X}_{2j}^* = \begin{pmatrix} x_{-j}^{(21)} & x_{-j}^{(22)} & \cdots & x_{-j}^{(2n)} \\ x_{-j+1}^{(21)} & x_{-j+1}^{(22)} & \cdots & x_{-j+1}^{(2n)} \\ x_{-j+2}^{(21)} & x_{-j+2}^{(22)} & \cdots & x_{-j+2}^{(2n)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-j}^{(21)} & x_{T-j}^{(22)} & \cdots & x_{T-j}^{(2n)} \end{pmatrix},$$

where $k = 1, 2, \ldots, a$; $l = 1, 2, \ldots, b$; $i = 1, 2, \ldots, c$; and $j = 1, 2, \ldots, d$.

We state the general conditions for these instruments in terms of the joint matrices $\mathbf{X}$ and $\mathbf{Z}$ though, in practice, only subsets of these matrices will be used in estimated models. It is generally necessary that

$$\text{plim}\left(T^{-1}\mathbf{Z}'\mathbf{Z}\right) = \lim_{T\to\infty}\left(T^{-1}\mathbf{Z}'\mathbf{Z}\right) = \mathbf{\Sigma}_{ZZ},$$

and also that

$$\text{plim}\left(T^{-1}\mathbf{Z}'\mathbf{X}\right) = \lim_{T\to\infty}\left(T^{-1}\mathbf{Z}'\mathbf{X}\right) = \mathbf{\Sigma}_{ZX},$$

where $\mathbf{\Sigma}_{ZZ}$ and $\mathbf{\Sigma}_{ZX}$ are positive definite matrices. These conditions will generally hold for the case of lagged instruments given they satisfy certain stochastic conditions. In addition, we assume homoscedastic residuals, i.e., $E\left[\mathbf{u}_i\mathbf{u}'_j\right] = \sigma_{ij}\mathbf{I}$ and, specially, $E\left[\mathbf{Z}'\mathbf{u}_i\right] = \mathbf{0}$.

To assure the consistency of the IV estimator we will need to make the following assumption about the stochastic properties of the observed variables.

Assumption 5.2.2.1 *For stochastic processes $\{y_t\}$ and $\{x_t\}$ suppose that:*

A1. $E\left[y_{ijt}\right] = \mu_{ij}^{(y)}, \quad \forall t$

A2. $E\left[x_{ijt}\right] = \mu_{ij}^{(x)}, \quad \forall t$

A3. $E\left[(y_{ij,t-r}-\mu_{ij}^{(y)})(y_{ef,t-w}-\mu_{ef}^{(y)})\right]=\gamma_{|r-w|}^{(ijef)},\quad \forall t$

A4. $E\left[(x_{ij,t-r}-\mu_{ij}^{(x)})(x_{ef,t-w}-\mu_{ef}^{(x)})\right]=\delta_{|r-w|}^{(ijef)},\quad \forall t$

A5. $E\left[(y_{ij,t-r}-\mu_{ij}^{(y)})(x_{ef,t-w}-\mu_{ef}^{(x)})\right]=\psi_{|r-w|}^{(ijef)},\quad \forall t$

A6. $\sum_{k=0}^{\infty}\gamma_k^{(\cdot)}<\infty,\ \sum_{k=0}^{\infty}\delta_k^{(\cdot)}<\infty,\ \sum_{k=0}^{\infty}\psi_k^{(\cdot)}<\infty$

We will also need the following two lemmas.

Lemma 5.2.2.2 *Let w_t be a covariance-stationary process with finite fourth moments and absolutely summable autocovariances. Then the sample mean satisfies $T^{-1}\sum_{t=1}^{T}w_t\overset{m.s.}{\rightarrow}\mu_w$ where m.s. denotes convergence in mean square.*

Proof. Omitted. See Hamilton (1994: 188), Proposition 7.5.

Lemma 5.2.2.3 *Let y_t and x_t be stochastic processes satisfying Assumption (5.2.2.2). Then the following convergence results hold:*

(i) $\frac{1}{T}\sum_{t=0}^{T}y_{ij,t-s}\overset{p}{\rightarrow}E\left[y_{ijt}\right]=\mu_{ij}^{(y)}$

(ii) $\frac{1}{T}\sum_{t=0}^{T}y_{ij,t-s}^{2}\overset{p}{\rightarrow}E\left[y_{ijt}^{2}\right]=\gamma_0^{(ij)}+(\mu_{ij}^{(y)})^2$

(iii) $\frac{1}{T}\sum_{t=0}^{T}y_{ij,t-r}y_{ef,t-w}\overset{p}{\rightarrow}E\left[y_{ij,t-r}y_{ij,t-w}\right]=\gamma_{|r-w|}^{(ijef)}+\mu_{ij}^{(y)}\mu_{ef}^{(y)}$

(vi) $\frac{1}{T}\sum_{t=0}^{T}x_{ij,t-s}\overset{p}{\rightarrow}E\left[x_{ijt}\right]=\mu_{ij}^{(x)}$

(v) $\frac{1}{T}\sum_{t=0}^{T}x_{ij,t-s}^{2}\overset{p}{\rightarrow}E\left[x_{ijt}^{2}\right]=\delta_0^{(ij)}+(\mu_{ij}^{(x)})^2$

(vi) $\frac{1}{T}\sum_{t=0}^{T}x_{ij,t-r}x_{ef,t-w}\overset{p}{\rightarrow}E\left[x_{ij,t-r}x_{ij,t-w}\right]=\delta_{|r-w|}^{(ijef)}+\mu_{ij}^{(x)}\mu_{ef}^{(x)}$

(vii) $\frac{1}{T}\sum_{t=0}^{T}y_{ij,t-r}x_{ef,t-w}\overset{p}{\rightarrow}E\left[y_{ij,t-r}x_{ef,t-w}\right]=\psi_{|r-w|}^{(ijef)}+\mu_{ij}^{(y)}\mu_{ef}^{(x)}$

Proof See Appendix §5A.

Proposition 5.2.2.4 *Let $\mathbf{X}\equiv(\iota,\ \mathbf{Y}_j,\ \mathbf{X}_j)$ where $\mathbf{Y}_{1j}\equiv(\mathbf{Y}_{1jt},\mathbf{Y}_{1jt-k})$ and $\mathbf{X}_{1j}\equiv(\mathbf{X}_{1jt},\mathbf{X}_{1jt-k})$. Let $\mathbf{Z}$ be a matrix of valid instruments defined as $\mathbf{Z}\equiv(\mathbf{Y}_1^*,\ \mathbf{Y}_2^*,\ \mathbf{X}_1^*,\ \mathbf{X}_2^*)$. Assuming that $E\left[\mathbf{u}_i\mathbf{u'}_j\right]=\sigma_{ij}\mathbf{I}$, the following result holds*

(i) $\text{plim}\left(\frac{1}{T}\mathbf{Z'Z}\right)=\mathbf{\Sigma}_{ZZ}$

(ii) $\text{plim}\left(\frac{1}{T}\mathbf{Z'X}\right)=\mathbf{\Sigma}_{ZX}$

(iii) $E\left[\mathbf{Z'u}_i\right]=\mathbf{0}$

Proof See Appendix §5B.

However, using lagged instruments reduces the effective sample size available for estimation, thus we might consider filling the missing observations with the leads (future values) of the observable indicators. The use of leads along with lags in the instrumental variables estimation was suggested by Griliches and Hausman (1986), see also (Wansbeek and Meijer 2000, Wansbeek 2001, Arellano 2003).

In particular, for a DSEM(p, q) model lagged observable indicators $x_{t-s}, x_{t-s-1}, \dots$ and $y_{t-s}, y_{t-s-1}, \dots$ for $s = \max(p, q) + 1$ will be valid instruments. Assuming a causal process, leads or future values of x_t and y_t will also be valid instruments. Generally, the set of valid instruments might include $x_0, x_1, \dots, x_{t-s-1}, x_{t+1}, x_{t+2}, \dots, x_T$ and $y_0, y_1, \dots, y_{t-s-1}, y_{t+1}, y_{t+2}, \dots, y_T$.

Using the operator

$$\boldsymbol{S}^j_{IV} = \boldsymbol{I}_N \otimes \left[\begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_j & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \boldsymbol{S}^j_T \right], \tag{5.20}$$

the instrumental variables can be constructed as combinations of lags and leads stacked together into vectors of instruments $\boldsymbol{x}^{(j)}_{IV}$ using the $\boldsymbol{S}^j_{IV}$ operator as

$$\boldsymbol{x}^{(j)}_{IV} = \boldsymbol{S}^j_{IV} \boldsymbol{x}_T. \tag{5.21}$$

This has the effect of replacing the missing values of the lag-only instruments with the future values (leads) as

$$\begin{pmatrix} x^{i=1}_{t=1} & \boxed{NA} & \boxed{NA} \\ x^{i=1}_{t=2} & x^{i=1}_{t=1} & \boxed{NA} \\ x^{i=1}_{t=3} & x^{i=1}_{t=2} & x^{i=1}_{t=1} \\ x^{i=1}_{t=4} & x^{i=1}_{t=3} & x^{i=1}_{t=2} \\ x^{i=2}_{t=1} & \boxed{NA} & \boxed{NA} \\ x^{i=2}_{t=2} & x^{i=2}_{t=1} & \boxed{NA} \\ x^{i=2}_{t=3} & x^{i=2}_{t=2} & x^{i=2}_{t=1} \\ x^{i=2}_{t=4} & x^{i=2}_{t=3} & x^{i=2}_{t=2} \end{pmatrix} \rightarrow \begin{pmatrix} x^{i=1}_{t=1} & \boxed{x^{i=1}_{t=2}} & \boxed{x^{i=1}_{t=2}} \\ x^{i=1}_{t=2} & x^{i=1}_{t=1} & \boxed{x^{i=1}_{t=3}} \\ x^{i=1}_{t=3} & x^{i=1}_{t=2} & x^{i=1}_{t=1} \\ x^{i=1}_{t=4} & x^{i=1}_{t=3} & x^{i=1}_{t=2} \\ x^{i=2}_{t=1} & \boxed{x^{i=2}_{t=2}} & \boxed{x^{i=2}_{t=2}} \\ x^{i=2}_{t=2} & x^{i=2}_{t=1} & \boxed{x^{i=2}_{t=3}} \\ x^{i=2}_{t=3} & x^{i=2}_{t=2} & x^{i=2}_{t=1} \\ x^{i=2}_{t=4} & x^{i=2}_{t=3} & x^{i=2}_{t=2} \end{pmatrix}$$

The above results allow consistent GIVE estimation of the OF equations using the available, model-implied (lagged) instruments contained in $\mathbf{Z}$, which includes all available eligible instruments that do not come from outside the modelled data. It must be mentioned that nothing precludes availability of valid instruments that are not merely lags of the modelled variables. However, the nature of structural equation models with latent variables casts doubt that such variables will be available. In any case, valid variables will satisfy the same conditions, but we have shown that available instruments already might exist in the used data in forms of lagged values not already included in the model.

5.2.3 Consistent generalised instrumental variable estimation of the OF equations

Formulation and estimation of the OF equations requires reliance on specific structure and status of the modelled variables. This structure is determined by the latent-form specification and makes specification of the OF equations rather complex. In order to derive generalised instrumental variable estimators (GIVE) for the OF equations, we start from the system of equations given in (5.17), (5.18), and (5.19) and write it by positioning its matrix and vector elements in the way that will facilitate the use of more concise notation, i.e.,

$$\begin{aligned} \mathbf{y}_{1j} &= \boldsymbol{\alpha}_{1j}^{(y)} + \mathbf{Y}_{1j}\boldsymbol{\beta}_j + \mathbf{X}_{1j}\boldsymbol{\gamma}_j + \mathbf{u}_{1j} \\ \mathbf{y}_{2j} &= \boldsymbol{\alpha}_{2j}^{(y)} + \mathbf{Y}_{1jt}\boldsymbol{\lambda}_j^{(y)} + \mathbf{u}_{2j} \\ \mathbf{x}_{2j} &= \boldsymbol{\alpha}_{2j}^{(x)} + \mathbf{X}_{1jt}\boldsymbol{\lambda}_j^{(x)} + \mathbf{u}_{3j} \end{aligned} \tag{5.22}$$

We are now able to simplify our notation by stacking all of the right-hand-side variables of each of the three parts of the system (5.22) by making the following definitions: $\mathbf{W}_{1j} \equiv (\boldsymbol{\iota}, \mathbf{Y}_{1j}, \mathbf{X}_{1j})$, $\mathbf{W}_{2j} \equiv (\boldsymbol{\iota}, \mathbf{Y}_{1jt})$, $\mathbf{W}_{3j} \equiv (\boldsymbol{\iota}, \mathbf{X}_{1jt})$, $\boldsymbol{\delta}_{1j}^{(y)} \equiv \left(\boldsymbol{\alpha}_{1j}^{(y)\prime}, \boldsymbol{\beta}_j', \boldsymbol{\gamma}_j'\right)'$, $\boldsymbol{\delta}_{2j}^{(y)} \equiv \left(\boldsymbol{\alpha}_{2j}^{(y)\prime}, \boldsymbol{\lambda}_{2j}^{(y)\prime}\right)'$, and $\boldsymbol{\delta}_{2j}^{(x)} \equiv \left(\boldsymbol{\alpha}_{2j}^{(x)\prime}, \boldsymbol{\lambda}_{2j}^{(x)\prime}\right)'$. It is now possible to re-write the system (5.22) in a simpler, more concise notation as

$$\begin{aligned} \mathbf{y}_{1j} &= \mathbf{W}_{1j}\boldsymbol{\delta}_{1j}^{(y)} + \mathbf{u}_{1j} \\ \mathbf{y}_{2j} &= \mathbf{W}_{2j}\boldsymbol{\delta}_{2j}^{(y)} + \mathbf{u}_{2j} \\ \mathbf{x}_{2j} &= \mathbf{W}_{3j}\boldsymbol{\delta}_{2j}^{(x)} + \mathbf{u}_{3j} \end{aligned} \tag{5.23}$$

An appropriate matrix of instruments $\mathbf{Z}$ need not contain all available eligible instruments, but it needs to have at least as many of them as there are endogenous variables in each equation. The matrix of instruments $\mathbf{Z}$ can differ across different (individual) equations of the system (5.23). For simplicity we assume that $\mathbf{Z}$ is correctly specified.

We proceed in defining the GIVE estimator. First, by premultiplying each part of the system by $\mathbf{Z}$ we obtain matrix equations $\mathbf{Z}'\mathbf{y}_{1j} = \mathbf{Z}'\mathbf{W}_{1j}\boldsymbol{\delta}_{1j}^{(y)} + \mathbf{Z}'\mathbf{u}_{1j}$, $\mathbf{Z}'\mathbf{y}_{2j} = \mathbf{Z}'\mathbf{W}_{2j}\boldsymbol{\delta}_{2j}^{(y)} + \mathbf{Z}'\mathbf{u}_{2j}$, and $\mathbf{Z}'\mathbf{x}_{2j} = \mathbf{Z}'\mathbf{W}_{3j}\boldsymbol{\delta}_{2j}^{(x)} + \mathbf{Z}'\mathbf{u}_{3j}$. We now define usual GIVE estimators for coefficient vectors $\hat{\boldsymbol{\delta}}_{1j}^{(y)}$, $\hat{\boldsymbol{\delta}}_{2j}^{(y)}$, and $\hat{\boldsymbol{\delta}}_{2j}^{(x)}$ as

$$\hat{\boldsymbol{\delta}}_{1j}^{(y)} = \left[\mathbf{W'}_{1j}\,\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{W}_{1j}\right]\mathbf{W'}_{1j}\,\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{y}_{1j}, \tag{5.24}$$

$$\delta_{2j}^{(y)} = \left[\mathbf{W'}_{2j}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'W}_{2j}\right]\mathbf{W'}_{2j}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'y}_{2j}, \tag{5.25}$$

and

$$\delta_{2j}^{(x)} = \left[\mathbf{W'}_{3j}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'W}_{3j}\right]\mathbf{W'}_{3j}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'x}_{2j}. \tag{5.26}$$

It is easy to show that the GIVE estimators given in (5.24), (5.25), and (5.26) are consistent estimators of the unknown coefficient vectors $\boldsymbol{\delta}_{1j}^{(y)}$, $\boldsymbol{\delta}_{2j}^{(y)}$, and $\boldsymbol{\delta}_{2j}^{(x)}$. To show this note that

$$\hat{\boldsymbol{\delta}}_{ij}^{(*)} = \boldsymbol{\delta}_{ij}^{(*)} + \left[\mathbf{W'}_{ij}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'W}_{ij}\right]\mathbf{W'}_{ij}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'u}_{ij}$$

Taking probability limits we obtain

$$\begin{aligned}
\operatorname{plim}\left(\hat{\boldsymbol{\delta}}_{ij}^{(*)}\right) &= \boldsymbol{\delta}_{ij}^{(*)} + \left[\operatorname{plim}\left(\tfrac{1}{T}\mathbf{W'}_{ij}\,\mathbf{Z}\right)\cdot\operatorname{plim}\left(\tfrac{1}{T}\left(\mathbf{Z'Z}\right)^{-1}\right)\operatorname{plim}\left(\tfrac{1}{T}\mathbf{Z'W}_{ij}\right)\right]^{-1} \\
&\quad \times \operatorname{plim}\left(\tfrac{1}{T}\mathbf{W'}_{ij}\,\mathbf{Z}\right)\cdot\operatorname{plim}\left(\tfrac{1}{T}\left(\mathbf{Z'Z}\right)^{-1}\right)\operatorname{plim}\left(\tfrac{1}{T}\mathbf{Z'u}_{ij}\right) \\
&= \boldsymbol{\delta}_{ij}^{(*)} + \left(\boldsymbol{\Sigma}_{W_{ij}Z}\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Sigma}_{ZW_{ij}}\right)^{-1}\boldsymbol{\Sigma}_{W_{ij}Z}\boldsymbol{\Sigma}_{ZZ}^{-1}\cdot\mathbf{0} \\
&= \boldsymbol{\delta}_{ij}^{(*)}
\end{aligned}$$

The above results holds for each of the vectors $\hat{\boldsymbol{\delta}}_{1j}^{(y)}$, $\hat{\boldsymbol{\delta}}_{2j}^{(y)}$, and $\hat{\boldsymbol{\delta}}_{2j}^{(x)}$, where superscripts (y, x) were replaced by asterisks, and subscripts $(1, 2)$ by i. For computational purposes, the GIVE estimators using the OF notation defined above can be written in more detail as follows. Firstly, the three sets of coefficient vectors in the structural part of the model are estimated by

$$\begin{pmatrix} \hat{\alpha}_{\eta j} \\ \hat{\boldsymbol{\beta}}_j \\ \hat{\boldsymbol{\gamma}}_j \end{pmatrix} = \begin{pmatrix} \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}'\boldsymbol{\iota} & \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'Y}_{1j} & \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'X}_{1j} \\ \mathbf{Y'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}'\boldsymbol{\iota} & \mathbf{Y'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'Y}_{1j} & \mathbf{Y'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'X}_{1j} \\ \mathbf{X'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}'\boldsymbol{\iota} & \mathbf{X'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'Y}_{1j} & \mathbf{X'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'X}_{1j} \end{pmatrix}^{-1} \times \begin{pmatrix} \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{1j} \\ \mathbf{Y'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{1j} \\ \mathbf{X'}_{1j}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{1j} \end{pmatrix}.$$

Secondly, the GIVE estimators of the measurement model are given by

$$\begin{pmatrix} \hat{\boldsymbol{\alpha}}_{2j}^{(y)} \\ \boldsymbol{\lambda}_{2j}^{(y)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}'\boldsymbol{\iota} & \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'Y}_{1jt} \\ \mathbf{Y'}_{1jt}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}'\boldsymbol{\iota} & \mathbf{Y'}_{1jt}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'Y}_{1jt} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{2j} \\ \mathbf{Y'}_{1jt}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{2j} \end{pmatrix},$$

and

$$\begin{pmatrix} \hat{\boldsymbol{\alpha}}_{2j}^{(y)} \\ \boldsymbol{\lambda}_{2j}^{(y)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}'\boldsymbol{\iota} & \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'Y}_{1jt} \\ \mathbf{Y'}_{1jt}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}'\boldsymbol{\iota} & \mathbf{Y'}_{1jt}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'Y}_{1jt} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\iota}'\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{2j} \\ \mathbf{Y'}_{1jt}\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{2j} \end{pmatrix}.$$

Asymptotic distribution of these estimators does not depend on the assumption that the modelled data is multivariate normal and, thus, GIVE estimators of the DSEM model are asymptotically distribution free. This is an advantage over the maximum likelihood estimator of the static structural equation model, and therefore, GIVE estimator can prove to be more robust to both misspecification of certain parts of the model and to departure from normality.[1]

The asymptotic distribution of the GIVE estimators is normal and it can be derived by noting that

$$\begin{aligned}\sqrt{T}\left(\hat{\boldsymbol{\delta}}_{ij}^{(*)}-\boldsymbol{\delta}_{ij}^{(*)}\right)&=\left[\left(\tfrac{1}{T}\mathbf{W'}_{ij}\mathbf{Z}\right)\left(\tfrac{1}{T}\left(\mathbf{Z'Z}\right)^{-1}\right)\left(\tfrac{1}{T}\mathbf{Z'W}_{ij}\right)\right]^{-1}\\&\times\left(\tfrac{1}{T}\mathbf{W'}_{ij}\mathbf{Z}\right)\left(\tfrac{1}{T}\left(\mathbf{Z'Z}\right)^{-1}\right)\left(\tfrac{1}{\sqrt{T}}\mathbf{Z'u}_{ij}\right).\end{aligned}$$

If we assume that $T^{-1/2}\mathbf{Z'u}_{ij}\xrightarrow{d}N\left[\mathbf{0},\sigma_{ij}\boldsymbol{\Sigma}_{\mathrm{ZZ}}\right]$, we can conclude that the asymptotic distribution of the DSEM coefficient estimates is

$$\sqrt{T}\left(\hat{\boldsymbol{\delta}}_{ij}^{(*)}-\boldsymbol{\delta}_{ij}^{(*)}\right)\xrightarrow{d}N\left[\mathbf{0},\ \sigma_{ij}\left(\boldsymbol{\Sigma}_{\mathrm{W}_{ij}\mathrm{Z}}\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Sigma}_{\mathrm{ZW}_{ij}}\right)^{-1}\right]$$

The asymptotic covariance matrix $\hat{\sigma}_{ij}\left[\mathbf{W'}_{ij}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'W}_{ij}\right]^{-1}$ can be estimated with $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\delta}}_{ij}^{(*)}}=\hat{\sigma}_{ij}\left[\mathbf{W'}_{ij}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'W}_{ij}\right]^{-1}$ where

$$\hat{\sigma}_{ij}=T^{-1}\hat{\mathbf{u}}_{ij}'\hat{\mathbf{u}}_{ij}=T^{-1}\left(\mathbf{y}_{ij}-\mathbf{W}_{ij}\hat{\boldsymbol{\delta}}_{ij}^{(*)}\right)'\left(\mathbf{y}_{ij}-\mathbf{W}_{ij}\hat{\boldsymbol{\delta}}_{ij}^{(*)}\right).$$

The empirical validity of instrumental variables, as opposite to their model-implied eligibility, is empirically testable. The validity of the choice of the instrumental variables can be tested by the Sargan (1988) χ^2 test. Applied to the OF equations, the Sargan test can be calculated as

$$\frac{\mathbf{y'}_{ij}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'y}_{ij}-\hat{\boldsymbol{\delta}}_{ij}^{(*)\prime}\left[\mathbf{W'}_{ij}\mathbf{Z}\left(\mathbf{Z'Z}\right)^{-1}\mathbf{Z'W}_{ij}\right]\hat{\boldsymbol{\delta}}_{ij}^{(*)}}{T^{-1}\hat{\mathbf{u}}_{ij}'\hat{\mathbf{u}}_{ij}}\underset{app}{\sim}\chi^2_{(d)},\tag{5.27}$$

where d is the number of over-identifying instruments, assumed to be independent of the equation error. It is important to note that selection of the IV's on the basis of the model-implied eligibility without testing for their empirical validity can result in considerable bias in the estimated coefficients. As the choice of instruments affects consistency of GIVE estimates, inappropriate IV selection might result in estimates that will not be robust to misspecification. Therefore, testing for the validity of IV's should be an important part in empirical estimation of DSEM models.

[1]Misspecification of one OF equation will not necessarily affect coefficients of other equations since these are estimated separately using a limited information estimator

5.3 Full information estimation (FIVE)

While consistent and possibly robust to certain forms of mis-specification, the GIVE estimator is not necessarily asymptotically efficient. A full-information instrumental variables (FIVE) efficient estimator for classical simultaneous equation systems is developed by Zellner and Theil (1962). The requirement for the use of the FIVE estimator we propose for the DSEM model is to have a common matrix of instruments ($\mathbf{Z}$) that can be applied to all equations in the OF system. Such matrix of instruments for dynamic models might be made out of lagged modelled variables that are eligible for all equations in the system.

FIVE is a systems estimator and can be obtained by staking all equations (premultiplied by the common IV matrix $\mathbf{Z}$) in a single matrix equation

$$\begin{pmatrix} \mathbf{Z}'\mathbf{y}_{11} \\ \vdots \\ \mathbf{Z}'\mathbf{y}_{1m} \\ \mathbf{Z}'\mathbf{y}_{21} \\ \vdots \\ \mathbf{Z}'\mathbf{y}_{2n} \\ \mathbf{Z}'\mathbf{x}_{21} \\ \vdots \\ \mathbf{Z}'\mathbf{x}_{2h} \end{pmatrix} = \mathrm{diag}\begin{pmatrix} (\mathbf{Z}'\mathbf{W}_{11})' \\ \vdots \\ (\mathbf{Z}'\mathbf{W}_{1m})' \\ (\mathbf{Z}'\mathbf{W}_{21})' \\ \vdots \\ (\mathbf{Z}'\mathbf{W}_{2n})' \\ (\mathbf{Z}'\mathbf{W}_{31})' \\ \vdots \\ (\mathbf{Z}'\mathbf{W}_{3h})' \end{pmatrix}' \begin{pmatrix} \delta_{11}^{(y)} \\ \vdots \\ \delta_{1m}^{(y)} \\ \delta_{21}^{(y)} \\ \vdots \\ \delta_{2n}^{(y)} \\ \delta_{21}^{(x)} \\ \vdots \\ \delta_{2h}^{(x)} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}'\mathbf{u}_{11}^{(\mathrm{y})} \\ \vdots \\ \mathbf{Z}'\mathbf{u}_{1m}^{(\mathrm{y})} \\ \mathbf{Z}'\mathbf{u}_{21}^{(\mathrm{y})} \\ \vdots \\ \mathbf{Z}'\mathbf{u}_{2n}^{(\mathrm{y})} \\ \mathbf{Z}'\mathbf{u}_{21}^{(\mathrm{x})} \\ \vdots \\ \mathbf{Z}'\mathbf{u}_{2h}^{(\mathrm{x})} \end{pmatrix}. \tag{5.28}$$

The matrix equation (5.28) can be simplified in the following way. Define

$$\tilde{\mathbf{X}} \equiv \begin{pmatrix} \mathbf{W}_{11t} \\ \vdots \\ \mathbf{W}_{1m} \\ \mathbf{W}_{21} \\ \vdots \\ \mathbf{W}_{2n} \\ \mathbf{W}_{31} \\ \vdots \\ \mathbf{W}_{3h} \end{pmatrix}, \quad \tilde{\mathbf{y}} \equiv \begin{pmatrix} \mathbf{y}_{11} \\ \vdots \\ \mathbf{y}_{1m} \\ \mathbf{y}_{21} \\ \vdots \\ \mathbf{y}_{2n} \\ \mathbf{x}_{21} \\ \vdots \\ \mathbf{x}_{2h} \end{pmatrix}, \quad \tilde{\boldsymbol{\delta}} \equiv \begin{pmatrix} \boldsymbol{\delta}_{11}^{(y)} \\ \vdots \\ \boldsymbol{\delta}_{1m}^{(y)} \\ \boldsymbol{\delta}_{21}^{(y)} \\ \vdots \\ \boldsymbol{\delta}_{2n}^{(y)} \\ \boldsymbol{\delta}_{21}^{(x)} \\ \vdots \\ \boldsymbol{\delta}_{2h}^{(x)} \end{pmatrix}, \quad \tilde{\mathbf{u}} \equiv \begin{pmatrix} \mathbf{u}_{11}^{(y)} \\ \vdots \\ \mathbf{u}_{1m}^{(y)} \\ \mathbf{u}_{21}^{(y)} \\ \vdots \\ \mathbf{u}_{2n}^{(y)} \\ \mathbf{u}_{21}^{(x)} \\ \vdots \\ \mathbf{u}_{2h}^{(x)} \end{pmatrix}. \tag{5.29}$$

Using definitions (5.29) we can re-write (5.28) as $(\mathbf{I}\otimes\mathbf{Z}')\,\tilde{\mathbf{y}} = (\mathbf{I}\otimes\mathbf{Z}')\,\tilde{\mathbf{X}}\tilde{\boldsymbol{\delta}} + (\mathbf{I}\otimes\mathbf{Z}')\,\tilde{\mathbf{u}}$. Note that $\dim(\tilde{\mathbf{X}}) = (p+1)m+(q+1)g+n+h$. It is also possible to write a compact expression for the asymptotic covariance matrix as

$$E\left[(\mathbf{I}\otimes\mathbf{Z}')\,\tilde{\mathbf{u}}\tilde{\mathbf{u}}'\,(\mathbf{I}\otimes\mathbf{Z}')'\right] = E\left[(\mathbf{I}\otimes\mathbf{Z}')\,(\boldsymbol{\Sigma}\otimes\mathbf{I})\,(\mathbf{I}\otimes\mathbf{Z}')'\right] = \boldsymbol{\Sigma}\otimes E\,[\mathbf{Z}'\mathbf{Z}].$$

We can now define the FIVE estimator for the OF equations as

$$\begin{aligned}\tilde{\boldsymbol{\delta}} &= \left\{\tilde{\mathbf{X}}'\,(\mathbf{I}\otimes\mathbf{Z}')'\left[\boldsymbol{\Sigma}^{-1}\otimes(\mathbf{Z}'\mathbf{Z})^{-1}\right](\mathbf{I}\otimes\mathbf{Z}')\,\tilde{\mathbf{X}}\right\}^{-1}\tilde{\mathbf{X}}\,(\mathbf{I}\otimes\mathbf{Z}')' \\ &\qquad\times\left[\boldsymbol{\Sigma}^{-1}\otimes(\mathbf{Z}'\mathbf{Z})^{-1}\right](\mathbf{I}\otimes\mathbf{Z}')\,\tilde{\mathbf{y}} \\ &= \left\{\tilde{\mathbf{X}}'\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\right]\tilde{\mathbf{X}}\right\}^{-1}\tilde{\mathbf{X}}'\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\right]\tilde{\mathbf{y}}.\end{aligned}\tag{5.30}$$

The matrix equation (5.30) enables non-iterative estimation of all structural parameters and measurement model coefficients, jointly. This procedure, given the **Z** matrix is valid for all equations in the system, yields consistent and efficient estimates of the parameters of a dynamic SEM model.

Finally we briefly discuss computation of the residual covariance matrix $\boldsymbol{\Sigma}$. First note that

$$\boldsymbol{\Sigma} = E\left(\tilde{\mathbf{u}}_1\tilde{\mathbf{u}}_1'\right) = E\begin{bmatrix}\mathbf{u}_1\mathbf{u}'_1 & \mathbf{u}_1\mathbf{u}'_2 & \mathbf{u}_1\mathbf{u}'_3\\ \mathbf{u}_2\mathbf{u}'_1 & \mathbf{u}_2\mathbf{u}'_2 & \mathbf{u}_2\mathbf{u}'_3\\ \mathbf{u}_3\mathbf{u}'_1 & \mathbf{u}_3\mathbf{u}'_2 & \mathbf{u}_3\mathbf{u}'_3\end{bmatrix}\in\mathbb{R}^{(m+n+h)\times(m+n+h)}.\tag{5.31}$$

The individual scalar elements of the block elements of $\boldsymbol{\Sigma}$ are calculated as follows. We have $\mathbf{u}_1\mathbf{u}'_1 = \left\{E\left(u_{1i}^{(y)}u_{1j}^{(y)}\right)\right\}\in\mathbb{R}^{m\times m}$ with typical element

$$\tilde{\sigma}_{11ij}^{(yy)} = T^{-1}\left(\mathbf{y}_{1j}-\mathbf{W}_{1j}\boldsymbol{\delta}_{1j}^{(y)}\right)'\left(\mathbf{y}_{1j}-\mathbf{W}_{1j}\boldsymbol{\delta}_{1j}^{(y)}\right).$$

Similarly, we have $\mathbf{u}_1\mathbf{u}'_2 = \left\{E\left(u_{1i}^{(y)}u_{2j}^{(y)}\right)\right\}\in\mathbb{R}^{m\times n}$, and symmetrically $\mathbf{u}_2\mathbf{u}'_1 = \left\{E\left(u_{2i}^{(y)}u_{1j}^{(y)}\right)\right\}\in\mathbb{R}^{n\times m}$ with typical elements of the form

$$\tilde{\sigma}_{12ij}^{(yy)} = T^{-1}\left(\mathbf{y}_{1j}-\mathbf{W}_{1j}\boldsymbol{\delta}_{1j}^{(y)}\right)'\left(\mathbf{y}_{2j}-\mathbf{W}_{2j}\boldsymbol{\delta}_{2j}^{(y)}\right)$$

and

$$\tilde{\sigma}_{21ij}^{(yy)} = T^{-1}\left(\mathbf{y}_{2j}-\mathbf{W}_{2j}\boldsymbol{\delta}_{2j}^{(y)}\right)'\left(\mathbf{y}_{1j}-\mathbf{W}_{1j}\boldsymbol{\delta}_{1j}^{(y)}\right),$$

respectively.

Note that the $(1,3)$ block element of (5.31) is merely a transpose of the block $(3,1)$, thus the individual elements can be estimated in the same way, namely $\mathbf{u}_1\mathbf{u}'_3 = \left\{E\left(u_{1i}^{(y)}u_{2j}^{(x)}\right)\right\}\in\mathbb{R}^{m\times h}$ and $\mathbf{u}_3\mathbf{u}'_1 = \left\{E\left(u_{2j}^{(x)}u_{1i}^{(y)}\right)\right\}\in\mathbb{R}^{h\times m}$, which has a typical element of the form

$$\tilde{\sigma}_{13ij}^{(yx)} = T^{-1}\left(\mathbf{y}_{1j} - \mathbf{W}_{1j}\boldsymbol{\delta}_{1j}^{(y)}\right)'\left(\mathbf{x}_{2j} - \mathbf{W}_{3j}\boldsymbol{\delta}_{2j}^{(x)}\right).$$

Finally, for the remaining two blocks we have $\mathbf{u}_2\mathbf{u}'_2 = \left\{E\left(u_{2i}^{(y)}u_{2j}^{(y)}\right)\right\} \in \mathbb{R}^{n\times n}$, where scalar elements can be estimated by

$$\tilde{\sigma}_{22ij}^{(yy)} = T^{-1}\left(\mathbf{y}_{2j} - \mathbf{W}_{2j}\boldsymbol{\delta}_{2j}^{(y)}\right)'\left(\mathbf{y}_{2j} - \mathbf{W}_{2j}\boldsymbol{\delta}_{2j}^{(y)}\right),$$

and $\mathbf{u}_3\mathbf{u}'_3 = \left\{E\left(u_{2i}^{(x)}u_{2j}^{(x)}\right)\right\} \in \mathbb{R}^{h\times h}$ with typical element estimated by

$$\tilde{\sigma}_{22ij}^{(xx)} = T^{-1}\left(\mathbf{x}_{2j} - \mathbf{W}_{3j}\boldsymbol{\delta}_{2j}^{(x)}\right)'\left(\mathbf{x}_{2j} - \mathbf{W}_{3j}\boldsymbol{\delta}_{2j}^{(x)}\right)$$

Consistency of the FIVE estimator of the OF model can be shown in a similar way as is usually shown for classical simultaneous equation systems.[2] Consistency of this estimator can be shown by noting that

$$\begin{aligned}\hat{\tilde{\boldsymbol{\delta}}} &= \left\{\tilde{\mathbf{X}}'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right]\tilde{\mathbf{X}}\right\}^{-1}\tilde{\mathbf{X}}'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right]\tilde{\mathbf{y}}\\ &= \tilde{\boldsymbol{\delta}} + \left\{\tilde{\mathbf{X}}'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right]\tilde{\mathbf{X}}\right\}^{-1}\tilde{\mathbf{X}}'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right]\tilde{\mathbf{u}}.\end{aligned}$$

Proposition (5.2.2.4)–(i) implies that

$$\text{plim}\left(T^{-1}\mathbf{Z}'\mathbf{Z}\right) = \boldsymbol{\Sigma}_{ZZ},$$

therefore it follows that plim $\left(\hat{\boldsymbol{\Sigma}}^{-1}\otimes T\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right) = \boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}_{ZZ}^{-1}$, and by (5.2.2.4)–(ii) it follows that

$$\begin{aligned}\text{plim}\left(\frac{1}{T}\mathbf{Z}'\mathbf{X}^*\right) &= \text{plim}\left(\frac{1}{T}\left(\mathbf{Z}'\boldsymbol{\iota},\mathbf{Z}'\mathbf{Y}_{1j},\mathbf{Z}'\mathbf{X}_{1j}\right)\right)\\ &= \text{plim}\left(\frac{1}{T}\left(\mathbf{Z}'\boldsymbol{\iota},\mathbf{Z}'\mathbf{Y}_{1j},\mathbf{Z}'\mathbf{Y}_{1j,t-i},\mathbf{Z}'\mathbf{X}_{1jt},\mathbf{Z}'\mathbf{X}_{1j,t-i}\right)\right)\\ &= \left(\boldsymbol{\Sigma}_{Z\boldsymbol{\iota}},\boldsymbol{\Sigma}_{ZY_t},\boldsymbol{\Sigma}_{ZY_{t-i}},\boldsymbol{\Sigma}_{ZX_t},\boldsymbol{\Sigma}_{ZX_{t-i}}\right) = \boldsymbol{\Sigma}_{ZX^*} \qquad (5.32)\end{aligned}$$

Using Proposition (5.2.2.4)–(iv) and expanding the Kronecker products we get the following convergence results

[2]However, in the OF case, similarly to the GIVE case, consistency will depend on the assumed properties of the model-implied (lagged) instruments.

$$\operatorname{plim}\frac{(\mathbf{I}\otimes\mathbf{Z}')\tilde{\mathbf{u}}}{T}=\begin{pmatrix}\operatorname{plim}\frac{\mathbf{Z}'\mathbf{u}_{11}^{(y)}}{T}\\ \vdots\\ \operatorname{plim}\frac{\mathbf{Z}'\mathbf{u}_{1m}^{(y)}}{T}\\ \operatorname{plim}\frac{\mathbf{Z}'\mathbf{u}_{21}^{(y)}}{T}\\ \vdots\\ \operatorname{plim}\frac{\mathbf{Z}'\mathbf{u}_{2n}^{(y)}}{T}\\ \operatorname{plim}\frac{\mathbf{Z}'\mathbf{u}_{31}^{(x)}}{T}\\ \vdots\\ \operatorname{plim}\frac{\mathbf{Z}'\mathbf{u}_{3h}^{(x)}}{T}\end{pmatrix}=\mathbf{0}$$

From (5.32) it follows that

$$\operatorname{plim}\frac{(\mathbf{I}\otimes\mathbf{Z}')\,\tilde{\mathbf{X}}}{T}=\operatorname{diag}\begin{pmatrix}\left(\operatorname{plim}\frac{\mathbf{Z}'\mathbf{W}_{11}}{T}\right)'\\ \vdots\\ \left(\operatorname{plim}\frac{\mathbf{Z}'\mathbf{W}_{1m}}{T}\right)'\\ \left(\operatorname{plim}\frac{\mathbf{Z}'\mathbf{W}_{21}}{T}\right)'\\ \vdots\\ \left(\operatorname{plim}\frac{\mathbf{Z}'\mathbf{W}_{2n}}{T}\right)'\\ \left(\operatorname{plim}\frac{\mathbf{Z}'\mathbf{W}_{31}}{T}\right)'\\ \vdots\\ \left(\operatorname{plim}\frac{\mathbf{Z}'\mathbf{W}_{3h}}{T}\right)'\end{pmatrix}'. \tag{5.33}$$

Therefore, we have specifically

$$\begin{aligned}\operatorname{plim}\left(\frac{\mathbf{Z}'\mathbf{W}_{1j}}{T}\right) &= \operatorname{plim}\left(\frac{\mathbf{Z}'\left(\boldsymbol{\iota},\mathbf{Y}_{1j},\mathbf{X}_{1j}\right)}{T}\right)\\ &= \left(\boldsymbol{\Sigma}_{Z\iota}^{(j)},\boldsymbol{\Sigma}_{ZY_t}^{(j)},\,\boldsymbol{\Sigma}_{ZY_{t-i}}^{(j)},\,\boldsymbol{\Sigma}_{ZX_t}^{(j)},\boldsymbol{\Sigma}_{ZX_{t-i}}^{(j)}\right)=\boldsymbol{\Sigma}_{ZW_1}^{(j)},\end{aligned}$$

and

$$\operatorname{plim}\left(\frac{\mathbf{Z}'\mathbf{W}_{2j}}{T}\right)=\operatorname{plim}\left(\frac{\mathbf{Z}'(\boldsymbol{\iota},\mathbf{Y}_{1jt})}{T}\right)=\left(\boldsymbol{\Sigma}_{\mathrm{Z}\iota}^{(j)},\,\boldsymbol{\Sigma}_{\mathrm{ZY}_t}^{(j)}\right)=\boldsymbol{\Sigma}_{\mathrm{ZW}_2}^{(j)}.$$

Finally, it follows that

$$\operatorname{plim}\left(\frac{\mathbf{Z}'\mathbf{W}_{3j}}{T}\right)=\operatorname{plim}\left(\frac{\mathbf{Z}'(\boldsymbol{\iota},\mathbf{X}_{1jt})}{T}\right)=\left(\boldsymbol{\Sigma}_{\mathrm{Z}\iota}^{(j)},\,\boldsymbol{\Sigma}_{\mathrm{ZX}_t}^{(j)}\right)=\boldsymbol{\Sigma}_{\mathrm{ZW}_1}^{(j)}.$$

We can now derive the probability limits for the matrix (5.33) as

$$\operatorname{plim}\frac{(\mathbf{I}\otimes\mathbf{Z}')\tilde{\mathbf{X}}}{T}=\operatorname{diag}\begin{pmatrix}\left(\boldsymbol{\Sigma}^{(1)}_{ZW_1}\right)'\\ \vdots\\ \left(\boldsymbol{\Sigma}^{(\mathrm{m})}_{ZW_1}\right)'\\ \left(\boldsymbol{\Sigma}^{(1)}_{ZW_2}\right)'\\ \vdots\\ \left(\boldsymbol{\Sigma}^{(\mathrm{n})}_{ZW_2}\right)'\\ \left(\boldsymbol{\Sigma}^{(1)}_{ZW_3}\right)'\\ \vdots\\ \left(\boldsymbol{\Sigma}^{(\mathrm{h})}_{ZW_3}\right)'\end{pmatrix}'=\boldsymbol{\Sigma}_{Z\tilde{X}}.$$

Using the above results, it follows that

$$\begin{aligned}\operatorname{plim}(\hat{\tilde{\boldsymbol{\delta}}}) &= \tilde{\boldsymbol{\delta}}+\operatorname{plim}\left\{\tilde{\mathbf{X}}'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right]\tilde{\mathbf{X}}\right\}^{-1}\\ &\qquad\times\tilde{\mathbf{X}}'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right]\tilde{\mathbf{u}}\\ &= \tilde{\boldsymbol{\delta}}+\operatorname{plim}\left\{\tilde{\mathbf{X}}'\left(\mathbf{I}\otimes\mathbf{Z}'\right)'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]\left(\mathbf{I}\otimes\mathbf{Z}'\right)\tilde{\mathbf{X}}\right\}^{-1}\\ &\qquad\times\tilde{\mathbf{X}}'\left(\mathbf{I}\otimes\mathbf{Z}'\right)'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]\left(\mathbf{I}\otimes\mathbf{Z}'\right)\tilde{\mathbf{u}}\\ &= \tilde{\boldsymbol{\delta}}+\left\{\left[\operatorname{plim}\tfrac{\tilde{\mathbf{X}}'(\mathbf{I}\otimes\mathbf{Z}')'}{T}\right]\operatorname{plim}\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes T\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]\left[\operatorname{plim}\tfrac{(\mathbf{I}\otimes\mathbf{Z}')\tilde{\mathbf{X}}}{T}\right]\right\}^{-1}\\ &\qquad\times\left[\operatorname{plim}\tfrac{\tilde{\mathbf{X}}'(\mathbf{I}\otimes\mathbf{Z}')'}{T}\right]\operatorname{plim}\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes T\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]\left[\operatorname{plim}\tfrac{(\mathbf{I}\otimes\mathbf{Z}')\tilde{\mathbf{u}}}{T}\right]\\ &= \tilde{\boldsymbol{\delta}}+\left\{\boldsymbol{\Sigma}_{\tilde{X}Z}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}_{ZZ}^{-1}\right)\boldsymbol{\Sigma}_{Z\tilde{X}}\right\}^{-1}\boldsymbol{\Sigma}_{\tilde{X}Z}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}_{ZZ}^{-1}\right)\cdot\mathbf{0}\\ &= \tilde{\boldsymbol{\delta}}\end{aligned}$$

Therefore, the FIVE estimates are consistent if applied to the OF model. It is also possible to show that FIVE is asymptotically more efficient then the GIVE estimator.[3]

The FIVE estimator, just like the GIVE estimator, is distribution-free in the sense that it is asymptotically normally distributed, with the assumption of Gaussian disturbances and no distributional assumptions about the modelled variables.[4] To see this, note that

$$\begin{aligned}\sqrt{T}(\hat{\tilde{\boldsymbol{\delta}}}-\tilde{\boldsymbol{\delta}}) &= \left\{T^{-1}\tilde{\mathbf{X}}'\left(\mathbf{I}\otimes\mathbf{Z}'\right)'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes T\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]T^{-1}\left(\mathbf{I}\otimes\mathbf{Z}'\right)\tilde{\mathbf{X}}\right\}^{-1}\\ &\qquad\times T^{-1}\tilde{\mathbf{X}}'\left(\mathbf{I}\otimes\mathbf{Z}'\right)'\left[\hat{\boldsymbol{\Sigma}}^{-1}\otimes T\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]T^{-1/2}\left(\mathbf{I}\otimes\mathbf{Z}'\right)\tilde{\mathbf{u}}\end{aligned}$$

Assuming Gaussian disturbances,

[3]Such proof is similar to the one for the ordinary simultaneous equation systems (see e.g. Judge et al., 1985.

[4]This is unlike to the Gaussian covariance structure based static SEM method which generally requires the modelled variables to be multivariate Gaussian.

$$\tilde{\mathbf{u}} \sim N\left[\mathbf{0},\ \boldsymbol{\Sigma}\otimes\mathbf{I}\right] \Rightarrow T^{-1/2}\left(\mathbf{I}\otimes\mathbf{Z}'\right)\tilde{\mathbf{u}} \sim N\left[\mathbf{0},\ \boldsymbol{\Sigma}\otimes T^{-1}\mathbf{Z}'\mathbf{Z}\right],$$

therefore asymptotic normality immediately follows, i.e.,

$$\sqrt{T}(\hat{\tilde{\boldsymbol{\delta}}} - \tilde{\boldsymbol{\delta}}) \sim N\left[\mathbf{0},\ \left(\boldsymbol{\Sigma}_{\tilde{X}Z}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}_{ZZ}^{-1}\right)\boldsymbol{\Sigma}_{Z\tilde{X}}\right)^{-1}\right],$$

where $\left[\boldsymbol{\Sigma}_{\tilde{X}Z}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}_{ZZ}^{-1}\right)\boldsymbol{\Sigma}_{Z\tilde{X}}\right]^{-1} \approx \left[\tilde{\mathbf{X}}'\left(\hat{\boldsymbol{\Sigma}}^{-1}\otimes\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right)\tilde{\mathbf{X}}\right]^{-1}$ and $\boldsymbol{\Sigma}'_{Z\tilde{X}} = \boldsymbol{\Sigma}_{\tilde{X}Z}$.

It is worth emphasizing that the above results apply only if the IV matrix $\mathbf{Z}$ is valid for all equations of the system. Practically, this means that for estimation of the j^{th} equation there might be eligible instruments that are not eligible for estimation of the i^{th} equation for $j \neq i$. Formally, $\mathbf{Z}$ that is eligible for the entire OF system contains the intersection of the rows of $\mathbf{Z}_\mathrm{j}$, i.e., instruments for each j^{th} equation in the system. If there is enough instruments[5] such matrix $\mathbf{Z}$ can be constructed so to enable identification of each equation in the system and consistent FIVE estimation.

However, as already mentioned, model-implied validity might by misleading if the model itself is mis-specified, thus empirical testing of IV's validity is essential.

5.4 Identification

Identification of the static structural equation models with latent variables is generally problematic. An early discussion of this topic can be found already in Wiley (1973), but a simple and straightforward procedure still does not exist. On the other hand, identification is well defined and straightforward in classical econometric simultaneous equation systems, and a similar approach can be developed for the OF equations.

We propose a simple procedure that uses only the coefficient matrices from the latent specification for identifying the OF estimation equations. The following technique provides sufficient conditions for identification of all equations in the systems.

Proposition 5.4.0.1 *Given a DSEM model with the structural equation of the form* $\boldsymbol{\eta}_t = \boldsymbol{\alpha}_\eta + \sum_{j=0}^{p}\mathbf{B}_j\boldsymbol{\eta}_{t-j} + \sum_{j=0}^{q}\boldsymbol{\Gamma}_j\boldsymbol{\xi}_{t-j} + \boldsymbol{\zeta}_t$ *and the measurement model given by* $\mathbf{x}_t = \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi}_t + \boldsymbol{\delta}_t$ *and* $\mathbf{y}_t = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t$ *define*

[5] Given sufficiently long time span of data, there will always be enough lagged variables to satisfy this requirement

$$\mathbf{K} = \begin{pmatrix} (\mathbf{I}-\mathbf{B}'_0) & -\mathbf{\Lambda}_2^{(y)\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_h \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} -\boldsymbol{\alpha}_\eta & -\boldsymbol{\alpha}_2^{(y)} & -\boldsymbol{\alpha}_2^{(x)} \\ -\mathbf{B}'_1 & \mathbf{0} & \mathbf{0} \\ -\mathbf{B}'_2 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\mathbf{B}'_p & \mathbf{0} & \mathbf{0} \\ -\mathbf{\Gamma}'_0 & \mathbf{0} & -\mathbf{\Lambda}_2^{(x)\prime} \\ -\mathbf{\Gamma}'_1 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\mathbf{\Gamma}'_q & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Then, the j^{th} equation of the system will be identified iff

$$\operatorname{rank}\left[\mathbf{R}_j \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix}\right] \geq m+n+h-1 \tag{5.34}$$

where $\mathbf{R}_j$ is a zero-one selection matrix having one's in places of omitted variables and one row for each omission. Note that if the equality holds the equation is exactly identified, otherwise it is overidentified.

Corollary 5.4.0.2 *A corollary to Proposition (5.4.0.1) states that unless*

$$\operatorname{rank}(\mathbf{R}_j) \geq m+n+h-1 \tag{5.35}$$

the j^{th} equation is not identified. The condition (5.35) is necessary for identification, while condition (5.34) is sufficient.

Proof See Appendix §5C.

It is therefore possible to use these rules to check for identification of each individual equation. The relevance of this approach lies in its ability to check for identification of the model that is specified in latent form and thus it avoids the need to derive the OF equations. In addition, this method is equally applicable for both static and dynamic structural equation models with latent variables.

Chapter 6

Empirical applications

6.1 Introduction

This section applies DSEM/PDSEM models to empirical data, focusing on fixed and random effects dynamic panel models, using the available SEM/econometric software packages.

This application has two main aims. First is to demonstrate the effect of using the starting values obtained by the methods proposed in chapter §5. Second aim is to test the capabilities of the existing commercial software packages, such as LISREL, in estimation of DSEM models for panel data.

Firstly, we will estimate a fixed-effects DPSEM model using a cross-country panel data set, where the sample size is moderate and time dimension is small. This is an example of a dynamic panel model with fixed effects, $T < N$, and both T and N relatively small.

The second application is to a micro (household) panel data using 13 available years of the British Household Panel Study (BHPS) to estimate a random effects dynamic micro consumption model. The BHPS data span over 13 years and include over 5,000 individuals, which are traced over time. This is, hence, an application to a very large data set with more pronounced temporal dynamics.

We will use the LISREL package (Jöreskog and Sörbom 1996b) for maximum likelihood estimation of the SEM models, and obtain starting values using instrumental variables methods and other software packages.

LISREL has its roots in a software package for estimation of structural equation models using ACS methods developed by Gruvaeus and Jöreskog (1970) as a Fortran-IV programme. Its predecessors were the ACOVS programme (for the analysis of covariance structures), and the FIELES programme (for the classical simultaneous equation models without measurement error) both due to Jöreskog et al. (1970). These programmes played a seminal role as ancestors of the LISREL package (Jöreskog and Sörbom 1996b), the programme that became synonymous with structural equation modelling. We will briefly describe this packages.

Model specification in LISREL is designed in a very general way, and in principle most linear models can be formulated and estimated with the LISREL syntax language. This syntax language is, however, designed primarily for cross-section (independent) data and hence the matrices that need to be specified as input to the programme refer those of a static model. Estimation of simple models for panel data is nevertheless possible by forcing the programme into estimating multiple equations with cross-equation equality constraints–a trick that specifies a dynamic covariance structure by treating the lagged endogenous variables as distinct exogenous variables. In summary, the specification of DPSEM models is not simple in the LISREL syntax, which is best suited for estimation of static SEM models (2.4)–(2.6), as the syntax refers only to the elements of the general $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ matrices. Though not easily, DPSEM models can be formulated in the LISREL syntax by treating all parameter matrices as belonging to a single matrix and then imposing various restrictions on the parameters to obtain the required DPSEM structure.

However, as the time dimension of the panel data increases, the number of equations with equality constraints also increases making the syntax very difficult to build and manipulate. In addition, dynamic panel models with pronounced time series dimension present a considerable optimisation challenge in practice.

LISREL uses numerical optimisation based on a modified Davidon-Fletcher-Powel (DFP) quasi-Newton algorithm (Jöreskog 1973, Jöreskog et al. 1970, Gruvaeus and Jöreskog 1970, Jöreskog 1977, Lee and Jennrich 1979). The modification due to Karl Jöreskog adds several iterations of the steepest decent preceding the DFP iterations, which lead to more rapid convergence.

The analytical first derivatives $\partial \ln L/\partial \boldsymbol{\theta}$ are obtained without taking into account equal elements in the symmetrical coefficient matrices, which are in turn handled by the equality constraints on the off-diagonal elements (Jöreskog 1977). Letting the unconstrained parameter vector is $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{r_1})$, suppose $r_1 - r_2$ elements of $\boldsymbol{\theta}$ are known constants (e.g. fixed to zero), with $r_2 \leq r_1$ free parameters $\boldsymbol{\theta}^* = (\theta_1^F, \ldots, \theta_{r_2}^F)$. Further, suppose that out of r_2 free parameters $r_2 - r_3$ are equal and $r_3 \leq r_2$ are distinct, with free distinct parameters now assembled in $\boldsymbol{\theta}^{**} = (\theta_1^{FD}, \ldots, \theta_{r_2}^{FD})$. If we let $\boldsymbol{R}$ $(r_1 \times r_2)$ be a zero-one selection matrix that deletes the elements of the score vector corresponding to the fixed parameter such that $\boldsymbol{R\theta} = \boldsymbol{\theta}^*$ the constant restrictions can be imposed by solving the restricted score equation $\boldsymbol{R}[\partial \ln L/\partial \boldsymbol{\theta}] = \boldsymbol{0}$, where L denotes the likelihood function. The equality restrictions can be imposed similarly by defining another zero-one matrix $\boldsymbol{K}$ $(r_2 \times r_3)$ with elements $\{k_{ij}\}$ such that

$$k_{ij} = \begin{cases} 1, & \theta_i^F = \theta_j^{FD} \\ 0, & \theta_i^F \neq \theta_j^{FD} \end{cases}$$

Therefore, the constant and equality restrictions can be imposed directly on the score vector as $\boldsymbol{K}'\boldsymbol{R}[\partial \ln L/\partial\boldsymbol{\theta}] = \mathbf{0}$.

Numerical optimisation algorithm implemented in LISREL uses an initial estimate $\boldsymbol{E}_0 = \boldsymbol{K}'\boldsymbol{R}E[(\partial \ln L/\partial\boldsymbol{\theta})(\partial \ln L/\partial\boldsymbol{\theta})']\boldsymbol{R}'\boldsymbol{K}$, which is updated by five iterations of the steepest decent algorithm (a modification due to Kark Jöreskog), followed by DFP iterations until convergence to the estimate $\tilde{\boldsymbol{E}}$.[1]

[1]While leading to faster convergence then the basic DFP algorithm, the modified-DFP is found to be inferior to Gauss-Newton algorithm and also it is known to produce incorrect standard errors (Lee and Jennrich 1979).

6.2 Application I: Modelling finance and growth

Panel models with simultaneity, dynamics, and latent variables are common place in empirical econometrics. A widely researched example is the relationship between financial development (FD) and growth. This is a theoretically ambiguous relationship since economic models indicating both positive and negative relationship exist in the literature. King and Levine (1993a), for example, suggest a positive FD-growth effect, while Bencivenga and Smith (1991) and Bencivenga et al. (1995) indicate a possibility of both positive and negative effects. Lucas (1988), on the other hand, dismisses the FD-growth effect altogether. Levine (2003) gives a detailed review of this literature. Without unambiguous theoretical implications, the finance-growth relationship thus remains an empirical issue. Nevertheless, the empirical literature failed to give a conclusive answer although preponderance of the empirical studies claim a positive FD-growth effect (Levine 1997, Levine and Zervos 1996, Demetriades and Hussein 1996, Levine and Zervos 1998, Neuser and Kugler 1998, Levine 1999, Rousseau and Wachtel 2000, Levine et al. 2000, Hali et al. 2002, Levine 2003).

The key statistical issues in the FD-growth research relate to the modelling and testing of the substantively implied latent structure of the unobservable (latent) financial development. While the mainstream FD-growth literature based on the IV/GMM methods does not explicitly test for the measurement errors by estimating formal statistical measurement models for the latent variable, it does suggest various observable FD indicators on the substantive grounds. Naturally, this introduces the problem of whether and how well the available indicators measure a single latent construct and how much error is contained in such indicators. In addition, the FD-growth simultaneity is held to be an important consideration and the dynamics and lagged feedback effects are both implied by the substantive theory.

Earlier studies (Levine 1997, Levine and Zervos 1998) used simple cross-country OLS regressions of GDP growth on the separate FD indicators without accounting for the cross-country heterogeneity or simultaneity problems. Separate growth regressions with individual observable indicators containing measurement error might result in the errors-in-variables problem and thus produce biased or inconsistent coefficient estimates. The inconsistency of the regression coefficients due to the measurement error is potentially considerable, which most profoundly concerns the actual relationship between the financial development and economic growth. In homogeneous random samples the measurement error biases regression coefficients towards zero, however, with heterogeneous cross-country data with fixed country-specific effects, the bias can go either way and the problem can be further magnified by the inclusion of other variables in a multiple regression setup (see e.g. Wansbeek and

Meijer (2000)). A major complication arises with heterogeneous samples (such as cross-sections of countries) where individual (fixed) effects might be correlated with the measurement-error components resulting from using noisy indicators in place of the (unobservable) latent variables (Griliches and Hausman 1986, Wansbeek 2001). Consequently, the more recent empirical literature uses panel data and instrumental variable methods (Rousseau and Wachtel 2000, Neusser and Kugler 1998, Levine 1999, Levine et al. 2000, Hali et al. 2002). While the panel studies suggested a similar positive finance-growth relationship, it was shown that even with similar methods and data different conclusions can be reached (Favara 2003). The most likely source of the problem is the failure to model the measurement structure of the latent financial development along with modelling the simultaneous and dynamic effects. Consequently, on the basis of such results we cannot assess validity of the substantively suggested FD indicators even if the errors-in-variables problem is corrected by using the IV methods.

The measurement error problem

There is a large body of empirical literature that investigates the FD-growth relationship using multiple observable indicators of the latent (unobservable) financial development. Commonly used indicators include various measures of the banking sector such as liabilities of commercial and central banks, domestic credit, and credit to the private sector (King and Levine 1993a, King and Levine 1993b, Levine 1997, Levine and Zervos 1998, Neusser and Kugler 1998, Levine 1999, Rousseau and Wachtel 2000, Levine et al. 2000, Hali et al. 2002, Levine 2003, Rousseau and Wachtel 2000, Neusser and Kugler 1998, Levine 1999, Levine et al. 2000, Hali et al. 2002, Favara 2003).

The observable indicators are generally identified on substantive grounds and used as individual regressors in separate growth regressions. The measurement issue is not addressed in this literature through statistical testing, which might have resulted in the collection of inappropriate indicators or produced wrong conclusions about the FD-growth relationship. This constitutes a major omission since the availability of multiple indicators allows identification of the measurement error components and statistical evaluation of the FD measurement models.

The errors-in-variables problem arising from the latent nature of the financial development can be generalised to the case of multiple observable indicators by a factor-analytic model. Suppose we can observe m_j noisy indicators x_{ij} of the unobservable variable ξ_j. Then we can specify a factor model

$$x_{ij} = \lambda_{ij}\xi_j + \delta_i, \quad i = 1, \dots, k, \quad j = 1, \dots, g, \tag{6.1}$$

where x_{ij} is the ith observable indicator of the jth latent variable ξ_j, and δ_{ij} is the measurement error. The error covariance matrix is required to be diagonal, $E[\boldsymbol{\delta}\boldsymbol{\delta}'] = \operatorname{diag}\left(\sigma^2_{\delta_1}, \ldots, \sigma^2_{\delta_m}\right)$. Though implicitly, a factor model for the latent FD variable is implied by the substantive theory which suggests multiple indicators and linear relationships between the indicators and the unobservable components. Obviously the classical errors-in-variables model $x = \xi + \delta$ is a special case of the general factor model with one observable indicator and λ fixed to 1.

Once the latent structure is explicitly recognized and modelled the main issue becomes whether and how well the observable indicators measure the postulated latent construct(s), which can be easily tested by simple confirmatory factor analysis. To illustrate these issues, we will give some new empirical results using the same data as in the existing literature.

For the first empirical illustration, consider the FD measurement models implied by Levine and Zervos (1998) who investigate the relationship between economic growth and various stock market development indicators. In addition, they also consider multiple indicators of economic development using the following observable variables in their analysis *GDP growth*, *capital stock growth*, *productivity growth*, *savings*, *capitalization*, *value traded*, *turnover*, *CAPM integration*, *ATP integration*. Using data from a cross-section of 47 countries, time-averaged over the 1976–1993 period, Levine and Zervos (1998) estimated a series of separate growth regressions of the particular economic growth indicators on the various stock market development indicators without testing the measurement models for the two latent concepts. The key underlying assumption was that these indicators indeed measure the economic growth and the stock market development, respectively. This implies a two-factor model with *GDP growth*, *capital stock growth*, and *productivity growth* measuring the latent economic growth and with *savings*, *capitalization*, *value traded*, *turnover*, *CAPM integration*, and *ATP integration* measuring stock market development. Using the same data as Levine and Zervos (1998), we fitted the two-factor model with maximum likelihood, which produced a χ^2 fit statistic of 125.81 with 26 degrees of freedom. This strongly rejects the model. Furthermore, the estimated error variance of the *GDP growth* is[2] -0.11 (0.09) while the correlation between the two latent variables is 0.33 (0.13). Individual (cross-sectional) correlations between growth indicators and FD indicators are all positive but the mis-fit of the measurement model is problematic. Namely, the postulated indicators of the financial development and the economic growth do not seem to measure the hypothesized latent variables well, which brings in question the conclusions about the FD-growth relationship made by Levine and Zervos (1998).

As a second example we take the Hali et al. (2002) study of the international

[2]Standard error is in the parentheses.

financial integration and economic growth, where the latent international financial integration is measured by several observable indicators. Hali et al. (2002) use panel data from 57 countries over five 5-year periods (1976-1980, 1981-1985, 1986-1990, 1991-1995, 1996-2000) and investigate the effect of the international financial integration on the GDP growth. The observable indicators are *capital account restriction measure*, *stock of accumulated capital flows divided by GDP*, *capital inflows and outflows divided by GDP*, *stock of accumulated capital inflows divided by GDP*, *capital inflows.* We fitted a single factor model to these indicators obtaining a χ^2 goodness-of-fit statistic of 725.793 (d.f. $= 5$), which strongly rejects the hypothesis that these five indicators measure a single latent variable. A trivial modelling exercise easily identifies the source of the problem which turns out to be associated with the *capital inflows* indicator. Re-estimating the model without *capital inflows* produced an insignificant χ^2 of 5.879 (d.f. $= 2$). These results suggests that *capital inflows* does not measure the same latent variable as the other indicators. Interestingly, the growth regressions estimated by Hali et al. (2002) using individual indicators in separate regressions find significant effect of financial integration on GDP growth across various specifications mainly when *capital inflows* is used the financial integration indicator.

The above two examples illustrate the likely drawback of not estimating the measurement errors and of selecting noisy indicators of latent variables without empirically testing the implied measurement models.

Our final example considers the possible bias of the regression coefficients due to the measurement error. It is known that measurement error in the regressors can bias the regression coefficients downwards (Aigner et al. 1984, Wansbeek and Meijer 2000). However, in heterogenous samples such as cross sections of countries, due to the possible correlation between the fixed effects and the measurement error, the direction of the bias cannot be easily determined. We will illustrate this problem in the context of the FD-growth models when financial development is unobservable but measured by various noisy indicators. We use the same data as Demirgüç-Knut and Levine (2001a), on 84 countries averaged from 1969 to 1995 where the variables are several indicators of the financial development, GDP growth (ΔGDP_i), logarithm of the initial GDP (i_i), government expenditure (gov_i), change in consumer prices (Δp_i) and a sum of exports plus imports divided by GDP ($trade_i$). We estimate a simple FD-growth model

$$\Delta GDP_i = \gamma_1 FD_i + \gamma_2 i_i + \gamma_3 gov_i + \gamma_4 \Delta p_i + trade_i, \tag{6.2}$$

as commonly done in the literature (e.g. Demirgüç-Knut and Levine (2001b)). Estimating the regression equation (6.2) by using individual noisy indicators such as liquid liabilities of the banks (l_i), share of domestic credit from deposit banks (b_i), or credit to private sector (p_i) produced three separate regression equations with γ_1 coefficients 1.92 (0.84), 3.767 (1.31), and 1.34 (0.765), with l_i, b_i, and p_i as regressors, respectively. When (6.2) is estimated as a SEM model with the latent financial development measured with all three observable indicators, the γ_1 coefficient is 1.21 (0.45). The coefficient estimates of γ_2, γ_3, and γ_4 were very similar across all four equations. It is immediately noticeable that γ_1 differs considerably in magnitude across different models, which is indicative of the measurement error bias. In this case the bias from using individual noisy indicators seems to be upward. However, it is difficult to make valid conclusions without modelling the possible feedback from growth to financial development with a temporal lag and without accounting for the country effects.

Data and variables

We will estimate an empirical DPSEM FD-growth model to illustrate the above discussed methods using panel data on 45 countries observed over 25 years, running from 1970 till 1995, and averaged over 5-year periods.[3] Our data come from the same sources as the data used by Demirgüç-Knut and Levine (2001b) and Levine et al. (2001), thereby avoiding possible data-induced effects in the empirical results. The empirical studies such as Beck et al. (2000) and Beck and Levine (2003) use data averaged over the five years periods in order to abstract from the business cycle effects and we follow the same approach here.

While a criticism that business cycle dynamics should be better modelled by using temporally less aggregated data (e.g. quarterly or annual series), the use of a relatively small number of time averages does not itself cause asymptotic difficulties for our purposes. While the maximum likelihood estimator of the fixed effects requires the "$T \to \infty$" asymptotics for the consistent estimation of the time means, this primarily concerns the time span of the data rather then how the series were aggregated.[4]

We estimate a simple FD-growth model that accounts for the dynamics and the

[3]For 25 years of annual data the use of the 5-year averages requires computing $\bar{w}_1 = \frac{1}{5}\sum_{i=1}^{5} w_i$, $\bar{w}_2 = \frac{1}{5}\sum_{i=1}^{5} w_{5+i}$, $\bar{w}_3 = \frac{1}{5}\sum_{i=1}^{5} w_{10+i}$, $\bar{w}_4 = \frac{1}{5}\sum_{i=1}^{5} w_{15+i}$, and $\bar{w}_5 = \frac{1}{5}\sum_{i=1}^{5} w_{20+i}$.

[4]Generally, for the l-period time averages, the overall time mean can be written as

$$\frac{1}{T}\sum_{t=1}^{T} w_t = \frac{1}{T}\sum_{j=1}^{T/l}\sum_{i=1}^{l} w_{jl+i},$$

which implies that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} w_t = \lim_{T\to\infty}\frac{1}{T}\sum_{j=1}^{T/l}\sum_{i=1}^{l} w_{jl+i}.$$

Therefore, the use of time-averaged data does not introduce the " short T" problem in respect to the maximum likelihood estimator of the individual fixed effects since the consistency of this estimator will still depend on the length of the original (un-averaged) time series of length T.

measurement error. Formulating such model as a DPSEM model enables us to simultaneously model the measurement structure of the latent financial development and its possible effects on the economic growth. Since DPSEM is a multi-equation model, it is straightforward to include the second equation in which financial development is endogenous, possibly affected by the lagged economic growth. The variable definitions are given in Table 6.1.

Table 6.1: Observable and latent variables

Symbol	Definition of variable
	Observable variables
b_t	Deposit bank domestic credit divided by the sum of deposit bank domestic credit and central bank domestic credit
p_t	Currency plus demand and interest-bearing liabilities of banks and nonbank financial intermediaries divided by GDP
l_t	Value of credits by financial intermediaries to the private sector divided by GDP
g_t	Rate of real per capita GDP growth
i_t	Log of real GDP per capita in beginning of the period
	Latent variables
G_t	Economic growth
F_t	Financial system development
I_t	Initial economic development

The indicators of the financial system development are constructed in the same way as the indicators in the mainstream empirical FD-growth literature to avoid introduction of data-specific differences in the results (see e.g. Back et al. (2000) and Demirgüç-Knut and Levine (2001b)). The empirical density plots (Figure 6.1) suggest the observable indicators have reasonably symmetric bell-shaped distributions, thus can be treated as approximately normally distributed, which might be of concern in maximum likelihood estimation given the modest sample size.

A naive finance-growth model

We will illustrate this important difference in the approach by considering a simple "naive" finance-growth model

$$GDP_t = \alpha + \beta_1 b_t + \beta_2 i_t + \gamma_1 GDP_{t-1} \tag{6.3}$$

that uses a single financial development indicator (e.g. b_t) and ignores the possible measurement error problem.

Consider first the panel data set consisting of N observations on each each of the variables from (6.3), observed at T time points and placed in the "wide" panel matrix $\boldsymbol{X}_W$ of the form

$$
\boldsymbol{X}_W = \begin{pmatrix} GDP_{t=1}^{n=1} & \cdots & GDP_{t=5}^{n=1} & b_{t=1}^{n=1} & \cdots & b_{t=5}^{n=1} & i_{t=1}^{n=1} & \cdots & i_{t=5}^{n=1} \\ GDP_{t=1}^{n=2} & \cdots & GDP_{t=5}^{n=N} & b_{t=1}^{n=2} & \cdots & b_{t=5}^{n=2} & i_{t=1}^{n=2} & \cdots & i_{t=5}^{n=2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ GDP_{t=1}^{n=N} & \cdots & GDP_{t=5}^{n=N} & b_{t=1}^{n=N} & \cdots & b_{t=5}^{n=N} & i_{t=1}^{n=N} & \cdots & i_{t=5}^{n=N} \end{pmatrix} \tag{6.4}
$$

which allows computation of the covariance matrix

$$
\boldsymbol{S}_W = \frac{1}{N-1} \boldsymbol{X}'_W \boldsymbol{X}_W \tag{6.5}
$$

The standard approach to estimate the parameters of model (6.3) with standard SEM software package such as LISREL is to specify a multi-equation model and impose cross-equation equality constraints. In LISREL notation this model can be specified as

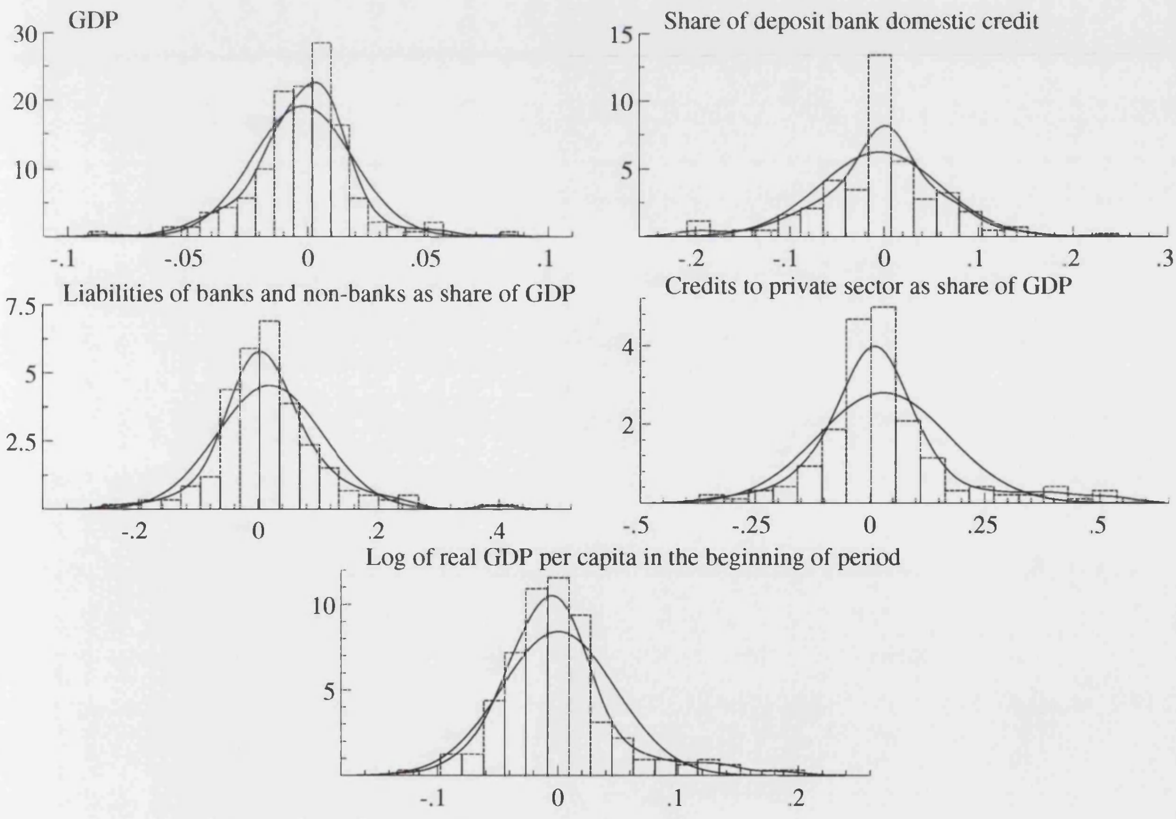

Figure 6.1: Empirical density of the observable variables

$$
\underbrace{\begin{pmatrix} GDP_{t=2} \\ GDP_{t=3} \\ GDP_{t=4} \\ GDP_{t=5} \end{pmatrix}}_{\boldsymbol{\eta}} = \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 0 & 0 \\ 0 & \beta_{32} & 0 & 0 \\ 0 & 0 & \beta_{43} & 0 \end{pmatrix}}_{\boldsymbol{B}} \underbrace{\begin{pmatrix} GDP_{t=2} \\ GDP_{t=3} \\ GDP_{t=4} \\ GDP_{t=5} \end{pmatrix}}_{\boldsymbol{\eta}}
$$

$$
+ \underbrace{\begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_{24} & \gamma_{25} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma_{36} & \gamma_{37} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{48} & \gamma_{59} \end{pmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix} GDP_{t=1} \\ b_{t=2} \\ i_{t=2} \\ b_{t=3} \\ i_{t=3} \\ b_{t=4} \\ i_{t=4} \\ b_{t=5} \\ i_{t=5} \end{pmatrix}}_{\boldsymbol{\xi}} + \underbrace{\begin{pmatrix} \zeta_{t=2} \\ \zeta_{t=3} \\ \zeta_{t=4} \\ \zeta_{t=5} \end{pmatrix}}_{\boldsymbol{\zeta}} \tag{6.6}
$$

subject to equality constraints

$$
\begin{aligned}
&\beta_{21} = \beta_{32} = \beta_{43} = \gamma_{11} \equiv \beta_1 \\
&\gamma_{12} = \gamma_{24} = \gamma_{36} = \gamma_{48} \equiv \gamma_1 \\
&\gamma_{13} = \gamma_{25} = \gamma_{37} = \gamma_{59} = \gamma_2 \equiv \gamma_2 \\
&\psi_{22} = \psi_{33} = \psi_{44} = \psi_{55}
\end{aligned}
$$

A path diagram corresponding to specification (6.6) is given in Figure (6.2).

The LISREL syntax needed to estimate this model can be written as

```
TI
DA NI=35 NO=225 NG=1 MA=CM
CM=widepanel.CM
SE 8 15 22 29 1 10 14 17 21 24 28 31 35 /
MO NX=9 NY=4 BE=FU,FI GA=FU,FI PH=SY,FR PS=DI,FR
FR BE(2,1) BE(3,2) BE(4,3) GA(1,1) GA(1,2) GA(1,3)
FR GA(2,4) GA(2,5) GA(3,6) GA(3,7) GA(4,8) GA(4,9)
EQ BE(2,1) BE(3,2) BE(4,3) GA(1,1)
EQ GA(1,2) GA(2,4) GA(3,6) GA(4,8)
EQ GA(1,3) GA(2,5) GA(3,7) GA(4,9)
EQ PS(1,1) PS(2,2) PS(3,3) PS(4,4)
OU ME=ML ND=5
```

where the "EQ" lines serve the purpose of imposing the cross-equation equality constraints on coefficients and error variances. We obtain the following empirical estimates

$$
GDP_t = \underset{(0.012)}{0.112}\, b_t - \underset{(0.016)}{0.043}\, i_t - \underset{(0.037)}{0.133}\, GDP_{t-1}. \tag{6.7}
$$

Similar results could be obtained by using standard econometric software such as PcGive or Ox and estimate a system of equations with cross-equation restrictions such as

$$\begin{aligned}
GDP_{t=2} &= \beta_1^{(1)} b_{t=2} + \beta_2^{(1)} i_{t=2} + \gamma_1^{(1)} GDP_{t=1} \\
GDP_{t=3} &= \beta_1^{(2)} b_{t=3} + \beta_2^{(2)} i_{t=3} + \gamma_1^{(2)} GDP_{t=2} \\
GDP_{t=4} &= \beta_1^{(3)} b_{t=4} + \beta_2^{(3)} i_{t=4} + \gamma_1^{(3)} GDP_{t=3} \\
GDP_{t=5} &= \beta_1^{(4)} b_{t=5} + \beta_2^{(4)} i_{t=5} + \gamma_1^{(4)} GDP_{t=4}
\end{aligned}$$

with cross-equation equality constraints

$$\begin{aligned}
\beta_1^{(1)} = \beta_1^{(2)} = \beta_1^{(3)} = \beta_1^{(4)} \equiv \beta_1 \\
\gamma_1^{(1)} = \gamma_1^{(2)} = \gamma_1^{(3)} = \gamma_1^{(4)} \equiv \gamma_1
\end{aligned}$$

In PcGive batch language, for example, this model can be estimated by running the following code, where letters "A", "B",...,"E" denote different time periods,

```
module("PcFiml");
```

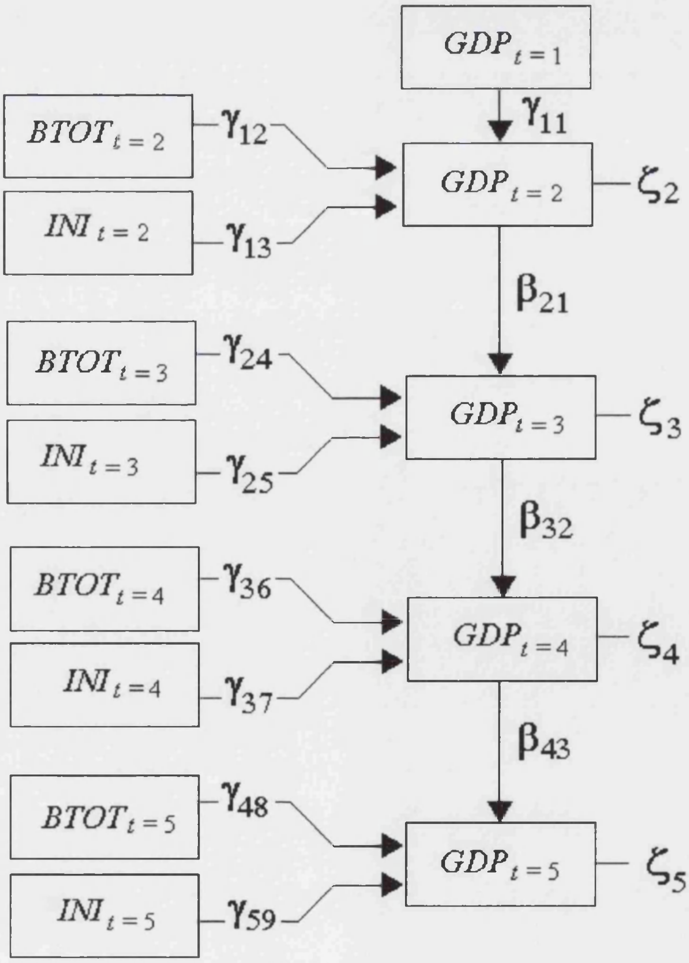

Figure 6.2: Single equation model

```
usedata("widepanel.xls");
system
{
    Y = BGDP, CGDP, DGDP, EGDP;
    Z = AGDP, BBTOT, BINI, CBTOT,
        CINI, DBTOT, DINI, EBTOT, EINI,
        Constant;
}
estsystem("OLS", 1, 1, 45, 1);
model
{
    BGDP = BBTOT, BINI, AGDP, Constant;
    CGDP = CBTOT, CINI, BGDP, Constant;
    DGDP = DBTOT, DINI, CGDP, Constant;
    EGDP = EBTOT, EINI, DGDP, Constant;
}
estmodel("FIML");
constraints
{
&0=&12; &4=&12; &8=&12;
&1=&13; &5=&13; &9=&13;
&2=&14; &6=&14; &10=&14;
&3=&12; &7=&12; &11=&12;
}
estmodel("CFIML");
```

which computes constrained full-information maximum likelihood (CFIML) estimates, but does not impose equality constraints on the error variance across all time periods, hence the empirical estimates would be numerically slightly different from those obtained by LISREL. However, standard econometric packages such as PcGive are generally not used for estimation of "wide-panel" models, rather the common approach is to estimate the "long-panel" models. We consider this approach next.

The approach taken in the panel-econometrics literature is based on what might be called the "long panel" format. Let $\boldsymbol{Y}_{NT}^{(l)}$ denote a cross-section time series data on N individuals observed over T time periods for variable "l". Also let $\boldsymbol{L}^j$ be a lag operator resulting in jth lag of $\boldsymbol{Y}_{NT}^{(l)}$. For a simple case with $N = 2$, $T = 3$ we have

$$\boldsymbol{Y}_{23}^{(l)} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix}, \quad \boldsymbol{Y}_{2(3-1)}^{(l)} \equiv \boldsymbol{L}^1 \boldsymbol{Y}_{23}^{(l)} = \begin{pmatrix} 0 \\ y_{11} \\ y_{12} \\ 0 \\ y_{21} \\ y_{22} \end{pmatrix}, \quad \boldsymbol{Y}_{2(3-2)}^{(l)} \equiv \boldsymbol{L}^2 \boldsymbol{Y}_{23}^{(l)} = \begin{pmatrix} 0 \\ 0 \\ y_{11} \\ 0 \\ 0 \\ y_{21} \end{pmatrix}$$

Since by assumption the pre-sample observations are zero, applying the jth lag operator will have the effect of shifting the values in the variable vector down j places and replacing first j places with zeros for each $i = 1, \ldots, N$. This operation can be concisely written in matrix notation as

$$\left(\boldsymbol{S}_{NT}^{j} \boldsymbol{Y}_{NT}^{(l)}\right) \odot \left(\boldsymbol{\iota}_N \otimes \boldsymbol{S}_T^{j} \boldsymbol{\iota}_T\right) \tag{6.8}$$

where $\odot$ denotes the Hadamard product. For the special case with $N = 2$, $T = 3$, we have

$$\boldsymbol{S}_{23}^{1} \boldsymbol{Y}_{23}^{(l)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \end{pmatrix},$$

and

$$\boldsymbol{\iota}_2 \otimes \boldsymbol{S}_3^{1} \boldsymbol{\iota}_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix},$$

therefore,

$$\left(\boldsymbol{S}_{23}^{1} \boldsymbol{Y}_{23}^{(l)}\right) \odot \left(\boldsymbol{\iota}_2 \otimes \boldsymbol{S}_3^{1} \boldsymbol{\iota}_3\right) = \begin{pmatrix} 0 \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \end{pmatrix} \odot \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ y_{11} \\ y_{12} \\ 0 \\ y_{21} \\ y_{22} \end{pmatrix} = \boldsymbol{L}^{1} \boldsymbol{Y}_{23}^{(l)}.$$

Note that

$$\boldsymbol{I}_2 \otimes \boldsymbol{S}_3^{1} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{3-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 \\ y_{11} \\ y_{12} \\ 0 \\ y_{21} \\ y_{22} \end{pmatrix},$$

hence we can alternatively write

$$\left(\boldsymbol{S}^{j}_{NT}\boldsymbol{Y}^{(l)}_{NT}\right) \odot \left(\boldsymbol{\iota}_N \otimes \boldsymbol{S}^{j}_{T}\boldsymbol{\iota}_T\right) = \left[\boldsymbol{I}_N \otimes \boldsymbol{S}^{j}_{T}\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{T-j} \end{pmatrix}\right] \boldsymbol{S}^{j}_{NT}\boldsymbol{Y}^{(l)}_{NT}, \tag{6.9}$$

which gives us a concise full-sample matrix formula for the jth panel lag operator

$$\boldsymbol{L}^{j} = \left(\boldsymbol{I}_N \otimes \boldsymbol{S}^{j}_{T}\boldsymbol{D}_j\right) \boldsymbol{S}^{j}_{NT}, \tag{6.10}$$

where

$$\boldsymbol{D}_j \equiv \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{T-j} \end{pmatrix}.$$

Note we can write an S/R function that will compute panel lags as

```
lag.panel <- function(x, p, n, t)
{
    d1 <- power.shift(n * t, p) % * % x
    d2 <- kronecker(rep(1, n), c(rep(NA, p), rep(1, t-p)))
    d1 * d2
}
```

We can now construct a "long" panel data set as

$$\mathbf{X}_L \equiv \left[\mathrm{vec}\,\mathbf{X}'_{GDP} : \left(\mathbf{I}_N \otimes \mathbf{S}^1_T\mathbf{D}_1\right)\mathbf{S}^1_{NT}\mathrm{vec}\,\mathbf{X}'_{GDP} : \mathrm{vec}\,\mathbf{X}'_b : \mathrm{vec}\,\mathbf{X}'_i\right]. \tag{6.11}$$

Note we used the panel lag operator $\left(\mathbf{I}_N \otimes \mathbf{S}^j_T\mathbf{D}_j\right)\mathbf{S}^j_{NT}$ only to lag GDP variable once as this is the only lagged variable we use in the model. Our data matrix for the special case with $T = 5$ is thus given by

$$\boldsymbol{X}_L = \begin{pmatrix}
GDP^{n=1}_{t=1} & 0 & b^{n=1}_{t=1} & i^{n=1}_{t=1} \\
GDP^{n=1}_{t=2} & GDP^{n=1}_{t=1} & b^{n=1}_{t=2} & i^{n=1}_{t=2} \\
\vdots & \vdots & \vdots & \vdots \\
GDP^{n=1}_{t=5} & GDP^{n=1}_{t=4} & b^{n=1}_{t=T} & i^{n=1}_{t=T} \\
\vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
GDP^{n=N}_{t=1} & 0 & b^{n=N}_{t=1} & i^{n=N}_{t=1} \\
GDP^{n=N}_{t=2} & GDP^{n=N}_{t=1} & b^{n=N}_{t=2} & i^{n=N}_{t=2} \\
\vdots & \vdots & \vdots & \vdots \\
GDP^{n=N}_{t=5} & GDP^{n=N}_{t=4} & b^{n=N}_{t=T} & i^{n=N}_{t=T}
\end{pmatrix}. \tag{6.12}$$

Thus, the model (6.3) can be written in full-sample notation as

$$\begin{aligned}\mathrm{vec}\,\boldsymbol{X}'_{GDP} &= \alpha + \beta_1 \mathrm{vec}\,\boldsymbol{X}'_b + \beta_2 \mathrm{vec}\,\boldsymbol{X}'_i + \gamma_1 \mathrm{vec}\,\boldsymbol{X}'_{GDP} + \boldsymbol{\zeta} \\ &= \boldsymbol{Z\theta} + \boldsymbol{\zeta}\end{aligned}$$

where $\boldsymbol{Z} \equiv (\boldsymbol{\iota} : \mathrm{vec}\,\boldsymbol{X}'_b : \mathrm{vec}\,\boldsymbol{X}'_{GDP})$ and $\boldsymbol{\theta} \equiv (\alpha, \beta_1, \beta_2, \gamma_1)'$. A standard econometric approach to estimate $\boldsymbol{\theta}$ would be by OLS using the formula

$$\hat{\boldsymbol{\theta}}_{OLS} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\mathrm{vec}\,\boldsymbol{X}'_{GDP}. \tag{6.13}$$

Estimating the single equation model (6.3) with OLS using PcGive, with data in the "long panel" format, gives the following estimates

$$GDP_t = \alpha + \underset{(0.025)}{0.121}\, b_t - \underset{(0.0136)}{0.032}\, i_t - \underset{(0.077)}{0.122}\, GDP_{t-1}$$

Such model can, of course, be estimated using LISREL language using the following simple programme

```
TI
DA NI=4 NO=225 NG=1 MA=CM
CM=longpanel.CM
LA
g b i g_1
SE
1 2 3 4 /
MO NX=2 NY=2 BE=FU,FI GA=FU,FI PH=SY,FR PS=DI,FR
FR BE(1,2) GA(1,1) GA(1,2) GA(2,1) GA(2,2)
OU ME=ML ND=5
```

where the maximum likelihood would give the OLS estimates. Estimates obtained from LISREL are virtually identical, namely

$$GDP_t = \underset{(0.001)}{-0.001} + \underset{(0.025)}{0.121}\, b_t + \underset{(0.0317)}{0.032}\, i_t - \underset{(0.077)}{0.122}\, GDP_{t-1}$$

We can note that the simple naive model using a single mis-measured financial development indicator leads to the conclusion that financial development (proxied by b_t) has a positive and significant effect on economic growth. We will reconsider this conclusion later one after estimating the model with multiple indicators of financial development explicitly treated as a latent variable.

To conclude the comparison of the two approaches to the analysis of panel data we can point out that the use of standard SEM software such as LISREL for the estimation of simple single-equation models, such as (6.3), has no advantage over the classical econometric approach and software. Namely, such model can be easily estimated with OLS (or IV if desired) using the "long panel" data vector of length NT. However, the comparison is a useful exercise that illuminates the specifics of each approach. It is easy to see that the reason why SEM software packages, such as LISREL, need the "wide" approach is because a sufficient statistic in form of the covariance matrix is needed for discrepancy function estimation implemented in LISREL (minimising distance between theoretical and empirical covariances), which cannot be be computed from data matrices made out of "long panel" NT-vectors.

We will now compare these two approaches in the case when more then two equations are estimated simultaneously using full-information methods. While we showed that single equations without measurement errors can be equivalently estimated in both "long" and "wide" approaches, where the "long" approach leads to the simple OLS estimator, the two approaches might lead to different results when systems estimators are used.

Estimation of simultaneous equations with cross-equation restrictions is the area where SEM software has a strong advantage even for models without latent variables (or measurement error) when the "wide panel" approach is taken. This is simply because standard SEM software packages allow cross-equation restrictions in respect to all components of the model, including error variances, and not only the regression coefficients. On the other hand, pooling the data into the "long panel" format and estimating a model that does not need to impose any cross-equation restrictions might be simpler to implement, but it could lead to different results. We will illustrate these issues on a real-data example.

Consider a simple two-equation simultaneous equation model

$$GDP_t = \alpha_0 + \alpha_1 b_t - \alpha_2 i_t - \alpha_3 GDP_{t-1} \tag{6.14}$$

$$b_t = \alpha_4 + \alpha_5 i_t - \alpha_6 GDP_{t-1}, \tag{6.15}$$

which simply adds to (6.3) the equation for financial development as a function of the initial growth and lagged economic development. The "wide panel" SEM approach can be illustrated by specifying (6.14)–(6.15) as a LISREL model

$$
\underbrace{\begin{pmatrix} GDP_{t=2} \\ b_{t=2} \\ GDP_{t=3} \\ b_{t=3} \\ GDP_{t=4} \\ b_{t=4} \\ GDP_{t=5} \\ b_{t=4} \end{pmatrix}}_{\boldsymbol{\eta}} = \underbrace{\begin{pmatrix} 0 & \beta_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{31} & 0 & 0 & \beta_{34} & 0 & 0 & 0 & 0 \\ \beta_{41} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{53} & 0 & 0 & \beta_{56} & 0 & 0 \\ 0 & 0 & \beta_{63} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{75} & 0 & 0 & \beta_{78} \\ 0 & 0 & 0 & 0 & \beta_{85} & 0 & 0 & 0 \end{pmatrix}}_{\boldsymbol{B}} \underbrace{\begin{pmatrix} GDP_{t=2} \\ b_{t=2} \\ GDP_{t=3} \\ b_{t=3} \\ GDP_{t=4} \\ b_{t=4} \\ GDP_{t=5} \\ b_{t=5} \end{pmatrix}}_{\boldsymbol{\eta}}
+ \underbrace{\begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{33} & 0 & 0 \\ 0 & 0 & \gamma_{43} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{54} & 0 \\ 0 & 0 & 0 & \gamma_{64} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{78} \\ 0 & 0 & 0 & 0 & \gamma_{85} \end{pmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix} GDP_{t=1} \\ i_{t=2} \\ i_{t=3} \\ i_{t=4} \\ i_{t=5} \end{pmatrix}}_{\boldsymbol{\xi}} + \underbrace{\begin{pmatrix} \zeta_{1t=2} \\ \zeta_{2t=2} \\ \zeta_{1t=3} \\ \zeta_{2t=3} \\ \zeta_{1t=4} \\ \zeta_{2t=4} \\ \zeta_{1t=5} \\ \zeta_{2t=5} \end{pmatrix}}_{\boldsymbol{\zeta}}, \tag{6.16}
$$

and impose cross-equation equality constraints on regression coefficients and error variances as follows

$$
\begin{aligned}
\beta_{(31)} &= \beta_{(53)} = \beta_{(75)} = \gamma_{(11)} \\
\beta_{(41)} &= \beta_{(63)} = \beta_{(85)} = \gamma_{(21)} \\
\beta_{(12)} &= \beta_{(34)} = \beta_{(56)} = \beta_{(78)} \\
\gamma_{(12)} &= \gamma_{(33)} = \gamma_{(54)} = \gamma_{(75)} \\
\gamma_{(22)} &= \gamma_{(43)} = \gamma_{(64)} = \gamma_{(85)} \\
\psi_{(11)} &= \psi_{(33)} = \psi_{(55)} = \psi_{(77)} \\
\psi_{(22)} &= \psi_{(44)} = \psi_{(66)} = \psi_{(88)}
\end{aligned}
$$

Estimation in LISREL using maximum likelihood yields the following results

$$
\begin{aligned}
GDP_t &= \underset{(0.012)}{0.122}\, b_t - \underset{(0.015)}{0.043}\, i_t - \underset{(0.039)}{0.133}\, GDP_{t-1} \\
b_t &= -\underset{(0.047)}{0.275}\, i_t - \underset{(0.112)}{1.152}\, GDP_{t-1},
\end{aligned}
$$

which is obtained by running the LISREL syntax that imposes cross-equation restrictions by specifying an equation for each time point, namely

```
TI
DA NI=35 NO=225 NG=1 MA=CM
CM=widepanel.CM
SE
```

```
8 10 15 17 22 24 29 31 1 14 21 28 35 /
MO NX=5 NY=8 BE=FU,FI GA=FU,FI PH=SY,FR PS=DI,FR
FR BE(1,2) BE(3,1) BE(3,4) BE(4,1) BE(5,3) BE(5,6) BE(6,3)
FR BE(7,5) BE(7,8) BE(8,5) GA(1,1) GA(1,2) GA(2,1) GA(2,2)
FR GA(3,3) GA(4,3) GA(5,4) GA(6,4) GA(7,5) GA(8,5)
EQ GA(1,2) GA(3,3) GA(5,4) GA(7,5)
EQ BE(3,1) BE(5,3) BE(7,5) GA(1,1)
EQ BE(4,1) BE(6,3) BE(8,5) GA(2,1)
EQ BE(1,2) BE(3,4) BE(5,6) BE(7,8)
EQ GA(2,2) GA(4,3) GA(6,4) GA(8,5)
EQ PS(1,1) PS(3,3) PS(5,5) PS(7,7) PH(1,1)
EQ PS(2,2) PS(4,4) PS(6,6) PS(8,8)
EQ PH(2,2) PH(3,3) PH(4,4) PH(5,5)
OU ME=ML ND=5
```

noting that we specified $NT = 225$ as the sample size instead of $N = 45$ to make it comparable with the the sample size used in the "long panel" estimation.

The same model can be estimated using the "long panel" approach by computing

$$\boldsymbol{S}_L = (N-1)^{-1}\boldsymbol{X}_L\boldsymbol{X'}_L$$

as an input matrix for LISREL. Estimating (6.16) by LISREL using maximum likelihood yields the estimates

$$\begin{aligned} GDP_t &= \underset{(0.012)}{0.122}\, b_t - \underset{(0.032)}{0.032}\, i_t - \underset{(0.077)}{0.122}\, GDP_{t-1} \\ b_t &= -\underset{(0.093)}{0.299}\, i_t - \underset{(0.219)}{1.053}\, GDP_{t-1}, \end{aligned}$$

which was produced by running a considerably simpler LISREL syntax that does not specify the model for each time point and hence does no impose any cross-equation restrictions, i.e.,

```
TI
DA NI=4 NO=225 NG=1 MA=CM
CM=longpanel.CM
LA
g b i g_1
SE
1 2 3 4 /
MO NX=2 NY=2 BE=FU,FI GA=FU,FI PH=SY,FR PS=DI,FR
FR BE(1,2) GA(1,1) GA(1,2) GA(2,1) GA(2,2)
OU ME=ML ND=5
```

Here we have used the 4×4 $\boldsymbol{S}_L$ as an input covariance matrix "longpanel.CM", letting LISREL treat it as a 'sufficient statistic' covariance matrix. The results are similar to those obtained before by using the 15×15 $\boldsymbol{S}_L$ matrix ("widepanel.CM").

Similar results can be obtained using PcFiml or any other package for systems estimation that uses full-sample maximum likelihood rather then sufficient statistics such as $\boldsymbol{S}_L$ or $\boldsymbol{S}_W$. This conclusion changes when latent variables are present, as we will show next. Before proceeding, we should additionally note that we encountered no convergence problems while estimating the above simple models. It turns out that this also will change with inclusion of latent variables and complication of the model, which calls for starting values different from those automatically generated by LISREL.

Empirical modelling

We consider the the measurement model for the latent financial development by using the observable indicators b_t, p_t, and l_t. Beck et al. (2000), for example, run three different sets of growth regressions using b_t, p_t, and l_t, which importantly assumes that these three indicators indeed measure financial development. A factor-analytic interpretation of the first assumption is that these indicators measure a single latent variable (factor) or that a single latent variable accounts for the observed correlations among b_t, p_t, and l_t. To this end we specify the following measurement model

$$\begin{pmatrix} b_t \\ p_t \\ l_t \end{pmatrix} = \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{pmatrix} F_t + \begin{pmatrix} \varepsilon_t^{(b)} \\ \varepsilon_t^{(p)} \\ \varepsilon_t^{(l)} \end{pmatrix}, \tag{6.17}$$

where the measurement error covariance matrix is of the form

$$\boldsymbol{\Theta}_\varepsilon = \begin{pmatrix} \theta_{11}^{(b)} & 0 & 0 \\ 0 & \theta_{22}^{(p)} & 0 \\ 0 & 0 & \theta_{33}^{(l)} \end{pmatrix}. \tag{6.18}$$

We allow a third-order autocorrelation process in I_t, which can be specified as[5]

$$\sum_{j=0}^{3} \left(\boldsymbol{S}_5^j \otimes \boldsymbol{\Phi}_j\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \phi_1 & 1 & 0 & 0 & 0 \\ \phi_2 & \phi_1 & 1 & 0 & 0 \\ \phi_3 & \phi_2 & \phi_1 & 1 & 0 \\ 0 & \phi_3 & \phi_2 & \phi_1 & 1 \end{pmatrix}. \tag{6.19}$$

This specification implies that the observable indicators measure a single latent variable over the entire sample period. Correlated measurement errors are not permitted but (6.19) allows fairly general dynamics in the exogenous latent variable process.

As the first step, we estimate only the measurement model (6.17) obtaining the maximum likelihood estimates reported in table 6.2. This was achieved using

[5]We only specify the lower triangular of this autocorrelation matrix due to symmetry.

LISREL 8.54 with default starting values (i.e. computed by the programme), which encountered no convergence problems.

The estimated coefficients (table 6.2) are all of the same sign and statistically significant. The overall fit of the model, however, is rather poor with the χ^2 fit statistic nearly five times greater than its degrees of freedom parameter. This brings in question the empirical results based on the separate growth regressions, but it also calls for considerable extension of the FD-growth research framework in the direction of searching for additional or better FD indicators. Recalling the example of the Hali et al. (2002) study where we showed how dropping a single indicator can considerably improve the fit of the model, the search for better indicators might be awarding in this case too. Another immediate implication for the empirical literature would be in using formal statistical procedures for the assessment of the measurement models as tools for selecting the observable indicators rather then guiding the selection only on the substantive grounds.

Table 6.2: FD measurement model estimates

	All countries		Developed countries		Developing countries	
$\boldsymbol{\theta}_i$	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
λ_{11}	0.018	(0.003)	0.007	(0.002)	0.026	(0.005)
λ_{21}	0.063	(0.005)	0.077	(0.009)	0.053	(0.006)
λ_{31}	0.096	(0.007)	0.106	(0.012)	0.076	(0.008)
$\sigma^2_{\varepsilon_1}$	0.004	(0.000)	0.001	(0.000)	0.006	(0.001)
$\sigma^2_{\varepsilon_2}$	0.003	(0.000)	0.003	(0.000)	0.002	(0.000)
$\sigma^2_{\varepsilon_3}$	0.006	(0.001)	0.007	(0.001)	0.004	(0.001)
ϕ_1	0.023	(0.012)	0.023	(0.018)	0.022	(0.017)
ϕ_2	−0.682	(0.031)	−0.662	(0.041)	−0.691	(0.044)
ϕ_3	−0.671	(0.031)	−0.652	(0.042)	−0.678	(0.045)
χ^2	543.489		266.492		333.820	
d.f.	111		111		111	

Next we divided the countries into developed and developing (Table 6.3), considering the possibility that these two groups of possibly quite different countries might have differently measured financial development.

The estimates in Table 6.2 indeed suggest that separate models fit better. The error variances and autocovariances of the latent variable F_t are fairly close between the two groups, though some differences can be observed in the factor loadings, which might be one of the sources of the improved fit. Namely, it seems that l_t (value of credits by financial intermediaries to the private sector) has greater weight in measuring financial development for developed countries, while the opposite holds for b_t (ratio of domestic and domestic plus central bank credit).

Finally, we aim to estimate the full DPSEM model including economic growth

Table 6.3: Country groups

Developed countries		Developing countries		
Australia	UK	Cameroon	Kenya	Syria
Austria	Greece	Colombia	Korea	Thailand
Belgium	Ireland	Costa Rica	Sri Lanka	Trinidad & T.
Canada	Italy	Ecuador	Malaysia	Venezuela
Switzerland	Japan	Egypt	Pakistan	South Africa
Germany	Netherlands	Ghana	Philippines	–
Denmark	Norway	Guatemala	Papua N.G.	–
Spain	New Zealand	Honduras	Rwanda	–
Finland	Sweden	India	Senegal	–
France	USA	Jamaica	El Salvador	–

and an additional exogenous control variable, the *initial GDP per capita.* The first equation is a dynamic FD-growth relationship, which includes lagged economic growth, while the second equation accounts for the possible feedback from the lagged growth back to the current financial development. The DSEM specification is of the model is as follows. Structural model is given by

$$\underbrace{\begin{pmatrix} G_t \\ F_t \end{pmatrix}}_{\boldsymbol{\eta}_t} = \underbrace{\begin{pmatrix} 0 & \beta_{12}^{(0)} \\ 0 & 0 \end{pmatrix}}_{\boldsymbol{B}_0} \underbrace{\begin{pmatrix} G_t \\ F_t \end{pmatrix}}_{\boldsymbol{\eta}_t} + \underbrace{\begin{pmatrix} \beta_{11}^{(1)} & 0 \\ \beta_{21}^{(1)} & 0 \end{pmatrix}}_{\boldsymbol{B}_1} \underbrace{\begin{pmatrix} G_{t-1} \\ F_{t-1} \end{pmatrix}}_{\boldsymbol{\eta}_{t-1}} + \underbrace{\begin{pmatrix} \gamma_{11}^{(0)} \\ \gamma_{21}^{(0)} \end{pmatrix}}_{\boldsymbol{\Gamma}_0} \underbrace{I_t}_{\boldsymbol{\xi}_t} + \underbrace{\begin{pmatrix} \zeta_t^{(G)} \\ \zeta_t^{(F)} \end{pmatrix}}_{\boldsymbol{\zeta}_t}, \tag{6.20}$$

while the measurement model assumes that *economic growth* (G_t) and *initial GDP* (I_t) are measured without error, while the financial development is measured by the same three observable indicators as before,

$$\underbrace{\begin{pmatrix} g_t \\ b_t \\ p_t \\ l_t \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \end{pmatrix}}_{\boldsymbol{\Lambda}_y} \underbrace{\begin{pmatrix} G_t \\ F_t \end{pmatrix}}_{\boldsymbol{\eta}_t} + \underbrace{\begin{pmatrix} 0 \\ \varepsilon_t^{(b)} \\ \varepsilon_t^{(p)} \\ \varepsilon_t^{(l)} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}. \tag{6.21}$$

$$\underbrace{i_t}_{\boldsymbol{x}_t} = \underbrace{I_t}_{\boldsymbol{\xi}_t} \tag{6.22}$$

This specification postulates a possible FD-growth effect, while in the same time it considers the alternative explanation, namely that higher levels of financial development occur in those countries which had higher economic growth in the recent past (i.e. over the past five years period). The parameter matrices to be estimated are specified as follows

$$\boldsymbol{B}_0 = \begin{pmatrix} 0 & \beta_{12}^{(0)} \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{B}_1 = \begin{pmatrix} \beta_{11}^{(1)} & 0 \\ \beta_{21}^{(1)} & 0 \end{pmatrix}, \quad \boldsymbol{\Gamma}_0 = \begin{pmatrix} \gamma_{11}^{(0)} \\ \gamma_{21}^{(0)} \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} \psi_{11}^{(G)} & 0 \\ 0 & \psi_{22}^{(F)} \end{pmatrix},$$

$$\sum_{j=0}^{3} \left(\mathbf{S}_5^j \otimes \boldsymbol{\Phi}_j\right) = \begin{pmatrix} \phi_0 & 0 & 0 & 0 & 0 \\ \phi_1 & \phi_0 & 0 & 0 & 0 \\ \phi_2 & \phi_1 & \phi_0 & 0 & 0 \\ \phi_3 & \phi_2 & \phi_1 & \phi_0 & 0 \\ 0 & \phi_3 & \phi_2 & \phi_1 & \phi_0 \end{pmatrix}, \quad \boldsymbol{\Theta}_\varepsilon = \begin{pmatrix} \theta_{11}^{(b)} & 0 & 0 \\ 0 & \theta_{11}^{(p)} & 0 \\ 0 & 0 & \theta_{11}^{(l)} \end{pmatrix}.$$

The above specification utilised the DSEM parameter matrices which require recursive specification of the relationship using the "t-notation". Estimation in LISREL requires re-casting of the DSEM model in the form of a LISREL model, which can be best illustrated in the standard LISREL path diagram. Using LISREL-style notation ($b_t \equiv y_t^{(1)}$, $p_t \equiv y_t^{(2)}$, $l_t \equiv y_t^{(3)}$, $g_t \equiv z_t$, $i_t \equiv x_t$, $G_t \equiv \eta_t^{(1)}$, $F_t \equiv \eta_t^{(2)}$, $I_t \equiv \xi_t$) the above model can be represented with the path diagram shown in Figure 6.3.

The LISREL syntax corresponding to the path diagram in Figure 6.3 is given in Appendix §6B. However, running this syntax with the default starting values generated by LISREL does not lead to convergence, which appears to be the case even after several thousand iterations. Some experimentation with random or arbitrary starting values equally lead to convergence failure.

The starting values that enable fast convergence of the LISREL's algorithm can, nevertheless, be obtained by using the instrumental variables estimates obtained by estimating the model in its OF form using the methods outlined in chapter §5. To obtain the OF-IV estimates we need to re-write the DSEM model (6.20)–(6.21) in the observed form. Using l_t as a unit-loading indicator (proxy) for F_t, the structural equation model can be written as

$$\begin{pmatrix} g_t \\ l_t - \varepsilon_t^{(l)} \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12}^{(0)} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} g_t \\ l_t - \varepsilon_t^{(l)} \end{pmatrix} + \begin{pmatrix} \beta_{11}^{(1)} & 0 \\ \beta_{21}^{(1)} & 0 \end{pmatrix} \begin{pmatrix} g_{t-1} \\ l_{t-1} - \varepsilon_{t-1}^{(l)} \end{pmatrix} + \begin{pmatrix} \gamma_{11}^{(0)} \\ \gamma_{21}^{(0)} \end{pmatrix} i_t + \begin{pmatrix} \zeta_t^{(G)} \\ \zeta_t^{(F)} \end{pmatrix},$$

which can be re-written as

$$\begin{pmatrix} g_t \\ l_t \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12}^{(0)} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} g_t \\ l_t \end{pmatrix} + \begin{pmatrix} \beta_{11}^{(1)} & 0 \\ \beta_{21}^{(1)} & 0 \end{pmatrix} \begin{pmatrix} g_{t-1} \\ l_{t-1} \end{pmatrix} + \begin{pmatrix} \gamma_{11}^{(0)} \\ \gamma_{21}^{(0)} \end{pmatrix} i_t + \begin{pmatrix} \zeta_t^{(G)} - \beta_{12}^{(0)} \varepsilon_t^{(l)} \\ \zeta_t^{(F)} + \varepsilon_t^{(l)} \end{pmatrix}.$$

Collecting the measurement errors that appear in each equation, we can determine which lagged errors are uncorrelated, i.e.,

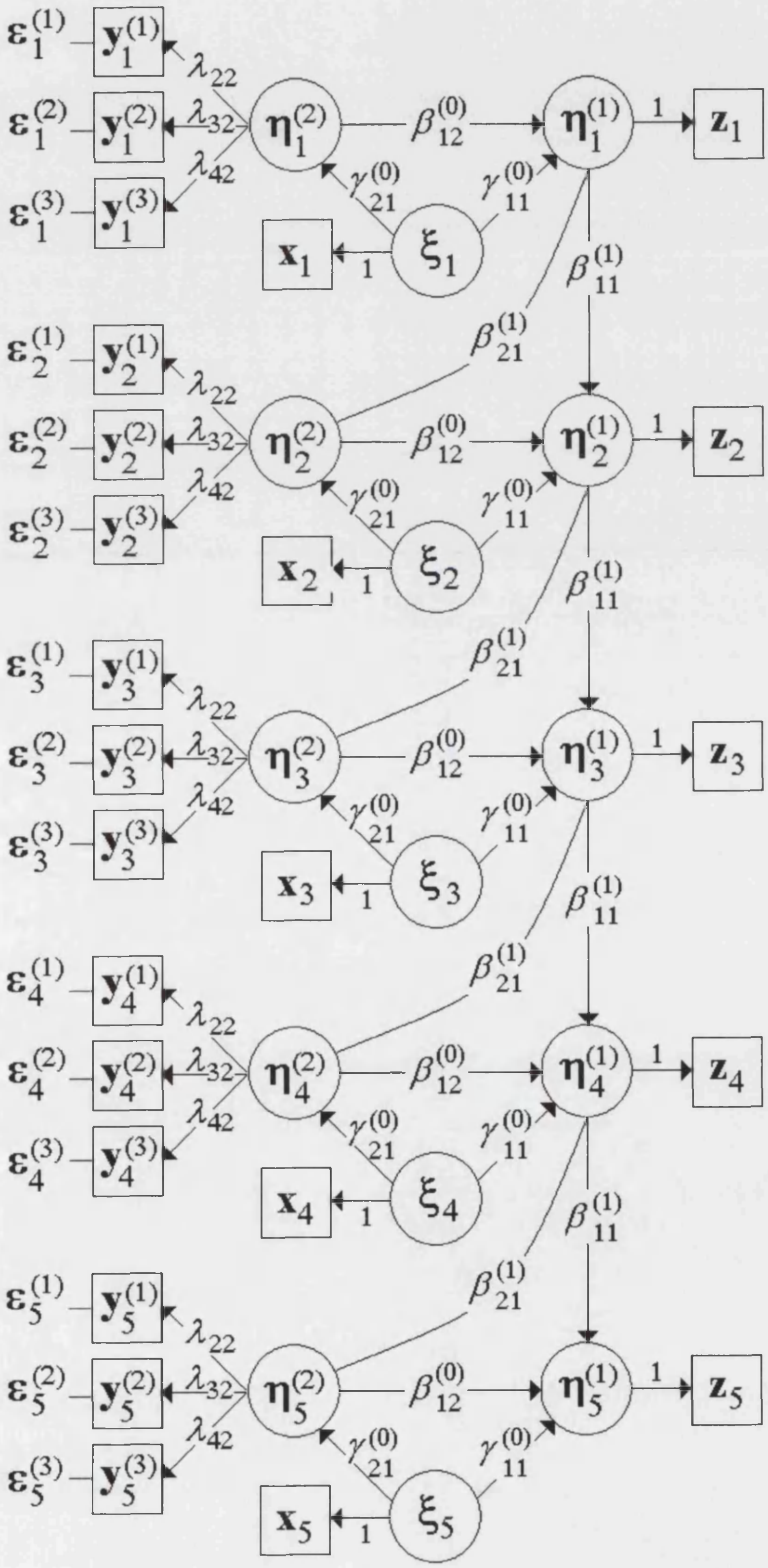

Figure 6.3: FD path diagram

which implies that, e.g., $\varepsilon_t^{(l)}$ appears in the first structural equation, and it is uncorrelated with $\varepsilon_{t-j}^{(b)}$, $\varepsilon_{t-j}^{(p)}$, and $\varepsilon_{t-1-j}^{(l)}$ for $j = 1, 2, \ldots$. We can use similar notation to identify the model-implied valid instruments as

$$\begin{pmatrix} g_t \\ l_t \end{pmatrix} \rightleftarrows \begin{pmatrix} b_{t-j}^{(IV)}, p_{t-j}^{(IV)}, l_{t-1-j}^{(IV)} \\ b_{t-j}^{(IV)}, p_{t-j}^{(IV)}, l_{t-1-j}^{(IV)} \end{pmatrix}, \quad j \geq 0,$$

which indicates that, e.g., $b_{t-j}^{(IV)}$, $p_{t-j}^{(IV)}$, and $l_{t-1-j}^{(IV)}$ for $j \geq 0$ are valid instruments in the equation for l_t. It is easy to see that in the case of the two variables that are assumed to be measured without error, g_{t-j} and i_{t-j} are valid instruments for all j.

The measurement model (6.21) can be written in the observed form by substituting l_t for F_t as

$$\begin{pmatrix} g_t \\ b_t \\ p_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \end{pmatrix} \begin{pmatrix} g_t \\ l_t \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_t^{(b)} - \lambda_{22}\varepsilon_t^{(l)} \\ \varepsilon_t^{(p)} - \lambda_{32}\varepsilon_t^{(l)} \end{pmatrix}.$$

Lagged errors that should be uncorrelated with the measurement errors appearing in the equations are

$$\begin{pmatrix} 0 \\ \varepsilon_t^{(b)} - \lambda_{22}\varepsilon_t^{(l)} \\ \varepsilon_t^{(p)} - \lambda_{32}\varepsilon_t^{(l)} \end{pmatrix} \rightleftarrows \begin{pmatrix} \varepsilon_{t-j}^{(g)}, \varepsilon_{t-j}^{(b)}, \varepsilon_{t-j}^{(p)}, \varepsilon_{t-j}^{(l)} \\ \varepsilon_{t-j}^{(g)}, \varepsilon_{t-1-j}^{(b)}, \varepsilon_{t-j}^{(p)}, \varepsilon_{t-1-j}^{(l)} \\ \varepsilon_{t-j}^{(g)}, \varepsilon_{t-j}^{(b)}, \varepsilon_{t-1-j}^{(p)}, \varepsilon_{t-1-j}^{(l)} \end{pmatrix}, \quad j \geq 0,$$

which implies the following lagged indicators as valid instruments

$$\begin{pmatrix} g_t \\ b_t \\ p_t \end{pmatrix} \rightleftarrows \begin{pmatrix} g_{t-j}^{(IV)}, b_{t-j}^{(IV)}, p_{t-j}^{(IV)}, l_{t-j}^{(IV)} \\ g_{t-j}^{(IV)}, b_{t-1-j}^{(IV)}, p_{t-j}^{(IV)}, l_{t-1-j}^{(IV)} \\ g_{t-j}^{(IV)}, b_{t-j}^{(IV)}, p_{t-1-j}^{(IV)}, l_{t-1-j}^{(IV)} \end{pmatrix}, \quad j \geq 0. \tag{6.23}$$

Of course, instrument validity is an empirical issue and we should proceed by testing it by considering different sets of the model-implied instruments. We construct the instruments for an indicator $\boldsymbol{x}_t$ by applying the instrument generating operator $\boldsymbol{S}_{IV}^j$ defined in (5.20) thus constructing the instruments as $\boldsymbol{S}_{IV}^j \boldsymbol{x}_t$ for $j \geq i$ where i is the minimum lag length implied by (6.23). The panel lags are computed with the S+ function `panel.lag()`, and the panel instruments are computed with the S+ function `instrument()`, both defined in Appendix §6A.

Selection of instruments in practice is often a subjective process based on trial and error, where an instrumental validity statistic, such as the Sargan (1988) test (5.27), is used to asses empirical validity of a particular instrument set. However, given our intention to use the OF-IV estimates as starting values for the maximum likelihood estimation we need simple rules for testing alternative sets of instruments. If the instrument-selection rules can be made automatic, there would be a potential for programming an algorithm that can generate OF-IV starting values automatically.

We suggest the following simple method for grouping the potential instruments. First, divide all eligible instruments into groups of instruments lagged once, twice, and three times. We restrict the maximum lag length to three here given $T = 5$. For example, using only the indicators of the latent financial development as potential instruments, we can define the j-th lagged set of instruments as

$$IV_j = \left(\boldsymbol{S}_{IV}^j \boldsymbol{b}_t : \boldsymbol{S}_{IV}^j \boldsymbol{p}_t : \boldsymbol{S}_{IV}^j \boldsymbol{l}_t\right). \tag{6.24}$$

It is possible to consider sets of instruments at given lag lengths combined together, which aids the simplicity and automatisation of the selection procedure. We will consider a minor complication in adding instruments based on g_t and i_t necessary to decrease the Sargan test statistic. As a simple decision rule for selection of the "best" set of instruments we suggest using a tradeoff of the minimised Sargan's χ^2 test criterion and the minimum regression standard error ($\hat{\sigma}$). The results of the IV tests shown in Tables 6.4 and 6.5 were obtained by the GiveWin 1.30 using the batch code given in Appendix §6B

Table 6.4: IV validity tests: Structural equations

Structural equation g_t			
Instruments sets	IV validity χ^2	d.f.	$\hat{\sigma}$
IV_1, IV_2, IV_3	20.805	8	0.02054
IV_1, IV_2	15.748	5	0.02070
IV_1, IV_3	13.794	5	0.02047
IV_2, IV_3	14.140	5	0.02047
IV_1	9.388	2	0.02066
IV_2 ✓	2.176	2	0.02772
IV_3	5.161	2	0.02060
Structural equation l_t			
Instruments sets	IV validity χ^2	d.f.	$\hat{\sigma}$
IV_1, IV_2, IV_3	86.678	12	0.144074
IV_1, IV_2	69.606	7	0.149933
IV_1, IV_3	67.744	7	0.151216
IV_2, IV_3	39.697	8	0.15143
IV_1	39.628	2	0.175956
IV_2* ✓	3.834	3	0.141764
IV_3	27.407	3	0.167316

Using the selected instrument sets (marked with ✓ in Tables 6.4, 6.5) we obtain the GIVE and FIVE (3SLS) estimates reported in Table 6.6. Using the GIVE estimates as starting values in LISREL enabled convergence in only 22 iterations, with the resulting ML estimates given in Table 6.6. Using FIVE estimates as starting

Table 6.5: IV validity tests: Measurement equations

Measurement equation for b_t			
Instruments sets	IV validity χ^2	d.f.	$\hat{\sigma}$
IV_1, IV_2, IV_3	42.246	8	0.06326
IV_1, IV_2	40.753	5	0.06335
IV_1, IV_3	9.328	5	0.06347
IV_2, IV_3	36.393	5	0.06323
IV_1	6.578	2	0.06402
IV_2	28.398	2	0.06998
IV_3 ✓	5.690	2	0.06289
Measurement equation for p_t			
Instruments sets	IV validity χ^2	d.f.	$\hat{\sigma}$
IV_1, IV_2, IV_3	76.452	8	0.061783
IV_1, IV_2	64.454	5	0.062349
IV_1, IV_3	54.493	5	0.061419
IV_2, IV_3	59.895	5	0.061671
IV_1	34.122	2	0.062206
IV_2 ✓	12.946	2	0.073943
IV_3	44.668	2	0.061336

* $g_j^{(IV)}$ and $i_j^{(IV)}$ included in the instruments set.

values enabled LISREL to converge in 18 iterations, however it was necessary to set the $\beta_{11}^{(1)}$ coefficient to zero, as the programme failed to converge using the FIVE estimate $\hat{\beta}_{11}^{(1)} = -0.1270$, which is of the wrong sign. The LISREL syntax along with the syntax that specifies starting values that use FIVE estimates is given in Appendix §6B.

Estimation of the DPSEM model (6.20)–(6.21) by maximum likelihood produces the estimates reported in table 6.6, which also reports the GIVE and FIVE estimates. It is notable that GIVE estimates have larger standard errors then either ML or FIVE estimates.

In addition to the full-sample estimates, we estimated two separate models for the sub-samples of developed and developing countries, with the results shown in Table 6.7.

Similarly to the results obtained above for the measurement model alone, the full model (6.20)–(6.21) fits considerably better in the two sub-samples than in the overall sample. The apparent lack of the close fit might be due to departures from normality, which is not an ignorable issue with small samples such as this. Thus, we test the normality of the model residuals (see figures 6.4 and 6.5).[6] Using the Doornik and Hansen (1994) normality test we obtain the normality χ^2 statistics with

[6]The residuals here refer to the differences between the corresponding elements of the fitted and observed covariance matrix.

Table 6.6: FD model estimates

	FIML		GIVE		FIVE (3SLS)	
Parameter	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
$\beta_{12}^{(0)}$	-0.0077	(0.0114)	-0.0068	(0.0210)	-0.0089	(0.0185)
$\beta_{11}^{(1)}$	0.0035	(0.0747)	0.0063	(0.0769)	-0.1270	(0.0706)
$\beta_{21}^{(1)}$	0.9841	(0.4560)	-0.5262	(2.0922)	0.4001	(0.5054)
$\gamma_{11}^{(0)}$	-0.0800	(0.0267)	-0.0742	(0.0370)	-0.0560	(0.0333)
$\gamma_{21}^{(0)}$	-0.7402	(0.1564)	-0.2413	(0.8133)	-0.8658	(0.2148)
λ_{22}**	0.1569	(0.0429)	0.0754	(0.0620)	0.1818	(0.0541)
λ_{32}	0.4349	(0.0808)	0.7326	(0.2672)	0.5246	(0.0536)
λ_{42}*	1.0000	–	1.0000	–	1.0000	–
$\theta_{11}^{(b)}$	0.0039	(0.0004)	0.0040	–	0.0041	–
$\theta_{22}^{(p)}$	0.0042	(0.0004)	0.0055	–	0.0039	–
$\theta_{33}^{(l)}$	0.0005	(0.0006)	–	–	–	–
$\psi_{11}^{(G)}$	0.0138	(0.0029)	0.0201	–	0.0189	–
$\psi_{22}^{(F)}$	0.0004	(0.0000)	0.0004	–	0.0004	–
ϕ_0**	0.0026	(0.0003)	–	–	–	–
ϕ_1**	-0.0004	(0.0001)	–	–	–	–
ϕ_2**	-0.0015	(0.0002)	–	–	–	–

* Fixed parameter.
** Sample (auto)covariance for GIVE and FIVE.

2 d.f. of 30.584, 2.840, and 49.816 for the full sample, developed, and developing countries' models, respectively. Clearly, we cannot reject the normality only for the model estimated with the sample of developed countries, hence caution is needed in interpreting the χ^2 fit statistics reported in table 6.7.

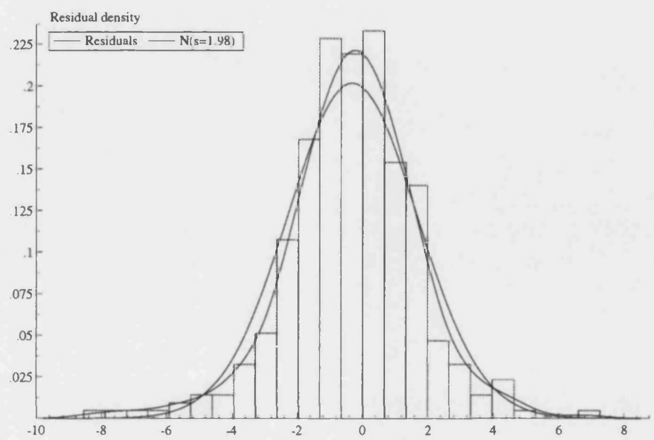

Figure 6.4: Density plot of the standardised residuals: Overall sample

Table 6.7: Sub-sample estimates

	Developed countries		Developing countries	
Parameter	Estimate	(SE)	Estimate	(SE)
$\beta_{12}^{(0)}$	0.0020	(0.0057)	−0.0044	(0.0253)
$\beta_{11}^{(1)}$	−0.2229	(0.1012)	0.0383	(0.1011)
$\beta_{21}^{(1)}$	5.0018	(1.7091)	0.4468	(0.4007)
$\gamma_{11}^{(0)}$	−0.2275	(0.0402)	−0.0742	(0.0392)
$\gamma_{21}^{(0)}$	−0.3048	(0.7068)	−0.5530	(0.1465)
λ_{22}	0.0387	(0.0277)	0.2901	(0.0898)
λ_{32}	0.4634	(0.1673)	0.4660	(0.1052)
λ_{42}	1.0000	–	1.0000	–
$\theta_{11}^{(b)}$	0.0011	(0.0001)	0.0056	(0.0008)
$\theta_{22}^{(p)}$	0.0057	(0.0011)	0.0029	(0.0006)
$\theta_{33}^{(l)}$	−0.0006	(0.0044)	0.0010	(0.0020)
$\psi_{11}^{(G)}$	0.0181	(0.0072)	0.0087	(0.0023)
$\psi_{22}^{(F)}$	0.0001	(0.0000)	0.0006	(0.0001)
ϕ_0	0.0004	(0.0001)	0.0038	(0.0005)
ϕ_1	−0.0001	(0.0000)	−0.0007	(0.0001)
ϕ_2	−0.0002	(0.0000)	−0.0020	(0.0003)
χ^2	310		310	
$d.f.$	622.6845		620.1287	

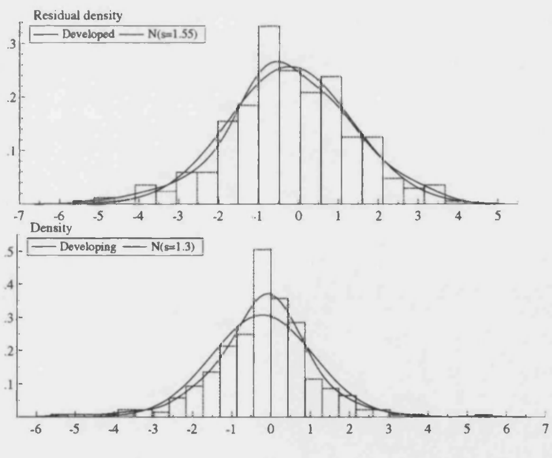

Figure 6.5: Density plot of the standardised residuals: Sub-samples

Conclusion

Despite the normality issues, the results contrast the mainstream empirical FD-growth literature. The first is a clear difference between the models for the two groups of countries, which suggest a more elaborative substantive theory should be developed to explain the FD-growth relationship relative to the level of development of the analysed countries. The second finding is that financial development has no significant impact on growth ($\beta_{12}^{(0)}$), while lagged growth has strong positive impact on the current financial development ($\beta_{11}^{(1)}$), which equally holds in the full sample as well as in the two sub-samples, separately. We also find that initial capital significantly affects both growth and financial development in the overall sample, but its effect on growth diminishes for the developed countries, while its effect on financial development is insignificant for the developing countries. The coefficients of the measurement model are similar to those estimated before, with generally significant loadings and error variances. We note that the smallest error variance belongs to y_3 (credit to private sector), which suggests that this indicator might be somewhat better then the other two.

6.3 Application II: UK micro consumption model

The relationship between consumption expenditure and personal disposable income is a widely researched topic in the empirical economics literature. The cornerstone of most of the debates has been the relationship between consumption and income in relation to the permanent income hypothesis (Friedman, 1957) and the life-cycle hypothesis (for a review see Deaton (1992)). This theoretical framework predicts a relationship between permanent income (annuity of the life-cycle income) and consumption, but does not predict strong relationship between current income and consumption. In theory, rational consumers should not respond to windfall gains and temporary income increase in increased consumption, rather their consumption should be smooth across the life-cycle, which is achieved by borrowing when income is low and repaying the debts when income increases (e.g., later in life). However, preponderance of the empirical studies using either micro or macro data reject the permanent income hypothesis insofar they find strong and statistically significant relationship between current income and consumption. Such empirical finding is known as excess sensitivity of consumption (Hayashi 1982, Campbell and Mankiw 1989, Campbell and Mankiw 1990, Campbell and Mankiw 1991, Deaton 1992, Browning and Lusardi 1996, Madsen and McAleer 2000, Madsen and McAleer 2001).

There are several theoretical explanations for the excess sensitivity of consumption in the literature. These explanations can be classified into three main groups, the liquidity constraint approach (Flavin 1981, Hubbard and Judd 1986, Jappelli and Pagano 1989, Scheinkman and Weiss 1986), the uncertainty hypothesis (Blanchard and Fischer 1989, Zeldes 1989, Deaton 1991, Aiyagari 1994, Muellbauer and Lattimore 1995, Carroll 1997, Ludvigson and Paxson 2001, Hahm and Steigerwarld 1999, Gourinchas and Parker 2002) and the behavioural life-cycle hypothesis (Madsen and McAleer 2001).

An important question is what happens to income elasticity (i.e. coefficient of income in the regression of consumption on income) if the consumption function is estimated as a latent-variable model.

The statistical explanation of excess sensitivity finding might rest in the effect of contemporaneous correlations among income and consumption indicators on the relationship between income and consumption itself. Larsen, (2002) suggested to estimate the latent total consumption in a household aiming at improving the accuracy of permanent income studies. He noted that, while the sum of individual expenditures (in a household) is an unbiased estimator of latent total household consumption, it is also in-optimal as such sum is an unweighted sum of components that contain measurement error. It can be added that not all expenditures are always reported, thus even if we accept to operate with an unweighted sum, such variable might still not resemble total household consumption expenditure, and a similar thing holds for income (where non-reporting of some types of incomes is a well known problem in household surveys). Larsen (2002) derived an alternative estimator of total household consumption, based on latent variable methods, that is unbiased and variance minimising. Essentially, by estimating a latent-type of model for household consumption, Larsen (2002) derived weights for various considered types of consumption expenditures (including also non-expenditure indicators). This line of research extended the previous efforts of estimating household consumption more precisely though still relying on total purchase expenditure (Kay et al. 1984, Aasness et al. 1993, Aasness et al. 1995).

Empirical studies that attempt to model the income-consumption relationship using latent variable techniques are scarce in the literature. There were few attempts to use the latent variable methods for this problem, primarily due to dynamic nature of the income-consumption relationship and inability of the typical covariance structure based models to handle data with pronounced dynamic component. For example, Ventura and Satorra (1998) use Spanish household data to estimate life-cycle effects on some product expenditures with only two years of data.

We will estimate a latent consumption function model that incorporates possible liquidity constraints effects using micro data from the British Household Panel Study survey and incorporating data for the 13 currently available waves (years).

Data and variables

The data for this empirical analysis comes from the British Household Panel Survey (Taylor 2005), which has 13 waves (years) of data available. For a number of variables all 13 years can be merged into a joint panel. The available variables on consumption expenditure and types of income, as well as potential liquidity constraints indicators vary across waves, and as our primary purpose is to illustrate dynamic latent variable modelling using data with pronounced time-series dimension we make a compromise by using only variables that were available across all 13 waves.

Specifically, we are forced to give up otherwise relevant durable expenditure data that are available only for the last six waves. The variables that we use in the model (with original BHPS codes) are shown in Table 6.8. By a BHPS convention, the variable codes are prefixed by wave identifiers a, $b, \ldots, m$. Household data (expenditures) were first spread onto individual level, and subsequently combined with the individual level income data, thus creating all-individual data files for all waves. Finally, wave-specific files were merged into a joint panel for all individuals across all waves in the both "long panel" and "wide panel" formats.

The data is a panel of 5,152 individuals observed over 13 years, hence $NT =$

Table 6.8: BHPS variables used in the model

BHPS code	Description
HSIZE	Number of persons in household
XPHSD1	Housing payments required borrowing
XPHSD2	Housing payments required cutbacks
XPHSDB	Been 2+ months late with housing payment
XPHSDF	Problems paying four housing over the year
XPFOOD	Total weekly food and grocery bill
XPHSG	Gross monthly housing costs
QFACHI	Highest academic qualification
HGEMP	Employment status
FYRL	Annual labour income
FIYRNL	Annual non-labour income
FIYRI	Annual investment income
SAVED	Amount saved each month

Table 6.9: Data transformation and variable names

Symbol	Description	Transformation
f_t	Annual personal food expenditure	52*(XPFOOD/HSIZE)
h_t	Annual personal housing costs	12*(XPHSG/HSIZE)
i_t	Annual labour income	FYRL
i_t	Annual investment income	FIYR
s_t	Annual personal savings	(FYRL + FIYRNL + FIYRI - XPFOOD - XPHSG)
r_t	Cumulative credit repayment problem	(YPHSD1 + XPHSD2 + XPHSDB + XPHSDF + HGEMP[a])
e_t	Highest level of academic education	GFACHI

[a] HGEMP was recoded so that 1 = working; 0 = not working

66,976. Data transformations used to create substantively relevant quantities are specified in Table 6.9. We create observable indicators for three latent variables: consumption, income, and liquidity constraints. The variable transformations allowed us to compute all quantities on annual, individual level. The variable *cumulative credit repayment problem* (r_t) was created in an attempt to extract the information on possible credit constraints; it sums indicators (0 and 1) of several types of credit repayment problems, thus cumulating to total number of credit difficulties (see Table 6.9). Personal savings is another variable frequently used in liquidity constraints modelling (Hayashi 1982) where individuals (or households) with positive saving rates are assumed to be liquidity un-constrained. Note that values for all variables were created on the same (annual) level thus the monthly (XPHSG) was multiplied by 12, and the weekly food expenditure variable (XPFOOD) was multiplied by 52. The expenditures were further divided by household size (HSIZE) to obtain an approximate estimate for individual household members.

Empirical modelling

We consider a dynamic structural equation model that relates current consumption, modelled as a latent variable, to current (latent) income, past consumption, past income, and current and past liquidity constraints. The model also relates current income to past income and education, which is assumed to be time-invariant and measured without error. Finally, the (latent) liquidity constraints are also modelled, and are assumed to depend on its own past as well as on the past consumption. Table 6.10 lists latent variables and their observable indicators.

Table 6.10: Latent and observable variables

Latent variable	Symbol	Observable indicators
Consumption	C_t	f_t, h_t
Income	I_t	l_t, i_t
Liquidity constraints	L_t	s_t, r_t

We suppose that excessive spending in one year causes greater degree of liquidity constraints in the following year(s). This model can be specified as a special case of the PDSEM model

$$\boldsymbol{\eta}_t = \sum_{j=0}^{5} \boldsymbol{B}_j \boldsymbol{\eta}_{t-j} + \boldsymbol{\Gamma}_0 \boldsymbol{\xi}_t + \boldsymbol{\zeta}_t \tag{6.25}$$

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_y \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t \tag{6.26}$$

$$\boldsymbol{x}_t = \boldsymbol{\xi}_t + \boldsymbol{\delta}_t \tag{6.27}$$

The measurement model(s) for the latent variables C_t, L_t, and I_t are specified as

$$\underbrace{\begin{pmatrix} f_t \\ h_t \\ l_t \\ i_t \\ s_t \\ r_t \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ \lambda_{21}^{(h)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{42}^{(i)} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{63}^{(r)} \end{pmatrix}}_{\boldsymbol{\Lambda}_y} \underbrace{\begin{pmatrix} C_t \\ I_t \\ L_t \end{pmatrix}}_{\boldsymbol{\eta}_t} + \underbrace{\begin{pmatrix} \varepsilon_t^{(f)} \\ \varepsilon_t^{(h)} \\ \varepsilon_t^{(l)} \\ \varepsilon_t^{(i)} \\ \varepsilon_t^{(s)} \\ \varepsilon_t^{(r)} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \tag{6.28}$$

while the single-indicator education variable E_t has the measurement model

$$\underbrace{e_t}_{\boldsymbol{x}_t} = \underbrace{E_t}_{\boldsymbol{\xi}_t} + \underbrace{\delta_t^{(e)}}_{\boldsymbol{\delta}_t}. \tag{6.29}$$

The structural part of the model describes the relationships among the latent variables and is specified as

$$
\begin{aligned}
\underbrace{\begin{pmatrix} C_t \\ L_t \\ I_t \end{pmatrix}}_{\boldsymbol{\eta}_t} &= \underbrace{\begin{pmatrix} 0 & \beta_{12}^{(0)} & \beta_{13}^{(0)} \\ 0 & 0 & \beta_{23}^{(0)} \\ 0 & 0 & 0 \end{pmatrix}}_{\boldsymbol{B}_0} \underbrace{\begin{pmatrix} C_t \\ L_t \\ I_t \end{pmatrix}}_{\boldsymbol{\eta}_t} + \underbrace{\begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} & \beta_{13}^{(1)} \\ \beta_{21}^{(1)} & \beta_{22}^{(1)} & \beta_{23}^{(1)} \\ 0 & 0 & \beta_{33}^{(1)} \end{pmatrix}}_{\boldsymbol{B}_1} \underbrace{\begin{pmatrix} C_{t-1} \\ L_{t-1} \\ I_{t-1} \end{pmatrix}}_{\boldsymbol{\eta}_{t-1}} \\
&+ \underbrace{\begin{pmatrix} \beta_{11}^{(2)} & \beta_{12}^{(2)} & \beta_{13}^{(2)} \\ \beta_{21}^{(2)} & \beta_{22}^{(2)} & \beta_{23}^{(2)} \\ 0 & 0 & \beta_{33}^{(2)} \end{pmatrix}}_{\boldsymbol{B}_2} \underbrace{\begin{pmatrix} C_{t-2} \\ L_{t-2} \\ I_{t-2} \end{pmatrix}}_{\boldsymbol{\eta}_{t-2}} + \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_{22}^{(3)} & \beta_{23}^{(3)} \\ 0 & 0 & \beta_{33}^{(3)} \end{pmatrix}}_{\boldsymbol{B}_3} \underbrace{\begin{pmatrix} C_{t-3} \\ L_{t-3} \\ I_{t-3} \end{pmatrix}}_{\boldsymbol{\eta}_{t-3}} \\
&+ \underbrace{\begin{pmatrix} \beta_{11}^{(4)} & 0 & 0 \\ 0 & \beta_{22}^{(4)} & 0 \\ 0 & 0 & \beta_{33}^{(4)} \end{pmatrix}}_{\boldsymbol{B}_4} \underbrace{\begin{pmatrix} C_{t-4} \\ L_{t-4} \\ I_{t-4} \end{pmatrix}}_{\boldsymbol{\eta}_{t-4}} + \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{33}^{(5)} \end{pmatrix}}_{\boldsymbol{B}_5} \underbrace{\begin{pmatrix} C_{t-5} \\ L_{t-5} \\ I_{t-5} \end{pmatrix}}_{\boldsymbol{\eta}_{t-5}} \\
&+ \underbrace{\begin{pmatrix} 0 \\ 0 \\ \gamma_{31}^{(0)} \end{pmatrix}}_{\boldsymbol{\Gamma}_0} \underbrace{E_t}_{\boldsymbol{\xi}_t} + \underbrace{\begin{pmatrix} \zeta_t^C \\ \zeta_t^L \\ \zeta_t^I \end{pmatrix}}_{\boldsymbol{\zeta}_t}, \qquad (6.30)
\end{aligned}
$$

The covariance matrix of the measurement errors does not permit correlation (and autocorrelation) among measurement errors and is thus specified as

$$
\boldsymbol{\Theta}_\varepsilon = \begin{pmatrix} \theta_{11}^{(f)} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{22}^{(h)} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{33}^{(l)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{44}^{(i)} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{55}^{(s)} & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{66}^{(r)} \end{pmatrix}. \qquad (6.31)
$$

Finally, the covariance matrix of the errors in the structural equations is specified as diagonal, i.e.,

$$
\boldsymbol{\Psi} = \begin{pmatrix} \psi_{11}^{(C)} & 0 & 0 \\ 0 & \psi_{22}^{(L)} & 0 \\ 0 & 0 & \psi_{33}^{(I)} \end{pmatrix}. \qquad (6.32)
$$

While the above (DPSEM) specification is relatively simple, specifying the same model in LISREL syntax is exceptionally tedious due to the necessity to specify all relations for each of the 13 time periods separately and impose cross-equation equality restrictions. To illustrate the drawbacks of LISREL specification consider the path diagram of the above model specified as LISREL path diagram. Figure 6.6 shows only one half of the full path diagram for the this model, i.e., for the first 7 time periods only, whereas the specification for the remaining periods remains the same.

Evidently, the descriptive clarity of SEM models' graphical representation by the means of path diagrams is largely lost for dynamic SEM models specified as LISREL path diagrams. Moreover, the LISREL package fails to generate or interactively build path diagrams beyond certain size and complexity, the diagram in figure 6.6 is an example of a path diagram that cannot be handles by LISREL, though the programme was capable of estimating the coefficients when good starting values were provided.

As in the previous section, we firstly estimate the OF model. The observed form of the structural part of the model is

$$
\begin{aligned}
\begin{pmatrix} f_t - \varepsilon_t^{(f)} \\ l_t - \varepsilon_t^{(l)} \\ s_t - \varepsilon_t^{(s)} \end{pmatrix} &= \begin{pmatrix} 0 & \beta_{12}^{(0)} & \beta_{13}^{(0)} \\ 0 & 0 & \beta_{23}^{(0)} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_t - \varepsilon_t^{(f)} \\ l_t - \varepsilon_t^{(l)} \\ s_t - \varepsilon_t^{(s)} \end{pmatrix} + \begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} & \beta_{13}^{(1)} \\ \beta_{21}^{(1)} & \beta_{22}^{(1)} & \beta_{23}^{(1)} \\ 0 & 0 & \beta_{33}^{(1)} \end{pmatrix} \begin{pmatrix} f_{t-1} - \varepsilon_{t-1}^{(f)} \\ l_{t-1} - \varepsilon_{t-1}^{(l)} \\ s_{t-1} - \varepsilon_{t-1}^{(s)} \end{pmatrix} \\
&+ \begin{pmatrix} \beta_{11}^{(2)} & \beta_{12}^{(2)} & \beta_{13}^{(2)} \\ \beta_{21}^{(2)} & \beta_{22}^{(2)} & \beta_{23}^{(2)} \\ 0 & 0 & \beta_{33}^{(2)} \end{pmatrix} \begin{pmatrix} f_{t-2} - \varepsilon_{t-2}^{(f)} \\ l_{t-2} - \varepsilon_{t-2}^{(l)} \\ s_{t-2} - \varepsilon_{t-2}^{(s)} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_{22}^{(3)} & \beta_{23}^{(3)} \\ 0 & 0 & \beta_{33}^{(3)} \end{pmatrix} \begin{pmatrix} f_{t-3} - \varepsilon_{t-3}^{(f)} \\ l_{t-3} - \varepsilon_{t-3}^{(l)} \\ s_{t-3} - \varepsilon_{t-3}^{(s)} \end{pmatrix} \\
&+ \begin{pmatrix} \beta_{11}^{(4)} & 0 & 0 \\ 0 & \beta_{22}^{(4)} & 0 \\ 0 & 0 & \beta_{33}^{(4)} \end{pmatrix} \begin{pmatrix} f_{t-4} - \varepsilon_{t-4}^{(f)} \\ l_{t-4} - \varepsilon_{t-4}^{(l)} \\ s_{t-4} - \varepsilon_{t-4}^{(s)} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{33}^{(5)} \end{pmatrix} \begin{pmatrix} f_{t-5} - \varepsilon_{t-5}^{(f)} \\ l_{t-5} - \varepsilon_{t-5}^{(l)} \\ s_{t-5} - \varepsilon_{t-5}^{(s)} \end{pmatrix} \\
&+ \begin{pmatrix} 0 \\ 0 \\ \gamma_{31}^{(0)} \end{pmatrix} E_t + \begin{pmatrix} \zeta_t^C \\ \zeta_t^L \\ \zeta_t^I \end{pmatrix},
\end{aligned} \tag{6.33}
$$

hence by collecting the terms, we can see that the composite additional error terms due to measurement error is given by

$$
\begin{pmatrix} \beta_{12}^{(0)} \varepsilon_t^{(l)} + \beta_{13}^{(0)} \varepsilon_t^{(s)} + \beta_{11}^{(1)} \varepsilon_{t-1}^{(f)} + \beta_{12}^{(1)} \varepsilon_{t-1}^{(l)} + \beta_{13}^{(1)} \varepsilon_{t-1}^{(s)} + \beta_{11}^{(2)} \varepsilon_{t-2}^{(f)} + \beta_{12}^{(2)} \varepsilon_{t-2}^{(l)} + \beta_{13}^{(2)} \varepsilon_{t-2}^{(s)} + \beta_{11}^{(4)} \varepsilon_{t-4}^{(f)} + \varepsilon_t^{(f)} \\ \beta_{23}^{(0)} \varepsilon_t^{(l)} + \beta_{21}^{(1)} \varepsilon_{t-1}^{(f)} + \beta_{22}^{(1)} \varepsilon_{t-1}^{(l)} + \beta_{23}^{(1)} \varepsilon_{t-1}^{(s)} + \beta_{21}^{(2)} \varepsilon_{t-2}^{(f)} + \beta_{22}^{(2)} \varepsilon_{t-2}^{(l)} + \beta_{23}^{(2)} \varepsilon_{t-2}^{(s)} + \beta_{22}^{(3)} \varepsilon_{t-3}^{(l)} + \beta_{23}^{(3)} \varepsilon_{t-3}^{(s)} + \beta_{22}^{(4)} \varepsilon_{t-4}^{(l)} + \varepsilon_t^{(l)} \\ \beta_{33}^{(1)} \varepsilon_{t-1}^{(s)} + \beta_{33}^{(2)} \varepsilon_{t-2}^{(s)} + \beta_{33}^{(3)} \varepsilon_{t-3}^{(s)} + \beta_{33}^{(4)} \varepsilon_{t-4}^{(s)} + \beta_{33}^{(5)} \varepsilon_{t-5}^{(s)} + \varepsilon_t^{(s)} \end{pmatrix},
$$

thus (6.33) can be rewritten by separating and collecting the errors as

$$
\begin{aligned}
\begin{pmatrix} f_t \\ l_t \\ s_t \end{pmatrix} &= \begin{pmatrix} 0 & \beta_{12}^{(0)} & \beta_{13}^{(0)} \\ 0 & 0 & \beta_{23}^{(0)} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_t \\ l_t \\ s_t \end{pmatrix} + \begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} & \beta_{13}^{(1)} \\ \beta_{21}^{(1)} & \beta_{22}^{(1)} & \beta_{23}^{(1)} \\ 0 & 0 & \beta_{33}^{(1)} \end{pmatrix} \begin{pmatrix} f_{t-1} \\ l_{t-1} \\ s_{t-1} \end{pmatrix} \\
&+ \begin{pmatrix} \beta_{11}^{(2)} & \beta_{12}^{(2)} & \beta_{13}^{(2)} \\ \beta_{21}^{(2)} & \beta_{22}^{(2)} & \beta_{23}^{(2)} \\ 0 & 0 & \beta_{33}^{(2)} \end{pmatrix} \begin{pmatrix} f_{t-2} \\ l_{t-2} \\ s_{t-2} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_{22}^{(3)} & \beta_{23}^{(3)} \\ 0 & 0 & \beta_{33}^{(3)} \end{pmatrix} \begin{pmatrix} f_{t-3} \\ l_{t-3} \\ s_{t-3} \end{pmatrix} \\
&+ \begin{pmatrix} \beta_{11}^{(4)} & 0 & 0 \\ 0 & \beta_{22}^{(4)} & 0 \\ 0 & 0 & \beta_{33}^{(4)} \end{pmatrix} \begin{pmatrix} f_{t-4} \\ l_{t-4} \\ s_{t-4} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{33}^{(5)} \end{pmatrix} \begin{pmatrix} f_{t-5} \\ l_{t-5} \\ s_{t-5} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \gamma_{31}^{(0)} \end{pmatrix} e_t \\
&- \begin{pmatrix} 0 & \beta_{12}^{(0)} & \beta_{13}^{(0)} \\ 0 & 0 & \beta_{23}^{(0)} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_t^{(f)} \\ \varepsilon_t^{(l)} \\ \varepsilon_t^{(s)} \end{pmatrix} - \begin{pmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} & \beta_{13}^{(1)} \\ \beta_{21}^{(1)} & \beta_{22}^{(1)} & \beta_{23}^{(1)} \\ 0 & 0 & \beta_{33}^{(1)} \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1}^{(f)} \\ \varepsilon_{t-1}^{(l)} \\ \varepsilon_{t-1}^{(s)} \end{pmatrix} \\
&- \begin{pmatrix} \beta_{11}^{(2)} & \beta_{12}^{(2)} & \beta_{13}^{(2)} \\ \beta_{21}^{(2)} & \beta_{22}^{(2)} & \beta_{23}^{(2)} \\ 0 & 0 & \beta_{33}^{(2)} \end{pmatrix} \begin{pmatrix} \varepsilon_{t-2}^{(f)} \\ \varepsilon_{t-2}^{(l)} \\ \varepsilon_{t-2}^{(s)} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_{22}^{(3)} & \beta_{23}^{(3)} \\ 0 & 0 & \beta_{33}^{(3)} \end{pmatrix} \begin{pmatrix} \varepsilon_{t-3}^{(f)} \\ \varepsilon_{t-3}^{(l)} \\ \varepsilon_{t-3}^{(s)} \end{pmatrix} \\
&- \begin{pmatrix} \beta_{11}^{(4)} & 0 & 0 \\ 0 & \beta_{22}^{(4)} & 0 \\ 0 & 0 & \beta_{33}^{(4)} \end{pmatrix} \begin{pmatrix} \varepsilon_{t-4}^{(f)} \\ \varepsilon_{t-4}^{(l)} \\ \varepsilon_{t-4}^{(s)} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{33}^{(5)} \end{pmatrix} \begin{pmatrix} \varepsilon_{t-5}^{(f)} \\ \varepsilon_{t-5}^{(l)} \\ \varepsilon_{t-5}^{(s)} \end{pmatrix} \\
&+ \begin{pmatrix} \varepsilon_t^{(f)} \\ \varepsilon_t^{(l)} \\ \varepsilon_t^{(s)} \end{pmatrix} + \begin{pmatrix} \zeta_t^C \\ \zeta_t^L \\ \zeta_t^I \end{pmatrix}
\end{aligned} \tag{6.34}
$$

hence the uncorrelated error components are

$$
\begin{pmatrix}
\overbrace{\varepsilon_t^{(l)}, \varepsilon_{t-1}^{(l)}, \varepsilon_{t-2}^{(l)}}^{\varepsilon_{t-j}^{(l)},\ j=0,1,2}, \overbrace{\varepsilon_t^{(f)}, \varepsilon_{t-1}^{(f)}, \varepsilon_{t-2}^{(f)}, \varepsilon_{t-4}^{(f)}}^{\varepsilon_{t-j}^{(f)},\ j=0,1,2,3,4}, \overbrace{\varepsilon_t^{(s)}, \varepsilon_{t-1}^{(s)}, \varepsilon_{t-2}^{(s)}}^{\varepsilon_{t-j}^{(s)},\ j=0,1,2} \\
\underbrace{\varepsilon_t^{(l)}, \varepsilon_{t-1}^{(l)}, \varepsilon_{t-2}^{(l)}, \varepsilon_{t-3}^{(l)}, \varepsilon_{t-4}^{(l)}}_{\varepsilon_{t-j}^{(l)},\ j=0,1,2,3,4}, \underbrace{\varepsilon_{t-1}^{(f)}, \varepsilon_{t-2}^{(f)}}_{\varepsilon_{t-j}^{(f)},\ j=1,2}, \underbrace{\varepsilon_{t-1}^{(s)}, \varepsilon_{t-2}^{(s)}, \varepsilon_{t-3}^{(s)}}_{\varepsilon_{t-j}^{(s)},\ j=1,2,3} \\
\underbrace{\varepsilon_t^{(s)}, \varepsilon_{t-1}^{(s)}, \varepsilon_{t-2}^{(s)}, \varepsilon_{t-3}^{(s)}, \varepsilon_{t-4}^{(s)}, \varepsilon_{t-5}^{(s)}}_{\varepsilon_{t-j}^{(s)},\ j=0,1,2,3,4,5}
\end{pmatrix}
\rightleftarrows
\underbrace{\begin{pmatrix}
\varepsilon_{t-5-j}^{(f)}, \varepsilon_{t-j}^{(h)}, \varepsilon_{t-3-j}^{(l)}, \varepsilon_{t-j}^{(i)}, \varepsilon_{t-3-j}^{(s)}, \varepsilon_{t-j}^{(r)} \\
\varepsilon_{t-3-j}^{(f)}, \varepsilon_{t-j}^{(h)}, \varepsilon_{t-5-j}^{(l)}, \varepsilon_{t-j}^{(i)}, \varepsilon_{t-4-j}^{(s)}, \varepsilon_{t-j}^{(r)} \\
\varepsilon_{t-j}^{(f)}, \varepsilon_{t-j}^{(h)}, \varepsilon_{t-j}^{(l)}, \varepsilon_{t-j}^{(i)}, \varepsilon_{t-6-j}^{(s)}, \varepsilon_{t-j}^{(r)}
\end{pmatrix}}_{j \geq 0},
$$

which implies the following valid instruments for the structural equations

$$
\begin{pmatrix} f_t \\ l_t \\ s_t \end{pmatrix} \rightleftarrows \begin{pmatrix}
f_{t-5-j}^{(IV)}, h_{t-j}^{(IV)}, l_{t-3-j}^{(IV)}, i_{t-j}^{(IV)}, s_{t-3-j}^{(IV)}, r_{t-j}^{(IV)} \\
f_{t-3-j}^{(IV)}, h_{t-j}^{(IV)}, l_{t-5-j}^{(IV)}, i_{t-j}^{(IV)}, s_{t-4-j}^{(IV)}, r_{t-j}^{(IV)} \\
f_{t-j}^{(IV)}, h_{t-j}^{(IV)}, l_{t-j}^{(IV)}, i_{t-j}^{(IV)}, s_{t-6-j}^{(IV)}, r_{t-j}^{(IV)}
\end{pmatrix}.
$$

The observed form of the measurement model is

$$\begin{pmatrix} h_t \\ i_t \\ r_t \end{pmatrix} = \begin{pmatrix} \lambda_{21}^{(h)} & 0 & 0 \\ 0 & \lambda_{42}^{(i)} & 0 \\ 0 & 0 & \lambda_{53}^{(r)} \end{pmatrix} \begin{pmatrix} f_t - \varepsilon_t^{(f)} \\ l_t - \varepsilon_t^{(l)} \\ s_t - \varepsilon_t^{(s)} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(h)} \\ \varepsilon_t^{(i)} \\ \varepsilon_t^{(r)} \end{pmatrix} \tag{6.35}$$

$$= \begin{pmatrix} \lambda_{21}^{(h)} & 0 & 0 \\ 0 & \lambda_{42}^{(i)} & 0 \\ 0 & 0 & \lambda_{53}^{(r)} \end{pmatrix} \begin{pmatrix} f_t \\ l_t \\ s_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(h)} - \lambda_{21}^{(h)} \varepsilon_t^{(f)} \\ \varepsilon_t^{(i)} - \lambda_{42}^{(i)} \varepsilon_t^{(l)} \\ \varepsilon_t^{(r)} - \lambda_{53}^{(r)} \varepsilon_t^{(s)} \end{pmatrix}, \tag{6.36}$$

hence the uncorrelated lagged errors are

$$\begin{pmatrix} \varepsilon_t^{(h)} - \lambda_{21}^{(h)} \varepsilon_t^{(f)} \\ \varepsilon_t^{(i)} - \lambda_{42}^{(i)} \varepsilon_t^{(l)} \\ \varepsilon_t^{(r)} - \lambda_{53}^{(r)} \varepsilon_t^{(s)} \end{pmatrix} \rightleftarrows \begin{pmatrix} \varepsilon_{t-1-j}^{(h)}, \varepsilon_{t-1-j}^{(f)}, \varepsilon_{t-i}^{(i)}, \varepsilon_{t-i}^{(l)}, \varepsilon_{t-i}^{(r)}, \varepsilon_{t-i}^{(s)} \\ \varepsilon_{t-i}^{(h)}, \varepsilon_{t-i}^{(f)}, \varepsilon_{t-1-j}^{(i)}, \varepsilon_{t-1-j}^{(l)}, \varepsilon_{t-i}^{(r)}, \varepsilon_{t-i}^{(s)} \\ \varepsilon_{t-i}^{(h)}, \varepsilon_{t-i}^{(f)}, \varepsilon_{t-i}^{(i)}, \varepsilon_{t-i}^{(l)}, \varepsilon_{t-1-j}^{(r)}, \varepsilon_{t-1-j}^{(s)} \end{pmatrix}, \quad j \geq 0,$$

which implies the following instruments

$$\begin{pmatrix} h_t \\ i_t \\ r_t \end{pmatrix} \rightleftarrows \begin{pmatrix} h_{t-1-j}^{(IV)}, f_{t-1-j}^{(IV)}, i_{t-i}^{(IV)}, l_{t-i}^{(IV)}, r_{t-i}^{(IV)}, s_{t-i}^{(IV)} \\ h_{t-i}^{(IV)}, f_{t-i}^{(IV)}, i_{t-1-j}^{(IV)}, l_{t-1-j}^{(IV)}, r_{t-i}^{(IV)}, s_{t-i}^{(IV)} \\ h_{t-i}^{(IV)}, f_{t-i}^{(IV)}, i_{t-i}^{(IV)}, l_{t-i}^{(IV)}, r_{t-1-j}^{(IV)}, s_{t-1-j}^{(IV)} \end{pmatrix}, \quad j \geq 0$$

We estimate the OF model (6.34) using the GIVE and FIVE methods suggested in chapter §5, using the instruments listed in table 6.11. As in the previous section, we use symbols for instruments such as $f_i^{(IV)}$ to denote f_t lagged 4 periods. The particular choices of instruments in table 6.11 are made by considering various combinations of eligible instruments and choosing the instruments sets that minimise the Sargan's validity of instruments χ^2 test. These instruments are selected for each equation separately, estimated with the limited-information GIVE methods. Hence, particular instruments sets are chosen for equations where h_t, i_t, etc. are endogenous variables.

The same approach to selection of instruments cannot be taken on the equation-by-equation basis when FIVE methods are used to estimate multiple equations jointly. Therefore, we chose the union of instruments sets to estimate multiple equations, where instruments used in GIVE estimation for each individual equation are used together in FIVE estimation.

However, the instruments sets used for estimation of measurement and structural equations are notably different thus we estimate these two sets of equations separately using the FIVE methods. While the entire model including both measurement and structural equations could be estimated jointly, we chose to estimate the two sets of equations separately aiming to improve the validity of instruments. The estimation code, with the instruments used in both GIVE and FIVE estimation is given in Appendix §6.

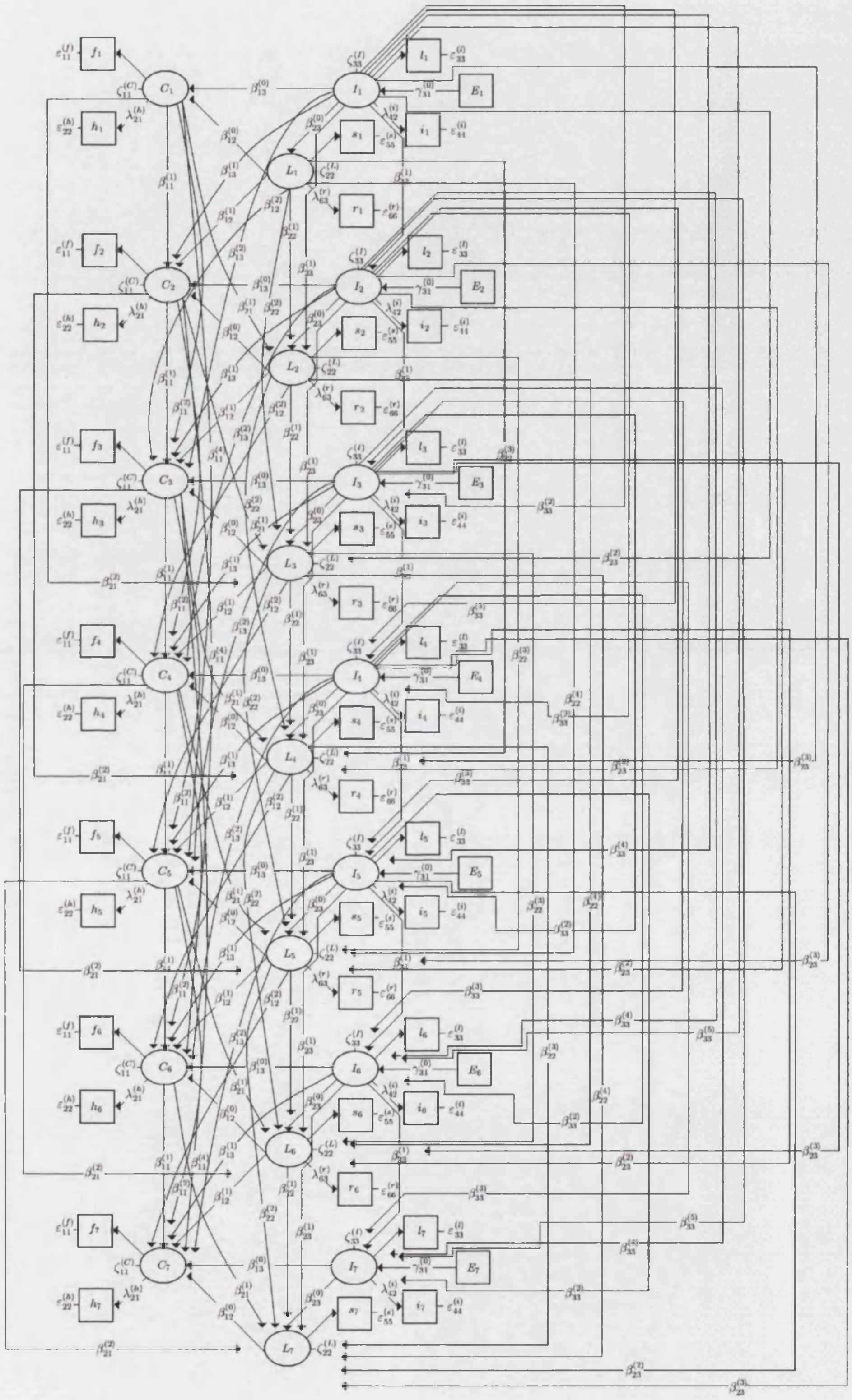

Figure 6.6: BHPS model path diagram

GIVE and FIVE estimates of the model (6.34) are reported in tables 6.12 and 6.13. Both sets of estimates can be used as starting values in LISREL maximum

Table 6.11: IV tests: BHPS model

Equation	Instruments	χ^2	d.f.	$\hat{\sigma}$
h_t	$f_4^{(IV)}, f_5^{(IV)}$	0.4597	1	0.9989
i_t	$l_5^{(IV)}, l_6^{(IV)}$	0.0609	1	0.9932
r_t	$s_5^{(IV)}, s_6^{(IV)}$	0.0058	1	0.9902
f_t	$f_5^{(IV)}, f_6^{(IV)}, f_7^{(IV)}$ $l_4^{(IV)}, l_5^{(IV)}, l_6^{(IV)}, l_7^{(IV)}$ $s_4^{(IV)}, s_5^{(IV)}, s_6^{(IV)}, s_7^{(IV)}$	0.1923	2	0.9986
s_t	$f_5^{(IV)}, f_6^{(IV)}, f_7^{(IV)}$ $l_5^{(IV)}, l_6^{(IV)}, l_7^{(IV)}$ $s_4^{(IV)}, s_5^{(IV)}, s_6^{(IV)}, s_7^{(IV)}$	0.7159	1	0.74969
l_t	$f_5^{(IV)}, f_6^{(IV)}, f_7^{(IV)}, f_8^{(IV)}$ $i_6^{(IV)}, i_7^{(IV)}, i_8^{(IV)}$	4.7685	2	0.9287

likelihood estimation. We can observe some differences between GIVE and FIVE estimates, mainly in the precision of the estimated coefficients, where FIVE coefficients have generally smaller standard errors. For example, the GIVE estimates of $\beta_{22}^{(3)}$ and $\beta_{22}^{(4)}$ (lagged autoregressive coefficients in the equation for income) are not significant and in fact are of the wrong sign, which is not the case with their FIVE estimates.

Maximum likelihood estimation using LISREL fails using the default starting values generated by LISREL, and this holds too when various arbitrary starting values such as setting all starting values to 0.5 or choosing randomly generated starting values are used. We also tried to estimate a "shorter" version of this model using only first several years of data thereby having to estimate a considerably smaller and simpler model. This also failed using the default starting values. On the other hand, the GIVE/FIVE estimates proved to be fairly successful as starting values when instruments are carefully chosen. The minimum Sargan χ^2 criterion appears to be sufficient for selecting the suitable instruments sets for this purpose.

The convergence of the LISREL's ML algorithm proved to be exceptionally sensitive to how close the starting values are to the maximum likelihood estimates. While we did not encounter multiple optima problems (all converged ML solutions converged to the same estimates), we found that even minimal alterations to the "working" starting values lead to non-convergence. In such cases, LISREL stops responding or crashes eventually if we specify a very high number of iterations.

The full information maximum likelihood (FIML) estimates reported in table 6.12 were obtained by LISREL with the FIVE estimates used as starting values. The use of GIVE estimates lead to the same solution, but some of the insignificant coefficients had to be set to zero to achieve convergence. We also set to zero error

Table 6.12: Coefficient estimates

Parameter	FIML Estimate	FIML (SE)	GIVE Estimate	GIVE (SE)	FIVE (3SLS) Estimate	FIVE (3SLS) (SE)
$\lambda_{21}^{(h)}$	0.1433	(0.0052)	0.2524	(0.0867)	0.1395	(0.0037)
$\lambda_{42}^{(i)}$	0.0573	(0.0047)	0.1738	(0.0058)	0.0835	(0.0038)
$\lambda_{63}^{(r)}$	−0.1469	(0.0047)	−0.2254	(0.0062)	−0.1503	(0.0036)
$\beta_{11}^{(1)}$	0.5377	(0.0093)	0.6776	(3.2684)	0.4181	(0.0039)
$\beta_{11}^{(2)}$	0.2217	(0.0059)	0.2830	(2.6170)	0.1230	(0.0039)
$\beta_{11}^{(4)}$	0.0838	(0.0063)	0.0110	(0.0392)	0.0025	(0.0039)
$\beta_{13}^{(0)}$	0.0999	(0.0130)	0.0322	(0.5893)	0.0225	(0.0431)
$\beta_{13}^{(1)}$	−0.0977	(0.0147)	−0.0575	(0.8276)	−0.0682	(0.0291)
$\beta_{13}^{(2)}$	−0.0606	(0.0139)	−0.0366	(0.5221)	−0.0121	(0.0205)
$\beta_{12}^{(0)}$	−0.0647	(0.0140)	−0.0088	(0.8990)	−0.0377	(0.0413)
$\beta_{12}^{(1)}$	0.0974	(0.0161)	0.0441	(0.2354)	0.0573	(0.0248)
$\beta_{12}^{(2)}$	0.0449	(0.0149)	0.0670	(0.2592)	−0.0061	(0.0194)
$\beta_{22}^{(1)}$	0.4935	(0.0121)	0.7095	(1.9488)	0.4510	(0.0041)
$\beta_{22}^{(2)}$	0.2280	(0.0069)	0.5721	(1.2515)	0.2405	(0.0046)
$\beta_{22}^{(3)}$	0.1382	(0.0075)	−0.1853	(1.0921)	0.1790	(0.0045)
$\beta_{22}^{(4)}$	0.0212	(0.0028)	−0.0396	(0.0687)	0.0390	(0.0024)
$\beta_{21}^{(1)}$	0.0098	(0.0020)	0.0529	(5.2649)	0.0047	(0.0012)
$\beta_{21}^{(2)}$	0.0096	(0.0021)	0.0973	(4.1640)	0.0033	(0.0012)
$\beta_{23}^{(0)}$	0.9067	(0.0045)	0.1672	(0.6912)	0.8707	(0.0059)
$\beta_{23}^{(1)}$	−0.4326	(0.0116)	−0.7318	(2.4364)	−0.3816	(0.0067)
$\beta_{23}^{(2)}$	−0.2187	(0.0062)	−0.2021	(1.3866)	−0.2271	(0.0054)
$\beta_{23}^{(3)}$	−0.1372	(0.0069)	−0.2884	(1.0763)	−0.1779	(0.0052)
$\beta_{33}^{(1)}$	0.5132	(0.0055)	0.7609	(0.4588)	0.6955	(0.0041)
$\beta_{33}^{(2)}$	−0.0313	(0.0053)	−0.0184	(0.5904)	−0.0168	(0.0054)
$\beta_{33}^{(3)}$	0.0386	(0.0052)	0.0778	(0.5799)	0.0534	(0.0061)
$\beta_{33}^{(4)}$	−0.0181	(0.0054)	−0.1931	(0.3195)	−0.0273	(0.0065)
$\beta_{33}^{(5)}$	−0.0528	(0.0050)	−0.0423	(0.3310)	−0.0491	(0.0057)
$\gamma_{31}^{(0)}$	0.4507	(0.0081)	0.0253	(0.0223)	0.4879	(0.0024)

variances that were not estimated by the IV methods. Convergence with FIVE estimates used as starting values was achieved in 37 iterations, which took 56.131 seconds on a Pentium(R) 4, 2.00 GHz CPU machine.

The reported model specification (6.30), path diagram (figure 6.6) and estimates in tables 6.12 and 6.13 resulted from a general-to-specific modelling approach we took initially by estimating model (6.30) without dynamic restrictions for up to 6 lags in each variable, and subsequently dropping the insignificant coefficients. Thus the reported model and estimates refer to a parsimonious specification arrived at

Table 6.13: Variance estimates

	FIML		GIVE		FIVE (3SLS)	
Parameter	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
$\psi_{11}^{(C)}$	0.4708	(0.0133)	0.5971	–	0.5243	–
$\psi_{22}^{(L)}$	0.0692	(0.0028)	0.5620	–	0.0901	–
$\psi_{33}^{(I)}$	0.2890	(0.0026)	0.8624	–	0.3263	–
$\theta_{11}^{(f)}$	0.0064	(0.0100)	–	–	–	–
$\theta_{22}^{(h)}$	0.9817	(0.0067)	0.9998	–	0.9991	–
$\theta_{33}^{(l)}$	0.0109	(0.0012)	–	–	–	–
$\theta_{44}^{(i)}$	0.9978	(0.0068)	0.9864	–	0.9938	–
$\theta_{55}^{(s)}$	0.0255	(0.0024)	–	–	–	–
$\theta_{66}^{(r)}$	0.9800	(0.0066)	0.9806	–	0.9781	–

empirically. We can note that our basic approach of estimating OF equations with IV methods and subsequently using the IV estimates as starting values in maximum likelihood estimation worked well for sequential reduction of the model, i.e., we did not encounter non-convergence difficulties as a consequence of changes in retained coefficients estimates after dropping insignificant coefficients. This should be expected if the dropped coefficients were numerically close to zero and uncorrelated with the retained coefficients. Thus, there would be no need to provide new or different starting values for the retained coefficients and hence a general-to-specific model reduction strategy, common in dynamic econometric modelling, proved to be feasible in this case.

Table 6.14 summarises the significant contemporaneous and lagged effects in all equations, where symbols $\oplus$ and $\ominus$ indicate significant positive and negative effects, respectively (lack of any indicator implies insignificant or zero effect). The results suggest that current income has positive effect on current consumption, however, one- and two-year lagged income has negative effect. It thus seems plausible that increase in current consumption due to higher current income might be offset by higher past income, which could have raised past consumptions above its long-run value. This negative feedback could be interpreted as a form of correction of overspending induced by windfall gains in income, which might be consistent with the permanent income hypothesis.

In conclusion, by a combination of IV and ML methods we were able to estimate a dynamic structural equation model with latent variables using 13 waves of the British Household Panel Study. If estimated as a single equation, ignorant of measurement error, the income-consumption relationship would be a trivial estimation task, even if dynamics are included. However, possible simultaneity and feedback effects in the consumption equation calls for simultaneous estimation of equations for

Table 6.14: Significant contemporaneous and dynamic effects

Dynamic effect	Coefficient	Lag length $(t-j)$, $j=0,1,\ldots,5$ 0	1	2	3	4	5
Income $\rightarrowtail$ Consumption	$\beta_{13}^{(j)}$	$\oplus$	$\ominus$	$\ominus$			
Income $\rightarrowtail$ Liquidity	$\beta_{23}^{(j)}$	$\oplus$	$\ominus$	$\ominus$	$\ominus$		
Income $\rightarrowtail$ Income	$\beta_{33}^{(j)}$		$\oplus$	$\ominus$	$\oplus$	$\ominus$	$\ominus$
Consumption $\rightarrowtail$ Liquidity	$\beta_{21}^{(j)}$		$\oplus$	$\oplus$			
Consumption $\rightarrowtail$ Consumption	$\beta_{11}^{(j)}$		$\oplus$	$\oplus$		$\oplus$	
Liquidity $\rightarrowtail$ Consumption	$\beta_{12}^{(j)}$	$\ominus$	$\oplus$	$\ominus$			
Liquidity $\rightarrowtail$ Liquidity	$\beta_{22}^{(j)}$		$\oplus$	$\oplus$	$\oplus$	$\oplus$	
Education $\rightarrowtail$ Income	$\gamma_{31}^{(j)}$	$\oplus$					

income and liquidity. Furthermore, consumption, income, and liquidity constraints are all classical examples of latent variables, hence any of their proxy measures such as various types of expenditures and personal income most likely will confound considerable measurement error. Dynamic effects in the form of lagged values of both endogenous and exogenous variables are essential in this type of models and hence undoubtedly need to be modelled.

DSEM model that meets all these requirements, as we have shown, can be estimated, yet we believe that capabilities of the currently available software packages most likely would not suffice for estimation of any more complex models, particulary with larger time series dimension. A major difference is made by using the starting values obtained with the OF-IV methods, without which successful estimation of PDSEM models with LISREL would be very difficult or impossible.

Chapter 7

Conclusion

We have shown that dynamic structural equation models with latent variables can be formulated and theoretically analysed in a unified statistical framework that can encompass different traditions arising from errors-in-variables, covariance structure analysis, and state-space form modelling literatures.

Using modern matrix algebra methods we have derived closed form of the likelihood and covariance structure of the general DSEM model and gave the vectorised analytical derivatives for this model. The practical utility of this approach is in that it allows maximum likelihood estimation of DSEM models without imposing cross-equation equality constraints over all T time points and provides analytical first and second derivatives that can be easily implemented in modern matrix languages such as S and R.

These methods were used to propose a new maximum likelihood estimator that uses raw-data rather then sufficient statistics based on closed form likelihood and analytical derivatives. A suggestion to programming the estimation functions in S/R language along with some simple practical examples were also given.

Non-parametric estimation methods based on limited and full information instrumental variables estimation were proposed and applied as both a stand-alone estimation approach and as an auxiliary method for obtaining suitable starting values for the maximum likelihood estimation using software packages such as LISREL.

Using empirical data, we have shown that certain DSEM models, namely with $N > T$ and T relatively small, can be estimated using the already existing analytical results implemented in SEM software packages such as LISREL. In general, we found the starting values based on the suggested IV methods make considerable difference in LISREL estimation thereby enabling estimation of more complex dynamic models.

Chapter 8

Technical Appendices

8.1 Chapter §2 appendices

Appendix §2A

Proof of Proposition 2.3.1.1 We will show that the log-likelihood (2.31) can be written as a sum of the conditional log-likelihood of of $\boldsymbol{V}_T$ given $\boldsymbol{\Xi}_T$ and the marginal log-likelihoods of $\boldsymbol{\Xi}_T$ and $\boldsymbol{Z}_T$. By Definition 2.3.0.4 the matrix $\boldsymbol{K}_S$ is upper triangular with identity matrices on the diagonal and from (2.24) $\boldsymbol{\Sigma}_L$ is block diagonal. It follows that the determinant of the product $\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S$ is equal to the product of the determinants of the block-diagonal elements of $\boldsymbol{\Sigma}_L$,

$$\begin{aligned} \left|\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S\right| &= \left|\boldsymbol{K}_S\right|\left|\boldsymbol{\Sigma}_L\right|\left|\boldsymbol{K}'_S\right| \\ &= \left|\boldsymbol{I}\right|\left|\boldsymbol{\Sigma}_L\right|\left|\boldsymbol{I}\right| \\ &= \left|\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon\right|\left|\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta\right|\left|\boldsymbol{\Sigma}_\Xi\right|\left|\boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right|, \end{aligned} \tag{8.1}$$

which further simplifies to $T^3\left|\boldsymbol{\Theta}_\varepsilon\right|\left|\boldsymbol{\Theta}_\delta\right|\left|\boldsymbol{\Psi}\right|\left|\boldsymbol{\Sigma}_\Xi\right|$. Note that

$$\boldsymbol{K}_S^{-1} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & -\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & -\boldsymbol{A}_\Xi^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & -\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}, \tag{8.2}$$

and

$$\boldsymbol{\Sigma}_L^{-1} = \begin{pmatrix} \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Psi} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta^{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_\Xi^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1} \end{pmatrix}.$$

Since $\left(\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S\right)^{-1} = \boldsymbol{K}_S^{-1\prime}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{K}_S^{-1}$ we can re-arrange the trace of the product

$$\begin{aligned} &\operatorname{tr}\left(\boldsymbol{Y}'_T : \boldsymbol{X}'_T : \boldsymbol{\Xi}'_T : \boldsymbol{Z}'_T\right)\left(\boldsymbol{K}_S\boldsymbol{\Sigma}_L\boldsymbol{K}'_S\right)^{-1}\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix} \\ &\qquad = \operatorname{tr}\left(\boldsymbol{Y}'_T : \boldsymbol{X}'_T : \boldsymbol{\Xi}'_T : \boldsymbol{Z}'_T\right)\boldsymbol{K}_S^{-1\prime}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{K}_S^{-1}\begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T \\ \boldsymbol{\Xi}_T \\ \boldsymbol{Z}_T \end{pmatrix}, \end{aligned} \tag{8.3}$$

and multiply

$$K_S^{-1}\begin{pmatrix} Y_T \\ X_T \\ \Xi_T \\ Z_T \end{pmatrix} = \begin{pmatrix} I & 0 & -A_\Xi^{(1)}A_\Xi^{(2)} & -A_\Xi^{(1)} \\ 0 & I & -I_T \otimes \Lambda_x & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}\begin{pmatrix} Y_T \\ X_T \\ \Xi_T \\ Z_T \end{pmatrix} = \begin{pmatrix} Y_T - A_\Xi^{(1)}\ A_\Xi^{(2)}\Xi_T + Z_T \\ X_T - (I_T \otimes \Lambda_x)\,\Xi_T \\ \Xi_T \\ Z_T \end{pmatrix}.$$

Thus (8.3) can be re-arranged as

$$\begin{aligned} &\operatorname{tr}\left(Y_T - A_\Xi^{(1)}\left(A_\Xi^{(2)}\Xi_T + Z_T\right)\right)'\left(I_T \otimes \Theta_\varepsilon^{-1}\right)\left(Y_T - A_\Xi^{(1)}\left(A_\Xi^{(2)}\Xi_T + Z_T\right)\right) \\ &+\ \operatorname{tr}\left(X_T - (I_T \otimes \Lambda_x)\,\Xi_T\right)'\left(I_T \otimes \Theta_\delta^{-1}\right)\left(X_T - (I_T \otimes \Lambda_x)\,\Xi_T\right) \\ &+\ \operatorname{tr}\left(\Xi_T \Xi'_T \Sigma_\Xi^{-1}\right) + \operatorname{tr}\left(Z_T Z'_T\left(I_T \otimes \Psi^{-1}\right)\right). \end{aligned} \tag{8.4}$$

Therefore, the joint log-likelihood (2.31) can be written using (8.1) and (8.4 as

$$\begin{aligned} \ell_S\left(F_T^{(S)};\theta\right) &= \alpha - \frac{1}{2}\ln|I_T \otimes \Theta_\varepsilon| - \frac{1}{2}\ln|I_T \otimes \Theta_\delta| - \frac{1}{2}\ln|\Sigma_\Xi| - \frac{1}{2}\ln|I_T \otimes \Psi| \\ &- \frac{1}{2}\operatorname{tr}\left(V_T - P\begin{pmatrix} \Xi_T \\ Z_T \end{pmatrix}\right)'\begin{pmatrix} I_T \otimes \Theta_\varepsilon^{-1} & 0 \\ 0 & I_T \otimes \Theta_\delta^{-1} \end{pmatrix}\left(V_T - P\begin{pmatrix} \Xi_T \\ Z_T \end{pmatrix}\right) \\ &- \frac{1}{2}\operatorname{tr}\left(\Xi_T \Xi'_T \Sigma_\Xi^{-1}\right) - \frac{1}{2}\operatorname{tr}\left(Z_T Z'_T\left(I_T \otimes \Psi^{-1}\right)\right). \end{aligned} \tag{8.5}$$

Note that the conditional log-likelihood of V_T given Ξ_T and Z_T is

$$\begin{aligned} &\ell_{V|\Xi,Z}\left(V_T|\Xi_T, Z_T;\theta_1\right) = -\frac{(n+k)T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\begin{pmatrix} I_T \otimes \Theta_\varepsilon^{-1} & 0 \\ 0 & I_T \otimes \Theta_\delta^{-1} \end{pmatrix}\right| \\ &-\frac{1}{2}\operatorname{tr}\left(V_T - A\begin{pmatrix} \Xi_T \\ Z_T \end{pmatrix}\right)'\begin{pmatrix} I_T \otimes \Theta_\varepsilon^{-1} & 0 \\ 0 & I_T \otimes \Theta_\delta^{-1} \end{pmatrix}\left(V_T - A\begin{pmatrix} \Xi_T \\ Z_T \end{pmatrix}\right), \end{aligned} \tag{8.6}$$

while the marginal log-likelihoods of Ξ_T and Z_T are

$$\ell_\Xi\left(\Xi_T;\theta_2\right) = -\frac{gT}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_\Xi| - \frac{1}{2}\operatorname{tr}\left(\Xi_T \Xi'_T \Sigma_\Xi^{-1}\right), \tag{8.7}$$

and

$$\ell_Z\left(Z_T;\theta_3\right) = -\frac{mT}{2}\ln(2\pi) - \frac{1}{2}\ln|I_T \otimes \Psi| - \frac{1}{2}\operatorname{tr}\left(Z_T Z'_T\left(I_T \otimes \Psi^{-1}\right)\right), \tag{8.8}$$

respectively. It follows that $\ell_S\left(F_T^{(S)};\theta\right) = \ell_{V|\Xi,Z}\left(V_T|\Xi_T, Z_T;\theta_1\right) + \ell_\Xi\left(\Xi_T;\theta_2\right) + \ell_Z\left(Z_T;\theta_3\right)$, as required.

Q.E.D.

Appendix §2B

Proof of Proposition 2.3.2.1 Firstly note that by Assumption 3.2.0.1 implies we have the following results for the time series processes $\{\zeta\}_1^T$, $\{\varepsilon\}_1^T$, and $\{\delta\}_1^T$,

$$
\begin{aligned}
E\left[\zeta_{t-k}\zeta_{t-s}'\right] &= \begin{cases} \boldsymbol{\Psi}, & k=s \\ \boldsymbol{0}, & k\neq s \end{cases} \Rightarrow E\left[\left(\zeta_1',\ldots,\zeta_T'\right)'\left(\zeta_1',\ldots,\zeta_T'\right)\right] = \boldsymbol{I}_T\otimes\boldsymbol{\Psi} \\
E\left[\varepsilon_{t-k}\varepsilon_{t-s}'\right] &= \begin{cases} \boldsymbol{\Theta}_\varepsilon, & k=s \\ \boldsymbol{0}, & k\neq s \end{cases} \Rightarrow E\left[\left(\varepsilon_1',\ldots,\varepsilon_T'\right)'\left(\varepsilon_1',\ldots,\varepsilon_T'\right)\right] = \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon \\
E\left[\delta_{t-k}\delta_{t-s}'\right] &= \begin{cases} \boldsymbol{\Theta}_\delta, & k=s \\ \boldsymbol{0}, & k\neq s \end{cases} \Rightarrow E\left[\left(\delta_1',\ldots,\delta_T'\right)'\left(\delta_1',\ldots,\delta_T'\right)\right] = \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta,
\end{aligned}
$$

therefore, in T-notation (Table 2.2) we have

$$
\begin{aligned}
E\left[\boldsymbol{Z}_T\boldsymbol{Z}_T'\right] &= E\left[\left(\operatorname{vec}\{\zeta_t\}_1^T\right)\left(\operatorname{vec}\{\zeta_t'\}_1^T\right)\right] = \left(\boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right) && (8.9)\\
E\left[\boldsymbol{E}_T\boldsymbol{E}_T'\right] &= E\left[\left(\operatorname{vec}\{\varepsilon_t\}_1^T\right)\left(\operatorname{vec}\{\varepsilon_t'\}_1^T\right)\right] = \left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon\right) && (8.10)\\
E\left[\boldsymbol{\Delta}_T\boldsymbol{\Delta}_T'\right] &= E\left[\left(\operatorname{vec}\{\delta_t\}_1^T\right)\left(\operatorname{vec}\{\delta_t'\}_1^T\right)\right] = \left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\delta\right). && (8.11)
\end{aligned}
$$

By the reduced-form equations (2.22) and (2.23) for $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T$ the block-elements of (2.34) can be derived as

$$
\begin{aligned}
\boldsymbol{\Sigma}_{11} &= E\left[\boldsymbol{Y}_T\boldsymbol{Y}_T'\right] \\
&= E\left[\left(\left(\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_y\right)\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}_T^j\otimes\boldsymbol{B}_j\right)^{-1}\left(\left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\boldsymbol{\Xi}_T+\boldsymbol{Z}_T\right)+\boldsymbol{E}_T\right)\right. \\
&\quad\left.\times\left(\left(\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_y\right)\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}_T^j\otimes\boldsymbol{B}_j\right)^{-1}\left(\left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\boldsymbol{\Xi}_T+\boldsymbol{Z}_T\right)+\boldsymbol{E}_T\right)'\right],
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma}_{12} &= E\left[\boldsymbol{Y}_T\boldsymbol{X}_T'\right] \\
&= E\left[\left(\left(\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_y\right)\left(\boldsymbol{I}_{mT}-\sum_{j=0}^{p}\boldsymbol{S}_T^j\otimes\boldsymbol{B}_j\right)^{-1}\left(\left(\sum_{j=0}^{q}\boldsymbol{S}_T^j\otimes\boldsymbol{\Gamma}_j\right)\boldsymbol{\Xi}_T+\boldsymbol{Z}_T\right)+\boldsymbol{E}_T\right)\right. \\
&\quad\left.\times\left(\left(\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x\right)\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T\right)'\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\Sigma}_{22} &= E\left[\boldsymbol{X}_T\boldsymbol{X}_T'\right] \\
&= E\left[\left(\left(\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x\right)\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T\right)\left(\left(\boldsymbol{I}_T\otimes\boldsymbol{\Lambda}_x\right)\boldsymbol{\Xi}_T+\boldsymbol{\Delta}_T\right)'\right],
\end{aligned}
$$

which by using (8.9)–(8.11) evaluate to (2.35), (2.36), and (2.37), respectively. Note that by covariance stationarity (Assumptions 3.2.0.1 and 2.2.0.2) $\boldsymbol{\Sigma}_{\Xi}$ has block-Toeplitz structure

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\Xi} &= \begin{pmatrix} \boldsymbol{\Phi}_0 & \boldsymbol{\Phi}'_1 & \boldsymbol{\Phi}'_2 & \cdots & \boldsymbol{\Phi}'_{T-1} \\ \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \ddots & \ddots & \vdots \\ \boldsymbol{\Phi}_2 & \ddots & \ddots & \boldsymbol{\Phi}'_1 & \boldsymbol{\Phi}'_2 \\ \vdots & \ddots & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 & \boldsymbol{\Phi}'_1 \\ \boldsymbol{\Phi}_{T-1} & \cdots & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_0 \end{pmatrix} \\
&= \sum_{j=0}^{T-1} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j \right) + \sum_{j=1}^{T-1} \left(\boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right) \\
&= \boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{T-1} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right), \qquad (8.12)
\end{aligned}
$$

and also note that $E\left[\boldsymbol{Z}_T\boldsymbol{Z}'_T\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Psi}$, $E\left[\boldsymbol{E}_T\boldsymbol{E}'_T\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon$, and $E\left[\boldsymbol{\Delta}_T\boldsymbol{\Delta}'_T\right] = \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta$. Typically, most of the block-elements $\boldsymbol{\Phi}_j$ of the second-moment matrix $E\left[\boldsymbol{\Xi}_T\boldsymbol{\Xi'}_T\right]$ will be zero, depending on the length of the memory in the process generating $\boldsymbol{\xi}_t$, which for the reason of simplicity we take to be q. Thus, for $j > q$, $\boldsymbol{\Phi}_j = \boldsymbol{0}$. It follows that (8.12) can be simplified to

$$
\begin{pmatrix} \boldsymbol{\Phi}_0 & \cdots & \boldsymbol{\Phi}'_q & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{\Phi}_0 & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{\Phi}_q & \ddots & \ddots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \ddots & \ddots & \ddots & \boldsymbol{\Phi}'_q \\ \vdots & \ddots & \ddots & \ddots & \boldsymbol{\Phi}_0 & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Phi}_q & \cdots & \boldsymbol{\Phi}_0 \end{pmatrix} = \boldsymbol{S}_T^0 \otimes \boldsymbol{\Phi}_0 + \sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right), \qquad (8.13)
$$

which consists of only $q+1$ symmetric matrices $\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_q$. Finally, note that $\boldsymbol{\Sigma}'_{12} = \boldsymbol{\Sigma}_{21}$.

Q.E.D.

Appendix §2C

Proof of Proposition 2.3.4.2 The proof proceeds similarly to the proof of Proposition 2.3.1.1. Firstly note that (2.74) can be written as

$$
\begin{aligned}
\ell_{OF}^{(S)}\left(\boldsymbol{F}_T^{(OF)};\boldsymbol{\theta}\right) &= -\frac{(n+k+g+m)T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}'\right| \\
&\quad -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{Y'}_T : \boldsymbol{X'}_T^{(\Lambda)} : \boldsymbol{X'}_T^{(U)} : \boldsymbol{\Delta'}_T^{(U)} : \boldsymbol{Z'}_T\right)\left(\boldsymbol{M}\boldsymbol{\Sigma}_L^*\boldsymbol{M}'\right)^{-1} \begin{pmatrix} \boldsymbol{Y}_T \\ \boldsymbol{X}_T^{(\Lambda)} \\ \boldsymbol{X}_T^{(U)} \\ \boldsymbol{\Delta}_T^{(U)} \\ \boldsymbol{Z}_T \end{pmatrix}, \qquad (8.14)
\end{aligned}
$$

which can be rearranged by following the same procedure we used to derive (8.5) as

$$
\begin{aligned}
&-\tfrac{(n+k+g+m)T}{2}\ln(2\pi)\\
&-\tfrac{1}{2}\ln \begin{matrix} \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}\end{matrix} \quad -\tfrac{1}{2}\ln\ \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\ -\tfrac{1}{2}\ln|\boldsymbol{\Sigma}_\Xi|-\tfrac{1}{2}\ln|\boldsymbol{I}_T\otimes\boldsymbol{\Psi}|\\
&-\tfrac{1}{2}\mathrm{tr}\left[\begin{matrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{matrix} \quad -\tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]' \begin{matrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)-1}\end{matrix}\Bigg)\left[\begin{matrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{matrix} \quad -\tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]\\
&-\tfrac{1}{2}\mathrm{tr}\quad (\boldsymbol{I}:-\boldsymbol{I})\quad \begin{matrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\end{matrix}\Bigg)\ \boldsymbol{X'}_T^{(U)}:\boldsymbol{\Delta'}_T^{(U)}\quad \begin{matrix}\boldsymbol{I}\\ -\boldsymbol{I}\end{matrix}\quad \boldsymbol{\Sigma}_\Xi^{-1}\Bigg)\\
&-\tfrac{1}{2}\mathrm{tr}\quad \boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\quad \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\ ^{-1}\quad -\tfrac{1}{2}\mathrm{tr}\quad \boldsymbol{Z}_T\boldsymbol{Z'}_T\quad \boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1}\quad ,
\end{aligned}
\tag{8.15}
$$

where

$$
\tilde{\boldsymbol{A}}\equiv\begin{pmatrix}\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & -\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)} & \boldsymbol{A}_\Xi^{(1)}\\ \boldsymbol{I}_T\otimes\tilde{\boldsymbol{\Lambda}}_x & \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}. \tag{8.16}
$$

Note that $\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Xi}_T=\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{X}_T^{(U)}-\boldsymbol{A}_\Xi^{(1)}\boldsymbol{A}_\Xi^{(2)}\boldsymbol{\Delta}_T^{(U)}$. Finally, we can observe that

$$
\begin{aligned}
\ell_{Y,X^\Lambda|X^U,\Delta^U,Z}\left(\boldsymbol{Y}_T,\boldsymbol{X}_T^{(\Lambda)}|\boldsymbol{X}_T^{(U)},\boldsymbol{\Delta}_T^{(U)},\boldsymbol{Z}_T;\boldsymbol{\theta}_1^*\right) &= -\tfrac{(n+k-g)T}{2}\ln(2\pi)\\
&-\tfrac{1}{2}\ln\left|\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}\end{pmatrix}\right|\\
&-\tfrac{1}{2}\mathrm{tr}\left[\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{pmatrix}-\tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]'\\
&\times\begin{pmatrix}\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)-1}\end{pmatrix}\left[\begin{pmatrix}\boldsymbol{Y}_T\\ \boldsymbol{X}_T^{(\Lambda)}\end{pmatrix}-\tilde{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\\ \boldsymbol{Z}_T\end{pmatrix}\right]
\end{aligned}
\tag{8.17}
$$

is the conditional log-likelihood of $\boldsymbol{Y}_T$ and $\boldsymbol{X}_T^{(\Lambda)}$ given $\boldsymbol{X}_T^{(U)}$, $\boldsymbol{\Delta}_T^{(U)}$, and $\boldsymbol{Z}_T$. The marginal log-likelihoods of $\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)}$, ℓ_{Δ^U}, and $\boldsymbol{Z}_T$ are given by

$$
\begin{aligned}
\ell_M\left(\boldsymbol{X}_T^{(U)}-\boldsymbol{\Delta}_T^{(U)};\boldsymbol{\theta}_2^*\right) &= -\frac{gT}{2}\ln(2\pi)-\frac{1}{2}\ln|\boldsymbol{\Sigma}_\Xi|\\
&- \frac{1}{2}\mathrm{tr}\left((\boldsymbol{I}:-\boldsymbol{I})\begin{pmatrix}\boldsymbol{X}_T^{(U)}\\ \boldsymbol{\Delta}_T^{(U)}\end{pmatrix}\left(\boldsymbol{X'}_T^{(U)}:\boldsymbol{\Delta'}_T^{(U)}\right)\begin{pmatrix}\boldsymbol{I}\\ -\boldsymbol{I}\end{pmatrix}\boldsymbol{\Sigma}_\Xi^{-1}\right),
\end{aligned}
$$

$$
\ell_M\left(\boldsymbol{\Delta}_T^{(U)};\boldsymbol{\theta}_3^*\right)=-\frac{gT}{2}\ln(2\pi)-\frac{1}{2}\ln\left|\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right|-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta'}_T^{(U)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)^{-1}\right), \tag{8.18}
$$

and

$$
\ell_M(\boldsymbol{Z}_T;\boldsymbol{\theta}_4^*)=-\frac{mT}{2}\ln(2\pi)-\frac{1}{2}\ln|\boldsymbol{I}_T\otimes\boldsymbol{\Psi}|-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{Z}_T\boldsymbol{Z'}_T\left(\boldsymbol{I}_T\otimes\boldsymbol{\Psi}^{-1}\right)\right), \tag{8.19}
$$

respectively. Hence (2.74) factorise into (2.75), as required.

Q.E.D.

Appendix §2D

Proof of Proposition 2.3.4.4 We will compare (2.77) and (2.50) by comparing their corresponding block-elements in expectation and probability limit. Recall that by Lemma (2.3.4.1) we have $\boldsymbol{\Sigma}_{\Xi} = \boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}$. Therefore, we can evaluate the block elements of (2.77) as follows.

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{YY} &= E\Big[\Big(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_T^{(U)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T + \boldsymbol{E}_T\Big) \\
&\quad \times \Big(\boldsymbol{X}'^{(U)}_T\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} - \boldsymbol{\Delta}'^{(U)}_T\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{Z}'_T\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{E}'_T\Big)\Big] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{Z}'_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_T^{(U)}\boldsymbol{E}'_T\right]}_{\boldsymbol{0}} \\
&- \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} \\
&- \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{Z}'_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(1)}_{\Xi} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{E}'_T\right]}_{\boldsymbol{0}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} - \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{Z}'_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Psi}}\boldsymbol{A}'^{(1)}_{\Xi} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{E}'_T\right]}_{\boldsymbol{0}} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} - \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} \\
&+ \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{Z}'_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(1)}_{\Xi} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{E}'_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon}} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\Big(\boldsymbol{A}_{\Xi}^{(2)}\Big(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\Big)\boldsymbol{A}'^{(2)}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Psi}\Big)\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon}. \qquad (8.20)
\end{aligned}
$$

By (2.63) for the random case or by (2.64) for the fixed case (8.20) becomes

$$\boldsymbol{A}_{\Xi}^{(1)}\Big(\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Sigma}_{\Xi}\boldsymbol{A}'^{(2)}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Psi}\Big)\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\varepsilon} = \bar{\boldsymbol{\Sigma}}_{YY}.$$

For $\tilde{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)}$ we have

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_{T}^{(U)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_{T}^{(U)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_{T} + \boldsymbol{E}_{T}\right)\right. \\
&\quad \left.\times\left(\boldsymbol{X}'^{(U)}_{T}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) - \boldsymbol{\Delta}'^{(U)}_{T}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) + \boldsymbol{\Delta}'^{(\Lambda)}_{T}\right)\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_{T}^{(U)}\boldsymbol{X}'^{(U)}_{T}\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_{T}^{(U)}\boldsymbol{\Delta}'^{(U)}_{T}\right]}_{\boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_{T}^{(U)}\boldsymbol{\Delta}'^{(\Lambda)}_{T}\right]}_{\boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_{T}^{(U)}\boldsymbol{X}'^{(U)}_{T}\right]}_{\boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_{T}^{(U)}\boldsymbol{\Delta}'^{(U)}_{T}\right]}_{\boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_{T}^{(U)}\boldsymbol{\Delta}'^{(\Lambda)}_{T}\right]}_{\boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_{T}\boldsymbol{X}'^{(U)}_{T}\right]}_{0}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) - \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_{T}\boldsymbol{\Delta}'^{(U)}_{T}\right]}_{0}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_{T}\boldsymbol{\Delta}'^{(\Lambda)}_{T}\right]}_{0} + \underbrace{E\left[\boldsymbol{E}_{T}\boldsymbol{X}'^{(U)}_{T}\right]}_{0}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) \\
&- \underbrace{E\left[\boldsymbol{E}_{T}\boldsymbol{\Delta}'^{(U)}_{T}\right]}_{0}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) + \underbrace{E\left[\boldsymbol{E}_{T}\boldsymbol{\Delta}'^{(\Lambda)}_{T}\right]}_{0} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) \qquad (8.21)
\end{aligned}
$$

Similarly, (8.21) evaluates to $\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Sigma}_{\Xi}\left(\boldsymbol{I}_{T}\otimes\bar{\boldsymbol{\Lambda}}'_{x}\right) = \bar{\boldsymbol{\Sigma}}_{YX}^{(\Lambda)}$. Next, for $\tilde{\boldsymbol{\Sigma}}_{YX}^{(U)}$ we have

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{YX}^{(U)} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{X}_{T}^{(U)} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_{T}^{(U)} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_{T} + \boldsymbol{E}_{T}\right)\boldsymbol{X}'^{(U)}_{T}\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{X}_{T}^{(U)}\boldsymbol{X}'^{(U)}_{T}\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_{T}^{(U)}\boldsymbol{X}'^{(U)}_{T}\right]}_{\boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_{T}\boldsymbol{X}'^{(UL)}_{T}\right]}_{0} + \underbrace{E\left[\boldsymbol{E}_{T}\boldsymbol{X}'^{(U)}_{T}\right]}_{0} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_{T}\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right), \qquad (8.22)
\end{aligned}
$$

which can be evaluated as $\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Sigma}_{\Xi} = \bar{\boldsymbol{\Sigma}}_{YX}^{(U)}$. For $\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)}$ we have

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)} &= E\left[\left(\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{X}_T^{(U)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{\Delta}_T^{(U)} + \boldsymbol{\Delta}_T^{(\Lambda)}\right)\right. \\
&\quad \left. \times \left(\boldsymbol{X}'^{(U)}_T \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{\Delta}'^{(U)}_T \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{\Delta}'^{(\Lambda)}_T\right)\right] \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{X}_T^{(U)} \boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{X}_T^{(U)} \boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\
&+ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{X}_T^{(U)} \boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{\Delta}_T^{(U)} \boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) \\
&+ \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{\Delta}_T^{(U)} \boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{\Delta}_T^{(U)} \boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)} \boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) - \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)} \boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)} \boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}} \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right) \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}, \qquad (8.23)
\end{aligned}
$$

which becomes $\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{\Sigma}_{\Xi} \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}'_x\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} = \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda\Lambda)}$. Similarly, for $\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)}$ it follows that

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)} &= E\left[\left(\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{X}_T^{(U)} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{\Delta}_T^{(U)} + \boldsymbol{\Delta}_T^{(\Lambda)}\right) \boldsymbol{X}'^{(U)}_T\right] \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{X}_T^{(U)} \boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{\Sigma}_{XX}^{(UU)}} - \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \underbrace{E\left[\boldsymbol{\Delta}_T^{(U)} \boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}} + \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)} \boldsymbol{X}'^{(U)}_T\right]}_{\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}} \\
&= \left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \left(\boldsymbol{\Sigma}_{XX}^{(UU)} - \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right) + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}, \qquad (8.24)
\end{aligned}
$$

which evaluates to $\left(\boldsymbol{I}_T \otimes \bar{\boldsymbol{\Lambda}}_x\right) \boldsymbol{\Sigma}_{\Xi} + \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \bar{\boldsymbol{\Sigma}}_{XX}^{(\Lambda U)}$. Finally,

$$
\tilde{\boldsymbol{\Sigma}}_{XX}^{(UU)} = E\left[\boldsymbol{X}_T^{(UL)} \boldsymbol{X}'^{(UL)}_T\right] = \boldsymbol{\Sigma}_{XX}^{(UU)}, \qquad (8.25)
$$

thus trivially we have $\tilde{\boldsymbol{\Sigma}}_{XX}^{(UU)} = \bar{\boldsymbol{\Sigma}}_{XX}^{(UU)}$. Therefore, $E[\tilde{\boldsymbol{\Sigma}}] = E[\bar{\boldsymbol{\Sigma}}]$ or $\underset{T\to\infty}{p\lim}\, 1/T\tilde{\boldsymbol{\Sigma}} = \underset{T\to\infty}{p\lim}\, 1/T\bar{\boldsymbol{\Sigma}}$, as required.

Q.E.D.

Appendix §2E

Proof of Proposition 2.3.4.5 We firstly derive (2.89)–(2.91) from (2.83)–(2.85) using Assumption 3.2.0.1. For $\boldsymbol{\Omega}_{YY}$ we have

$$\begin{aligned}
\boldsymbol{\Omega}_{YY} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{E}_T\right)\left(\boldsymbol{Z}'_T\boldsymbol{A}'^{(1)}_{\Xi} - \boldsymbol{\Delta}'^{(U)}_T\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{E}'_T\right)\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{Z}'_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Psi}}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{E}'_T\right]}_{\boldsymbol{0}} \\
&+ \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\boldsymbol{A}'^{(2)}_{\Xi}\boldsymbol{A}'^{(1)}_{\Xi} \\
&- \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{Z}'_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(1)}_{\Xi} - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{E}'_T\right]}_{\boldsymbol{0}} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{Z}'_T\right]}_{\boldsymbol{0}}\boldsymbol{A}'^{(1)}_{\Xi} + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{E}'_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon} \\
&= \boldsymbol{A}_{\Xi}^{(1)}\left(\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\boldsymbol{A}'^{(2)}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Psi}\right)\boldsymbol{A}'^{(1)}_{\Xi} + \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_\varepsilon, \qquad (8.26)
\end{aligned}$$

which gives (2.89). Next, for $\boldsymbol{\Omega}_{YX}^{(\Lambda)}$ we have

$$\begin{aligned}
\boldsymbol{\Omega}_{YX}^{(\Lambda)} &= E\left[\left(\boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{Z}_T - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\boldsymbol{\Delta}_T^{(U)} + \boldsymbol{E}_T\right)\left(\boldsymbol{\Delta}'^{(\Lambda)}_T - \boldsymbol{\Delta}'^{(U)}_T\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right)\right)\right] \\
&= \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{0}} - \boldsymbol{A}_{\Xi}^{(1)}\underbrace{E\left[\boldsymbol{Z}_T\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{0}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) + \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{0}} - \underbrace{E\left[\boldsymbol{E}_T\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{0}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) \\
&= \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)}\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}\right), \qquad (8.27)
\end{aligned}$$

which, since $\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \boldsymbol{0}$, yields (2.90). Finally, $\boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)}$ can be evaluated as

$$\begin{aligned}
\boldsymbol{\Omega}_{XX}^{(\Lambda\Lambda)} &= E\left[\left(\boldsymbol{\Delta}_T^{(\Lambda)} - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\boldsymbol{\Delta}_T^{(U)}\right)\left(\boldsymbol{\Delta}'^{(\Lambda)}_T - \boldsymbol{\Delta}'^{(U)}_T\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right)\right)\right] \\
&= \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)}} - \underbrace{E\left[\boldsymbol{\Delta}_T^{(\Lambda)}\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(\Lambda)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}} \\
&+ \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\underbrace{E\left[\boldsymbol{\Delta}_T^{(U)}\boldsymbol{\Delta}'^{(U)}_T\right]}_{\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}}\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) \\
&= \boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} + \left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right) - \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(U\Lambda)}\right) \\
&+ \left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x\right)\left(\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(UU)}\right)\left(\boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}'_x\right), \qquad (8.28)
\end{aligned}$$

yielding (2.91) again by noting that $\boldsymbol{I}_T\otimes\boldsymbol{\Theta}_{\delta\delta}^{(\Lambda U)} = \boldsymbol{0}$. Secondly, we derive (2.92) as follows. By Definition 2.3.0.4,

$$\boldsymbol{K}_R \equiv \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{A}_{\Xi}^{(1)}\boldsymbol{A}_{\Xi}^{(2)} & \boldsymbol{A}_{\Xi}^{(1)} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I}_T\otimes\bar{\boldsymbol{\Lambda}}_x & \boldsymbol{0} \end{pmatrix}, \qquad (8.29)$$

hence

$$\bar{\boldsymbol{D}}_S \boldsymbol{\Sigma}_L \bar{\boldsymbol{D}}_S' \equiv \begin{pmatrix} \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(\Lambda\Lambda)} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Theta}_{\delta\delta}^{(UU)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_T \otimes \boldsymbol{\Psi} \end{pmatrix}, \tag{8.30}$$

where $\boldsymbol{\Sigma}_L$ is defined as in (2.24). Finally, premultiplying and postmultiplying (8.30) by (8.29) yields (2.92), as required.

Q.E.D.

Appendix §2F

Proof of Proposition 2.3.5.3 We will show that the general DSEM model (2.1)–(2.2) can be written in the state-space form (2.95)–(2.96). Firstly, the structural part of the general DSEM model (2.1) and the VAR(q) process for $\boldsymbol{\xi}_t$ (2.97) can be written as a system

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} &= \begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} \\ &+ \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} + \ldots + \begin{pmatrix} \boldsymbol{B}_r & \boldsymbol{\Gamma}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{v}_t \end{pmatrix}, \end{aligned} \tag{8.31}$$

or equivalently as

$$\begin{aligned} \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0) & -\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} &= \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} \\ &+ \ldots + \begin{pmatrix} \boldsymbol{B}_r & \boldsymbol{\Gamma}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{v}_t \end{pmatrix}. \end{aligned} \tag{8.32}$$

Therefore, the reduced form of (8.31) is

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} &= \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0) & -\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}^{-1} \left[\begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} \right. \\ &+ \left. \ldots + \begin{pmatrix} \boldsymbol{B}_r & \boldsymbol{\Gamma}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{v}_t \end{pmatrix} \right]. \end{aligned} \tag{8.33}$$

Note that

$$\begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0) & -\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} & (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix},$$

hence

$$\begin{aligned} &\begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} & (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}_j & \boldsymbol{\Gamma}_j \\ \boldsymbol{0} & \boldsymbol{R}_j \end{pmatrix} \\ &= \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{B}_j & (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{\Gamma}_j + (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{\Gamma}_0 \boldsymbol{R}_j \\ \boldsymbol{0} & \boldsymbol{R}_j \end{pmatrix}, \end{aligned} \tag{8.34}$$

and

$$\begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0)^{-1} & (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_0 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\zeta}_t \\ \boldsymbol{v}_t \end{pmatrix} = \begin{pmatrix} (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\zeta}_t + (\boldsymbol{I}-\boldsymbol{B}_0)^{-1}\boldsymbol{\Gamma}_0\boldsymbol{v}_t \\ \boldsymbol{v}_t \end{pmatrix}. \tag{8.35}$$

Therefore, using (8.34) and (8.35), the reduced form of the system (8.31) can be written as

$$\begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{G}_1 \\ \boldsymbol{0} & \boldsymbol{R}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} + \ldots + \begin{pmatrix} \boldsymbol{\Pi}_r & \boldsymbol{G}_r \\ \boldsymbol{0} & \boldsymbol{R}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_{t-r} \\ \boldsymbol{\xi}_{t-r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{K}_t \\ \boldsymbol{v}_t \end{pmatrix}. \tag{8.36}$$

making use of the notation from Definition 2.3.5.2. Finally, we stack the current and lagged $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ into a single column vector, collect all coefficient matrices in a single block matrix, and stack the residuals into a single as

$$\boldsymbol{\vartheta}_t \equiv \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \\ \boldsymbol{\eta}_{t-1} \\ \boldsymbol{\xi}_{t-1} \\ \vdots \\ \boldsymbol{\eta}_{t-r+1} \\ \boldsymbol{\xi}_{t-r+1} \end{pmatrix}, \boldsymbol{H} \equiv \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{G}_1 & \cdots & \boldsymbol{\Pi}_{r-1} & \boldsymbol{G}_{r-1} & \boldsymbol{\Pi}_r & \boldsymbol{G}_r \\ \boldsymbol{0} & \boldsymbol{R}_1 & \cdots & \boldsymbol{0} & \boldsymbol{R}_{r-1} & \boldsymbol{0} & \boldsymbol{R}_r \\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \boldsymbol{w}_t \equiv \begin{pmatrix} \boldsymbol{K}_t \\ \boldsymbol{v}_t \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \tag{8.37}$$

and

$$\boldsymbol{W}_t \equiv \begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{x}_t \end{pmatrix}, \boldsymbol{F} \equiv \begin{pmatrix} \boldsymbol{\Lambda}_y & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_x & \cdots & \boldsymbol{0} \end{pmatrix}, \boldsymbol{u}_t \equiv \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\delta}_t \end{pmatrix},$$

therefore, (8.36) can be written in the state space form (2.95)–(2.96), as required.

Q.E.D.

8.2 Chapter §3 appendices

Appendix §3A

Proof of Lemma 3.2.2.1 Firstly, let $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_4$ be some zero-one matrices such that

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \boldsymbol{G}_1 \otimes \boldsymbol{\Sigma}_{11} + \boldsymbol{G}_2 \otimes \boldsymbol{\Sigma}_{21} + \boldsymbol{G}_3 \otimes \boldsymbol{\Sigma}_{12} + \boldsymbol{G}_4 \otimes \boldsymbol{\Sigma}_{22},$$

which, by applying the vec operator yields

$$\begin{aligned} \operatorname{vec}\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} &= \operatorname{vec}(\boldsymbol{G}_1 \otimes \boldsymbol{\Sigma}_{11}) + \operatorname{vec}(\boldsymbol{G}_2 \otimes \boldsymbol{\Sigma}_{21}) \\ &\quad + \operatorname{vec}(\boldsymbol{G}_3 \otimes \boldsymbol{\Sigma}_{12}) + \operatorname{vec}(\boldsymbol{G}_4 \otimes \boldsymbol{\Sigma}_{22}) \\ &= \boldsymbol{H}_1 \operatorname{vec}\boldsymbol{\Sigma}_{11} + \boldsymbol{H}_2 \operatorname{vec}\boldsymbol{\Sigma}_{21} + \boldsymbol{H}_3 \operatorname{vec}\boldsymbol{\Sigma}_{12} + \boldsymbol{H}_4 \operatorname{vec}\boldsymbol{\Sigma}_{22}, \end{aligned}$$

for some zero-one matrices $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_4$. Note that for any $\boldsymbol{G}_k$ $(a \times b)$ and $\boldsymbol{\Sigma}_{ij}$ $(c \times d)$ it holds that $\operatorname{vec}\boldsymbol{G}_k \otimes \boldsymbol{\Sigma}_{ij} = [(\boldsymbol{I}_b \otimes \boldsymbol{K}_{da})(\operatorname{vec}\boldsymbol{G}_k \otimes \boldsymbol{I}_d) \otimes \boldsymbol{I}_c] \operatorname{vec}\boldsymbol{\Sigma}_{ij}$, therefore $\boldsymbol{H}_k = [(\boldsymbol{I}_b \otimes \boldsymbol{K}_{da})(\operatorname{vec}\boldsymbol{G}_k \otimes \boldsymbol{I}_d) \otimes \boldsymbol{I}_c]$. Now, to show that $\operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be expressed as a linear function of the vectors $\operatorname{vec}\boldsymbol{\Sigma}_{ij} = \left(\boldsymbol{m}_1^{(ij)\prime} \cdots \boldsymbol{m}_{nT}^{(ij)\prime}\right)'$, $i, j = 1, 2$ we will show that $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_4$ are of the required form. Note that the dimensions of the blocks of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and their columns are

$$\begin{pmatrix} \underbrace{\boldsymbol{\Sigma}_{11}}_{nT \times nT} & \underbrace{\boldsymbol{\Sigma}_{12}}_{nT \times kT} \\ \underbrace{\boldsymbol{\Sigma}_{21}}_{kT \times nT} & \underbrace{\boldsymbol{\Sigma}_{22}}_{kT \times kT} \end{pmatrix} = \begin{pmatrix} \underbrace{\boldsymbol{m}_1^{(11)}}_{nT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{nT}^{(11)}}_{nT \times 1} & \underbrace{\boldsymbol{m}_1^{(12)}}_{nT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{kT}^{(12)}}_{nT \times 1} \\ \underbrace{\boldsymbol{m}_1^{(21)}}_{kT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{nT}^{(21)}}_{kT \times 1} & \underbrace{\boldsymbol{m}_1^{(22)}}_{kT \times 1} & \cdots & \underbrace{\boldsymbol{m}_{kT}^{(22)}}_{kT \times 1} \end{pmatrix}.$$

Applying the vec operator to the columns-partition (3.26) of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ produces a $T^2(n+k)^2$ vector

$$\operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \operatorname{vec}\begin{pmatrix} \boldsymbol{m}_1^{(11)} & \cdots & \boldsymbol{m}_{nT}^{(11)} & \boldsymbol{m}_1^{(12)} & \cdots & \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{m}_1^{(21)} & \cdots & \boldsymbol{m}_{nT}^{(21)} & \boldsymbol{m}_1^{(22)} & \cdots & \boldsymbol{m}_{kT}^{(22)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_1^{(11)} \\ \boldsymbol{m}_1^{(21)} \\ \vdots \\ \boldsymbol{m}_{nT}^{(11)} \\ \boldsymbol{m}_{nT}^{(21)} \\ \boldsymbol{m}_1^{(12)} \\ \boldsymbol{m}_1^{(22)} \\ \vdots \\ \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{m}_{kT}^{(22)} \end{pmatrix}.$$

Now we have

$$
\boldsymbol{H}_{11}\operatorname{vec}\boldsymbol{\Sigma}_{11} = \begin{pmatrix} \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1^{(11)} \\ \boldsymbol{m}_2^{(11)} \\ \boldsymbol{m}_3^{(11)} \\ \vdots \\ \boldsymbol{m}_{nT}^{(11)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_1^{(11)} \\ \boldsymbol{0} \\ \boldsymbol{m}_2^{(11)} \\ \boldsymbol{0} \\ \boldsymbol{m}_3^{(11)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{m}_{nT}^{(11)} \\ \boldsymbol{0} \\ \vdots \\ \vdots \\ \boldsymbol{0} \end{pmatrix},
$$

$$
\boldsymbol{H}_{21}\operatorname{vec}\boldsymbol{\Sigma}_{21} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1^{(21)} \\ \boldsymbol{m}_2^{(21)} \\ \boldsymbol{m}_3^{(21)} \\ \vdots \\ \boldsymbol{m}_{nT}^{(21)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{m}_1^{(21)} \\ \boldsymbol{0} \\ \boldsymbol{m}_2^{(21)} \\ \boldsymbol{0} \\ \boldsymbol{m}_3^{(21)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{m}_{nT}^{(21)} \\ \vdots \\ \vdots \\ \boldsymbol{0} \end{pmatrix},
$$

$$
\boldsymbol{H}_{12}\operatorname{vec}\boldsymbol{\Sigma}_{12} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{I}_{nT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{nT} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{nT} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{nT} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{m}_1^{(12)} \\ \boldsymbol{m}_2^{(12)} \\ \boldsymbol{m}_3^{(12)} \\ \vdots \\ \boldsymbol{m}_{kT}^{(12)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{m}_1^{(12)} \\ \boldsymbol{0} \\ \boldsymbol{m}_2^{(12)} \\ \boldsymbol{0} \\ \boldsymbol{m}_3^{(12)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{m}_{kT}^{(12)} \\ \boldsymbol{0} \end{pmatrix},
$$

and

$$
\boldsymbol{H}_{22}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{22}=\begin{pmatrix}
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\
\boldsymbol{I}_{kT} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\
\boldsymbol{0} & \boldsymbol{I}_{kT} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{kT} & \cdots & \boldsymbol{0}\\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_{kT}
\end{pmatrix}
\begin{pmatrix}
\boldsymbol{m}_1^{(22)}\\
\boldsymbol{m}_2^{(22)}\\
\boldsymbol{m}_3^{(22)}\\
\vdots\\
\boldsymbol{m}_{kT}^{(22)}
\end{pmatrix}
=\begin{pmatrix}
\boldsymbol{0}\\
\vdots\\
\vdots\\
\boldsymbol{0}\\
\boldsymbol{m}_1^{(22)}\\
\boldsymbol{0}\\
\boldsymbol{m}_2^{(22)}\\
\boldsymbol{0}\\
\boldsymbol{m}_3^{(22)}\\
\boldsymbol{0}\\
\vdots\\
\boldsymbol{m}_{kT}^{(22)}
\end{pmatrix},
$$

therefore, it is easy to see that

$$
\boldsymbol{H}_{11}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{11}+\boldsymbol{H}_{21}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{21}+\boldsymbol{H}_{12}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{12}+\boldsymbol{H}_{22}\,\mathrm{vec}\,\boldsymbol{\Sigma}_{22}=\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})
$$

as required.

Q.E.D.

Appendix §3B

Proof of Proposition 3.2.2.3 Firstly note that differentiating the log-likelihood (??) is equivalent to differentiating

$$
\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)}}=-\frac{N}{2}\frac{\partial \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial \boldsymbol{\theta}_j^{(*)}}-\frac{1}{2}\frac{\partial\,\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}}{\partial \boldsymbol{\theta}_j^{(*)}},
$$

where, by the chain rule for matrix calculus, the first term evaluates to

$$
\frac{\partial \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial \boldsymbol{\theta}_j^{(*)}}=\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\frac{\partial \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}\right)=\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}),
$$

and for the second term we obtain

$$
\begin{aligned}
\frac{\partial\,\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}}{\partial \boldsymbol{\theta}_j^{(*)}}
&=\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}\right)\left(\frac{\partial\,\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}}{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}\right)\\
&=-\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\mathrm{vec}\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT},
\end{aligned}
$$

where we used the results

$$
\frac{\partial \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})|}{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}=\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}),
$$

and

$$
\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}(\boldsymbol{\theta})}=-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}).
$$

Differentiating the log-likelihood now yields

$$
\begin{aligned}
\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)}} &= -\frac{N}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right) \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \\
&\quad + \frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right) \operatorname{vec} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \\
&= \frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left(\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \operatorname{vec} \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right) \\
&= \frac{1}{2}\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}}\right)\left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \quad (3.38)
\end{aligned}
$$

which is equivalent to (3.29), as required.

Q.E.D.

Appendix §3C

Proof of Proposition 3.2.2.4 We derive the components $\partial \operatorname{vec} \boldsymbol{\Sigma}_{ij} / \partial \boldsymbol{\theta}_j^{(*)}$ for each $\boldsymbol{\Sigma}_{ij}$ block, in turn. The derivatives for $\operatorname{vec} \boldsymbol{\Sigma}_{11}$ are obtained as follows. For $\boldsymbol{\Sigma}_{11}$ we obtain the derivative in respect to particular components of $\boldsymbol{\theta}$ as follows. Using the result that

$$
\frac{\partial \operatorname{vec}\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)}{\partial \operatorname{vec} \boldsymbol{B}_i} = -\frac{\partial \operatorname{vec}\left(\boldsymbol{S}_T^i \otimes \boldsymbol{B}_i\right)}{\partial \operatorname{vec} \boldsymbol{B}_i} = -\boldsymbol{K}_{T,m}^{*}\left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S}'^{i}_T\right) \otimes \boldsymbol{I}_m,
$$

we obtain the partial derivative in respect to $\operatorname{vec} \boldsymbol{B}_i$ as

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)'}{\partial \operatorname{vec} \boldsymbol{B}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)}{\partial \operatorname{vec} \boldsymbol{B}_i} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)} \right) \\
&\times \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1}}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}} \right) \\
&\times \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1}} \right) \\
&= \left[\boldsymbol{K}_{T,m}^{*} \left(\boldsymbol{I}_{mT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_m\right] \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \right] \\
&\times \left(\left[\boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \otimes \boldsymbol{I}_{mT} \right] + \left[\boldsymbol{Y}' \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \otimes \boldsymbol{I}_{mT} \right] \boldsymbol{K}_{mT,mT} \right) \\
&\times \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right) \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right).
\end{aligned}
$$

Next, we obtain

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \frac{\partial \operatorname{vec} \boldsymbol{\Lambda} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{F} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)' \boldsymbol{\Lambda}'}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{Y} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)} \right) \\
&\times \left(\frac{\partial \operatorname{vec} \boldsymbol{\Lambda} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{F} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)' \boldsymbol{\Lambda}'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{F} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)'} \right) \\
&= \left[\boldsymbol{K}_{T,g}^{*} \left(\boldsymbol{I}_{Tg} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_m\right] \\
&\times \left[\boldsymbol{Y} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)' \otimes \boldsymbol{I}_{mT} + \boldsymbol{Y}' \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)' \otimes \boldsymbol{I}_{mT}\right) \Big] \boldsymbol{K}_{mT,mT} \left(\boldsymbol{\Lambda}' \otimes \boldsymbol{\Lambda}'\right),
\end{aligned}
$$

where we used the result that

$$
\frac{\partial \operatorname{vec} \left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j \right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} = \frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} = \boldsymbol{K}_{T,g}^{*} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_m.
$$

The derivative in respect to $\operatorname{vec} \boldsymbol{\Lambda}_y$ is obtained as

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{X} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{X} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right)}\right) \\
&= \left(\boldsymbol{K}_{T,m}^{*} \otimes \boldsymbol{I}_n\right) \left(\left[\boldsymbol{X} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right) \otimes \boldsymbol{I}_{nT}\right] + \left[\boldsymbol{X}' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right) \otimes \boldsymbol{I}_{nT}\right] \boldsymbol{K}_{nT,nT}\right),
\end{aligned}
$$

$$
\operatorname{vec} \boldsymbol{\Sigma}_{11} = \operatorname{vec} \boldsymbol{L} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \boldsymbol{Y} \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \boldsymbol{L}' + \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon\right). \tag{8.39}
$$

To obtain the derivatives in respect to vech $\boldsymbol{\Phi}_0$ and vech $\boldsymbol{\Phi}_i$ firstly note that for a symmetrical $a \times a$ matrix $\boldsymbol{X}$, $\partial \operatorname{vec} \boldsymbol{X} / \partial \operatorname{vech} \boldsymbol{X} = \boldsymbol{D}_a'$. Hence we have

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_0}{\partial \operatorname{vech} \boldsymbol{\Phi}_0}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right) \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right) \\
&= \boldsymbol{D}_g' \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right) \boldsymbol{Z}'}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}\right) \\
&= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right] \left(\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T\right) \left(\boldsymbol{Z}' \otimes \boldsymbol{Z}'\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i}\right) \frac{\partial \operatorname{vec} \boldsymbol{Z} \left[\sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right] \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \\
&= \boldsymbol{D}_g' \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left[\sum_{j=1}^{q} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j\right)\right] \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} + \frac{\partial \operatorname{vec} \boldsymbol{Z} \left[\sum_{j=1}^{q} \left(\boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}_j'\right)\right] \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \boldsymbol{D}_g' \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right) \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} + \frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)' \boldsymbol{Z}'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \boldsymbol{D}_g' \left[\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right) \boldsymbol{Z}'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}\right)\right. \\
&\quad + \left.\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{Z} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)' \boldsymbol{Z}'}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'}\right)\right] \\
&= \boldsymbol{D}_g' \left[\boldsymbol{K}_{T,g}^{*} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_g\right] \left(\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}\right) \left(\boldsymbol{Z}' \otimes \boldsymbol{Z}'\right)
\end{aligned}
$$

while for vech $\boldsymbol{\Psi}$,

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Psi}} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Psi}}{\partial \operatorname{vech} \boldsymbol{\Psi}}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{D} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{D}'}{\partial \operatorname{vec} \boldsymbol{\Psi}}\right) \\
&= \boldsymbol{D}_m' \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)}{\partial \operatorname{vec} \boldsymbol{\Psi}}\right) \left(\frac{\partial \operatorname{vec} \boldsymbol{D} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right) \boldsymbol{D}'}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\right)}\right) \\
&= \boldsymbol{D}_m' \left[\boldsymbol{I}_m \otimes \left(\operatorname{vec} \boldsymbol{I}_T\right)'\right] \left(\boldsymbol{K}_{m,T} \otimes \boldsymbol{I}_T\right) \left(\boldsymbol{D}' \otimes \boldsymbol{D}'\right).
\end{aligned}
$$

Finally, we have

$$\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Theta}_\varepsilon}{\partial \operatorname{vech} \boldsymbol{\Theta}_\varepsilon}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\varepsilon\right)}{\partial \operatorname{vec} \boldsymbol{\Theta}_\varepsilon}\right) = \boldsymbol{D}_n' \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T).$$

The derivatives of $\operatorname{vec} \boldsymbol{\Sigma}_{12}$ are similarly obtained as

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{B}_i} &= \frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \left.\quad \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left.\quad \sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{F} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x')}{\partial \operatorname{vec} \boldsymbol{B}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left.\quad \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)}{\partial \operatorname{vec} \boldsymbol{B}_i}\right) \left(\frac{\partial \operatorname{vec} \left.\quad \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}}{\partial \operatorname{vec} \left.\quad \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)}\right) \\
&\times \frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \left.\quad \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left.\quad \sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{F} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x')}{\partial \operatorname{vec} \left.\quad \boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1}} \\
&= \left[\boldsymbol{K}_{T,m}^{*} \quad \boldsymbol{I}_{mT} \otimes \boldsymbol{S'}_T^{i} \quad \otimes \boldsymbol{I}_m\right] \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \otimes \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1}\right] \\
&\times \left(\left[\left(\sum_{j=0}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Gamma}_j\right) \boldsymbol{F} \quad \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x' \quad\right] \otimes (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)'\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} &= \frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Gamma}_i}\right) \left(\frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S}_T^j \otimes \boldsymbol{B}_j\right)^{-1} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right) \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Gamma}_i\right)}\right) \\
&= \left[\boldsymbol{K}_{T,g}^{\bar{\tau}T} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i\right) \otimes \boldsymbol{I}_m\right] \left(\left[\boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right] \otimes \left[\left(\boldsymbol{I}_{mT} - \sum_{j=0}^{p} \boldsymbol{S'}_T^j \otimes \boldsymbol{B}_j'\right)^{-1} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right]\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} &= \frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \boldsymbol{Q} \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y} \\
&= \left(\frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_y}\right) \left(\frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \boldsymbol{Q} \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y)}\right) \\
&= \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{n,T} \otimes \boldsymbol{I}_T) \left(\left[\boldsymbol{Q} \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right] \otimes \boldsymbol{I}_{nT}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{11}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \boldsymbol{Q} \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} \\
&= \left(\frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x}\right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)}\right) \left(\frac{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y) \boldsymbol{Q} \boldsymbol{F} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} (\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x)}\right) \\
&= \left[\boldsymbol{I}_n \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right] (\boldsymbol{K}_{k,T} \otimes \boldsymbol{I}_T) \boldsymbol{K}_{k,T} \left(\boldsymbol{I}_{gT} \otimes \boldsymbol{F} \boldsymbol{Q}' \left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right]\right),
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_0}{\partial \operatorname{vech} \boldsymbol{\Phi}_0}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right) \\
&= \boldsymbol{D}_g' \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}\right) \\
&= \boldsymbol{D}_g' \left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right]\left(\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T\right)\left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{Q}'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{12}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q}\left[\sum_{j=1}^{q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S}'^j_T \otimes \boldsymbol{\Phi}_j'\right)\right]\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right. \\
&+ \left.\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q}\left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right) \\
&= \boldsymbol{D}_g'\left[\left(\frac{\partial \operatorname{vec}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)}\right)\right. \\
&+ \left.\left(\frac{\partial \operatorname{vec}\left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i}\right)\left(\frac{\partial \operatorname{vec}\left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)'}{\partial \operatorname{vec}\left(\boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i\right)}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y\right) \boldsymbol{Q}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec}\left(\boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i\right)'}\right)\right] \\
&= \boldsymbol{D}_g'\left[\boldsymbol{K}^*_{T,g}\left(\boldsymbol{I}_{gT} \otimes \boldsymbol{S}'^i_T\right) \otimes \boldsymbol{I}_g\right]\left(\boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT}\right)\left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{Q}'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_y'\right)\right].
\end{aligned}
$$

Lastly, the derivatives of vec $\boldsymbol{\Sigma}_{22}$ are obtained as follows

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} &= \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x} \\
&= \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)}{\partial \operatorname{vec} \boldsymbol{\Lambda}_x}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)}\right) \\
&= \left(\boldsymbol{K}^*_{T,g} \otimes \boldsymbol{I}_k\right)\left(\left[\boldsymbol{F}\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{I}_{kT}\right] + \left[\boldsymbol{F}'\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \boldsymbol{I}_{kT}\right] \boldsymbol{K}_{k,T}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Phi}_0} &= \left(\frac{\partial \operatorname{vec} \boldsymbol{\Phi}_0}{\partial \operatorname{vech} \boldsymbol{\Phi}_0}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right) \\
&= \boldsymbol{D}_g'\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_0}\right)\left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Phi}_0\right)}\right) \\
&= \boldsymbol{D}_g'\left[\boldsymbol{I}_g \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right]\left(\boldsymbol{K}_{g,T} \otimes \boldsymbol{I}_T\right)\left[\left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right) \otimes \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x'\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} &= \frac{\partial \operatorname{vec} \boldsymbol{\Phi}_i}{\partial \operatorname{vech} \boldsymbol{\Phi}_i} \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \left[\sum\limits_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j + \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right] \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x\right)}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \\
&= \boldsymbol{D}'_g \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \left[\sum\limits_{j=1}^{q} \boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_j \right] \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right. \\
&+ \left. \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \left[\sum\limits_{j=1}^{q} \boldsymbol{S'}_T^j \otimes \boldsymbol{\Phi}'_j \right] \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \\
&= \boldsymbol{D}'_g \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \;\; \boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i \;\; \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} + \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \;\; \boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i \;\;' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \\
&= \boldsymbol{D}'_g \left[\left(\frac{\partial \operatorname{vec} \;\; \boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \;\; \boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i \;\; \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \;\; \boldsymbol{S}_T^j \otimes \boldsymbol{\Phi}_i} \right) \right. \\
&+ \left. \frac{\partial \operatorname{vec} \;\; \boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i}{\partial \operatorname{vec} \boldsymbol{\Phi}_i} \Bigg) \;\; \frac{\partial \operatorname{vec} \;\; \boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i \;\;'}{\partial \operatorname{vec} \;\; \boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i} \Bigg) \;\; \frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right) \;\; \boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i \;\;' \left(\boldsymbol{I}_T \otimes \boldsymbol{\Lambda}_x\right)'}{\partial \operatorname{vec} \;\; \boldsymbol{S}_T^i \otimes \boldsymbol{\Phi}_i \;\;'} \Bigg) \right] \\
&= \boldsymbol{D}'_g \left[\boldsymbol{K}^*_{T,g} \;\; \boldsymbol{I}_{gT} \otimes \boldsymbol{S'}_T^i \;\; \otimes \boldsymbol{I}_g \right] \;\; \boldsymbol{I}_{gT} + \boldsymbol{K}_{gT,gT} \quad \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x \;\; \otimes \;\; \boldsymbol{I}_T \otimes \boldsymbol{\Lambda}'_x \quad ,
\end{aligned}
$$

and

$$
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_{22}}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} = \left(\frac{\partial \operatorname{vec} \boldsymbol{\Theta}_\delta}{\partial \operatorname{vech} \boldsymbol{\Theta}_\delta} \right) \left(\frac{\partial \operatorname{vec} \left(\boldsymbol{I}_T \otimes \boldsymbol{\Theta}_\delta\right)}{\partial \operatorname{vec} \boldsymbol{\Theta}_\delta} \right) = \boldsymbol{D}'_k \left(\boldsymbol{I}_k \otimes (\operatorname{vec} \boldsymbol{I}_T)'\right) \left(\boldsymbol{K}_{kT} \otimes \boldsymbol{I}_T\right).
$$

The remaining derivatives are zero trivially in all cases where particular component of the parameter vector $\boldsymbol{\theta}$ is not contained in $\boldsymbol{\Sigma}_{ij}$.

Q.E.D.

Appendix §3D

Proof of Proposition 3.2.3.1 We obtain the general form of the second partial derivative (3.33) by differentiating the typical element of the score vector

$$
\frac{\partial \ln L\left(\tilde{\boldsymbol{W}}_{NT}\right)}{\partial \boldsymbol{\theta}_j^{(*)}} = \frac{1}{2} \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j^{(*)}} \right) \left[\operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\, \tilde{\boldsymbol{W}}_{NT} \tilde{\boldsymbol{W}}'_{NT} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N \operatorname{vec} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \right] \tag{8.40}
$$

in respect to the component $\boldsymbol{\theta}_i^{(*)}$ of the parameter vector $\boldsymbol{\theta}$. Note that (8.40) is the partial derivative of the log-likelihood (3.16) in respect to the component $\boldsymbol{\theta}_j^{(*)}$ of the parameter vector $\boldsymbol{\theta}$. We make use of the generalised product rule for matrix calculus

$$
\frac{\partial\, G(\boldsymbol{z}) h(\boldsymbol{z})}{\partial\, \boldsymbol{z}} = \frac{\partial \operatorname{vec} G(\boldsymbol{z})}{\partial\, \boldsymbol{z}} \left[h(\boldsymbol{z}) \otimes \mathbf{I}_d\right] + \frac{\partial\, h(\boldsymbol{z})}{\partial\, \boldsymbol{z}} G(\boldsymbol{z})' \tag{8.41}
$$

where $G(\boldsymbol{z})$ is a matrix function of the vector $\boldsymbol{z}$, $h(\boldsymbol{z})$ is a vector function of $\boldsymbol{z}$, and d is the dimension of the vector $\boldsymbol{z}$. Letting $\partial \operatorname{vec} \boldsymbol{\Sigma}(\boldsymbol{\theta}) / \partial\, \boldsymbol{\theta}_j^{(*)} \equiv G(\boldsymbol{z})$ and

$$\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - N\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \equiv h(\boldsymbol{z}),$$

we firstly differentiate the two additive components of $\partial\, h(\boldsymbol{z})/\partial\, \boldsymbol{z}$. Differentiating the first component of $h(\boldsymbol{z})$ in respect to $\boldsymbol{\theta}_i^{(*)}$ we obtain

$$\begin{aligned}
&\frac{\partial \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)}{\partial\,\boldsymbol{\theta}_i^{(*)}}\\
&= \left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\,\boldsymbol{\theta}_i^{(*)}}\right)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}\right)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}\right)\\
&= -\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\,\boldsymbol{\theta}_i^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}\right)\\
&= -\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\,\boldsymbol{\theta}_i^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)\\
&\times\left(\left[\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{I}_{mT}\right]+\left[\boldsymbol{I}_{mT}\otimes\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right), \qquad (8.42)
\end{aligned}$$

where we used the result

$$\begin{aligned}
\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})} &= \left[\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{I}_{mT}\right] \qquad (8.43)\\
&+\left[\boldsymbol{I}_{mT}\otimes\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \qquad (8.44)
\end{aligned}$$

For the second component we have

$$\begin{aligned}
\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial\,\boldsymbol{\theta}_i^{(*)}} &= -\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\,\boldsymbol{\theta}_i^{(*)}}\right)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})}{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}\right)\\
&= -\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\,\boldsymbol{\theta}_i^{(*)}}\right)\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)
\end{aligned}$$

Substituting (8.42) and (8.44) into (8.41) yields

$$\begin{aligned}
\frac{\partial^2 \ln L \quad \tilde{\boldsymbol{W}}_{NT}}{\partial \;_j^{(*)}\partial \;_i^{(*)\prime}} &= \frac{1}{2}\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}(\;)}{\partial \;_j^{(*)}\partial \;_i^{\prime(*)}}\right)\left[\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\;)\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\;) - N\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\;)\right]\otimes\boldsymbol{I}_{p_i}\\
&+ \frac{1}{2}\quad \frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\;)\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\;)}{\partial \;_i^{(*)}} - N\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}(\;)}{\partial \;_i^{(*)}}\Bigg)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\;)}{\partial \;_j^{(*)}}\right)'\\
&= \frac{1}{2}\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}(\;)}{\partial \;_j^{(*)}\partial \;_i^{\prime(*)}}\right)\left[\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\;)\,\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\;) - N\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\;)\right]\otimes\boldsymbol{I}_{p_i}\\
&- \frac{1}{2}\Bigg[\quad \frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\;)}{\partial \;_i^{(*)}}\Bigg)\;\boldsymbol{\Sigma}^{-1}(\;)\otimes\boldsymbol{\Sigma}^{-1}(\;)\\
&\times \quad \left[\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\;)\otimes\boldsymbol{I}_{mT}\right]-\left[\boldsymbol{I}_{mT}\otimes\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}'_{NT}\boldsymbol{\Sigma}^{-1}(\;)\right]\\
&- \quad N\quad \frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\;)}{\partial \;_i^{(*)}}\Bigg)\;\boldsymbol{\Sigma}^{-1}(\;)\otimes\boldsymbol{\Sigma}^{-1}(\;)\;\Bigg]\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}(\;)}{\partial \;_j^{(*)}}\right)',
\end{aligned}$$

which gives the expression (3.33), as required.

Q.E.D.

Appendix §3E

Proof of Proposition 3.2.3.2 We will show that the probability limit of the typical element of the Hessian matrix (3.33) is given by (3.34). By (3.24) and (3.25) it follows that

$$\underset{T,N\to\infty}{p\lim}\left[\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{1}{N}\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]=\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}),$$

and hence

$$\underset{T,N\to\infty}{p\lim}\left[\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{1}{N}\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]=\boldsymbol{0}. \tag{8.45}$$

Therefore, the first term converges in probability to zero,

$$\underset{T,N\to\infty}{p\lim}\frac{1}{2}\left(\frac{\partial^2\operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j^{(*)}\partial\boldsymbol{\theta}_i^{(*)\prime}}\right)\left(\left[\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{1}{N}\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})-\operatorname{vec}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\otimes\boldsymbol{I}_{p_i}\right)=\boldsymbol{0}.$$

Next, note that

$$\underset{T,N\to\infty}{p\lim}\left(\frac{1}{N}\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{I}_{mT}\right)=\boldsymbol{I}_{mT}\otimes\boldsymbol{I}_{mT},$$

and

$$\underset{T,N\to\infty}{p\lim}\left(\boldsymbol{I}_{mT}\otimes\frac{1}{N}\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)=\boldsymbol{I}_{mT}\otimes\boldsymbol{I}_{mT},$$

thus we have

$$\underset{T,N\to\infty}{p\lim}\left(\left[\frac{1}{N}\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{I}_{mT}\right]-\left[\boldsymbol{I}_{mT}\otimes\frac{1}{N}\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right)=\boldsymbol{0}.$$

This implies that the second term converges in probability to zero,

$$\begin{aligned}\underset{T,N\to\infty}{p\lim}\;&\frac{1}{2}\left(\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i^{(*)}}\right)\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\\&\times\frac{1}{N}\left(\left[\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{I}_{mT}\right]-\left[\boldsymbol{I}_{mT}\otimes\tilde{\boldsymbol{W}}_{NT}\tilde{\boldsymbol{W}}_{NT}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right)=\boldsymbol{0}\end{aligned}$$

This leaves us with the remaining term as required by (3.34).

Q.E.D

8.3 Chapter §5 appendices

Appendix §5A

Proof of Lemma 5.2.2.3 Results (i) and (iv) follow directly from assumption A1 and Lemma (5.2.2.2). To prove (ii) note that the convergence in mean square in Lemma (5.2.2.2) implies that

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_{ijt-r}-\mu_{ij}^{(y)}\right)^2\right] &= E\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_{ijt-r}^2-2y_{ijt-r}\mu_{ij}^{(y)}+\left(\mu_{ij}^{(y)}\right)^2\right)\right] \\ &= E\left[\frac{1}{T}\sum_{t=1}^{T}y_{ijt-r}^2\right]-\left(\mu_{ij}^{(y)}\right)^2, \end{aligned}$$

where for $r = w$, this is equal to γ_0, thus

$$E\left[\frac{1}{T}\sum_{t=1}^{T}y_{ijt-r}^2\right]=\gamma_0^{(ij)}+\left(\mu_{ij}^{(y)}\right)^2,$$

and similarly for (v),

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}\left(x_{ijt-r}-\mu_{ij}^{(x)}\right)^2\right] &= E\left[\frac{1}{T}\sum_{t=1}^{T}\left(x_{ijt-r}^2-2x_{ijt-r}\mu_{ij}^{(x)}+\left(\mu_{ij}^{(x)}\right)^2\right)\right] \\ &= E\left[\frac{1}{T}\sum_{t=1}^{T}x_{ijt-r}^2\right]-\left(\mu_{ij}^{(x)}\right)^2 \end{aligned}$$

$$\Rightarrow E\left[\frac{1}{T}\sum_{t=1}^{T}x_{ijt-r}^2\right]=\delta_0^{(ij)}+\left(\mu_{ij}^{(x)}\right)^2,$$

establishing (v).

To show (iii), the above is easily generalised for higher moments by noting that

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}(y_{ij,t-r}-\mu_{ij}^{(y)})(y_{ef,t-w}-\mu_{ef}^{(y)})\right] &= E\left[\frac{1}{T}\sum_{t=1}^{T}y_{ij,t-r}y_{ef,t-w}\right]-\mu_{ij}^{(y)}\mu_{ef}^{(y)} \\ &= \gamma_{|r-w|}^{(ijef)}-\mu_{ij}^{(y)}\mu_{ef}^{(y)}, \end{aligned}$$

and similarly for (vi)

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}(x_{ij,t-r}-\mu_{ij}^{(x)})(x_{ef,t-w}-\mu_{ef}^{(x)})\right] &= E\left[\frac{1}{T}\sum_{t=1}^{T}x_{ij,t-r}x_{ef,t-w}\right]-\mu_{ij}^{(x)}\mu_{ef}^{(x)} \\ &= \delta_{|r-w|}^{(ijef)}-\mu_{ij}^{(x)}\mu_{ef}^{(x)}, \end{aligned}$$

Finally, for (vii) we have

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}(y_{ij,t-r}-\mu_{ij}^{(y)})(x_{ef,t-w}-\mu_{ef}^{(x)})\right] &= E\left[\frac{1}{T}\sum_{t=1}^{T}y_{ij,t-r}x_{ef,t-w}\right]-\mu_{ij}^{(y)}\mu_{ef}^{(x)} \\ &= \psi_{|r-w|}^{(ijef)}-\mu_{ij}^{(x)}\mu_{ef}^{(x)}. \end{aligned}$$

Q.E.D.

Appendix §5B

Proof of Proposition 5.2.2.4 To show that $\operatorname{plim}\left(T^{-1}\mathbf{Z}'\mathbf{Z}\right) = \mathbf{\Sigma}_{ZZ}$, where $\mathbf{\Sigma}_{ZZ} \neq \mathbf{0}$, define a general matrix of eligible instruments $\mathbf{Z} \equiv (\mathbf{Y}_1^*, \mathbf{Y}_2^*, \mathbf{X}_1^*, \mathbf{X}_2^*)$. First, we have

$$\mathbf{Z}'\mathbf{Z} = \begin{pmatrix} \mathbf{Y}_1^{*\prime}\mathbf{Y}_1^* & \mathbf{Y}_1^{*\prime}\mathbf{Y}_2^* & \mathbf{Y}_1^{*\prime}\mathbf{X}_1^* & \mathbf{Y}_1^{*\prime}\mathbf{X}_2^* \\ \mathbf{Y}_2^{*\prime}\mathbf{Y}_1^* & \mathbf{Y}_2^{*\prime}\mathbf{Y}_2^* & \mathbf{Y}_2^{*\prime}\mathbf{X}_1^* & \mathbf{Y}_2^{*\prime}\mathbf{X}_2^* \\ \mathbf{X}_1^{*\prime}\mathbf{Y}_1^* & \mathbf{X}_1^{*\prime}\mathbf{Y}_2^* & \mathbf{X}_1^{*\prime}\mathbf{X}_1^* & \mathbf{X}_1^{*\prime}\mathbf{X}_2^* \\ \mathbf{X}_2^{*\prime}\mathbf{Y}_1^* & \mathbf{X}_2^{*\prime}\mathbf{Y}_2^* & \mathbf{X}_2^{*\prime}\mathbf{X}_1^* & \mathbf{X}_2^{*\prime}\mathbf{X}_2^* \end{pmatrix}$$

The upper-left block of $\mathbf{Z}'\mathbf{Z}$ is given by

$$\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^* = \begin{pmatrix} \mathbf{y}_{p-1}^{(1j)\prime} \\ \mathbf{y}_{p-2}^{(1j)\prime} \\ \vdots \\ \mathbf{y}_{p-a}^{(1j)\prime} \end{pmatrix} \begin{pmatrix} \mathbf{y}_{p-1}^{(1j)} & \mathbf{y}_{p-2}^{(1j)} & \cdots & \mathbf{y}_{p-a}^{(1j)} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{p-1}^{(1j)} & \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{p-2}^{(1j)} & \cdots & \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{p-a}^{(1j)} \\ \mathbf{y}_{p-2}^{(1j)\prime}\mathbf{y}_{p-1}^{(1j)} & \mathbf{y}_{p-2}^{(1j)\prime}\mathbf{y}_{p-2}^{(1j)} & \cdots & \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{p-a}^{(1j)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{p-a}^{(1j)\prime}\mathbf{y}_{p-1}^{(1j)} & \mathbf{y}_{p-a}^{(1j)\prime}\mathbf{y}_{p-2}^{(1j)} & \cdots & \mathbf{y}_{p-a}^{(1j)\prime}\mathbf{y}_{p-a}^{(1j)} \end{pmatrix}$$

Note that, e.g., the upper-left block element of the $\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^*$ matrix is of the form

$$\mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{p-1}^{(1j)} = \begin{pmatrix} y_{-p-1}^{(11)} & y_{1-p-1}^{(11)} & y_{2-p-1}^{(11)} & \cdots & y_{T-p-1}^{(11)} \\ y_{-p-1}^{(12)} & y_{1-p-1}^{(12)} & y_{2-p-1}^{(12)} & \cdots & y_{T-p-1}^{(12)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{-p-1}^{(1m)} & y_{1-p-1}^{(1m)} & y_{2-p-1}^{(1m)} & \cdots & y_{T-p-1}^{(1m)} \end{pmatrix} \begin{pmatrix} y_{-p-1}^{(11)} & y_{-p-1}^{(12)} & \cdots & y_{-p-1}^{(1m)} \\ y_{1-p-1}^{(11)} & y_{1-p-1}^{(12)} & \cdots & y_{1-p-1}^{(1m)} \\ y_{2-p-1}^{(11)} & y_{2-p-1}^{(12)} & \cdots & y_{2-p-1}^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-p-1}^{(11)} & y_{T-p-1}^{(12)} & \cdots & y_{T-p-1}^{(1m)} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{t=1}^{T}\left(y_{t-p-1}^{(11)}\right)^2 & \sum_{t=1}^{T} y_{t-p-1}^{(11)}y_{t-p-1}^{(12)} & \cdots & \sum_{t=1}^{T} y_{t-p-1}^{(11)}y_{t-p-1}^{(1m)} \\ \sum_{t=1}^{T} y_{t-p-1}^{(12)}y_{t-p-1}^{(11)} & \sum_{t=1}^{T}\left(y_{t-p-1}^{(12)}\right)^2 & \cdots & \sum_{t=1}^{T} y_{t-p-1}^{(12)}y_{t-p-1}^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^{T} y_{t-p-1}^{(1m)}y_{t-p-1}^{(11)} & \sum_{t=1}^{T} y_{t-p-1}^{(m1)}y_{t-p-1}^{(12)} & \cdots & \sum_{t=1}^{T}\left(y_{t-p-1}^{(1m)}\right)^2 \end{pmatrix}$$

Taking probability limits gives

$$\operatorname{plim}\left(\tfrac{1}{T}\mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{p-1}^{(1j)}\right) =$$
$$\begin{pmatrix} \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{t-p-1}^{(11)}\right)^2\right) & \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-p-1}^{(11)}y_{t-p-1}^{(12)}\right) & \cdots & \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-p-1}^{(11)}y_{t-p-1}^{(1m)}\right) \\ \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-p-1}^{(12)}y_{t-p-1}^{(11)}\right) & \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{t-p-1}^{(12)}\right)^2\right) & \cdots & \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-p-1}^{(12)}y_{t-p-1}^{(1m)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-p-1}^{(1m)}y_{t-p-1}^{(11)}\right) & \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-p-1}^{(m1)}y_{t-p-1}^{(12)}\right) & \cdots & \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{t-p-1}^{(1m)}\right)^2\right) \end{pmatrix}$$

We can now apply the convergence results implied by Lemma (5.2.2.3) to obtain

$$\text{plim}\left(\frac{1}{T}\mathbf{y}_{p-1}^{(1j)'}\mathbf{y}_{p-1}^{(1j)}\right) = \begin{pmatrix} E\left[\left(y_{t-p-1}^{(11)}\right)^2\right] & E\left[y_{t-p-1}^{(11)}y_{t-p-1}^{(12)}\right] & \cdots & E\left[y_{t-p-1}^{(11)}y_{t-p-1}^{(1m)}\right] \\ E\left[y_{t-p-1}^{(12)}y_{t-p-1}^{(11)}\right] & E\left[\left(y_{t-p-1}^{(12)}\right)^2\right] & \cdots & E\left[y_{t-p-1}^{(12)}y_{t-p-1}^{(1m)}\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[y_{t-p-1}^{(1m)}y_{t-p-1}^{(11)}\right] & E\left[y_{t-p-1}^{(m1)}y_{t-p-1}^{(12)}\right] & \cdots & E\left[\left(y_{t-p-1}^{(1m)}\right)^2\right] \end{pmatrix}$$

$$= \begin{pmatrix} \gamma_0^{(11)} + (\mu_{11}^{(y)})^2 & \gamma_0^{(11,12)} + \mu_{11}^{(y)}\mu_{12}^{(y)} & \cdots & \gamma_0^{(11,1m)} + \mu_{11}^{(y)}\mu_{1m}^{(y)} \\ \gamma_0^{(12,11)} + \mu_{12}^{(y)}\mu_{11}^{(y)} & \gamma_0^{(12)} + (\mu_{12}^{(y)})^2 & \cdots & \gamma_0^{(12,1m)} + \mu_{12}^{(y)}\mu_{1m}^{(y)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_0^{(1m,11)} + \mu_{1m}^{(y)}\mu_{11}^{(y)} & \gamma_0^{(1m,12)} + \mu_{1m}^{(y)}\mu_{12}^{(y)} & \cdots & \gamma_0^{(1m)} + (\mu_{1m}^{(y)})^2 \end{pmatrix},$$

which gives the required results for the upper-left block element of $\mathbf{Y}_1^{*'}\mathbf{Y}_1^*$. However, note that all other block elements of this matrix can be written in the general form

$$\mathbf{y}_{p-r}^{(1j)'}\mathbf{y}_{p-w}^{(1j)} = \begin{pmatrix} y_{-p-r}^{(11)} & y_{1-p-r}^{(11)} & y_{2-p-r}^{(11)} & \cdots & y_{T-p-r}^{(11)} \\ y_{-p-r}^{(12)} & y_{1-p-r}^{(12)} & y_{2-p-r}^{(12)} & \cdots & y_{T-p-r}^{(12)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{-p-r}^{(1m)} & y_{1-p-r}^{(1m)} & y_{2-p-r}^{(1m)} & \cdots & y_{T-p-r}^{(1m)} \end{pmatrix} \begin{pmatrix} y_{-p-w}^{(11)} & y_{-p-w}^{(12)} & \cdots & y_{-p-w}^{(1m)} \\ y_{1-p-w}^{(11)} & y_{1-p-w}^{(12)} & \cdots & y_{1-p-w}^{(1m)} \\ y_{2-p-w}^{(11)} & y_{2-p-w}^{(12)} & \cdots & y_{2-p-w}^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-p-w}^{(11)} & y_{T-p-w}^{(12)} & \cdots & y_{T-p-w}^{(1m)} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{t=1}^{T} y_{t-p-r}^{(11)}y_{t-p-w}^{(11)} & \sum_{t=1}^{T} y_{t-p-r}^{(11)}y_{t-p-w}^{(12)} & \cdots & \sum_{t=1}^{T} y_{t-p-r}^{(11)}y_{t-p-w}^{(1m)} \\ \sum_{t=1}^{T} y_{t-p-r}^{(12)}y_{t-p-w}^{(11)} & \sum_{t=1}^{T} y_{t-p-r}^{(12)}y_{t-p-w}^{(12)} & \cdots & \sum_{t=1}^{T} y_{t-p-r}^{(12)}y_{t-p-w}^{(1m)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^{T} y_{t-p-r}^{(1m)}y_{t-p-w}^{(11)} & \sum_{t=1}^{T} y_{t-p-r}^{(m1)}y_{t-p-w}^{(12)} & \cdots & \sum_{t=1}^{T} y_{t-p-r}^{(1m)}y_{t-p-w}^{(1m)} \end{pmatrix}$$

and thus, from Lemma (5.2.2.3), it follows that each block element of $\mathbf{Y}_1^{*'}\mathbf{Y}_1^*$ converges to

$$\text{plim}\left(\frac{1}{T}\mathbf{y}_{p-r}^{(1j)'}\mathbf{y}_{p-w}^{(1j)}\right) = \begin{pmatrix} E\left[y_{t-p-r}^{(11)}y_{t-p-w}^{(11)}\right] & E\left[y_{t-p-r}^{(11)}y_{t-p-w}^{(12)}\right] & \cdots & E\left[y_{t-p-r}^{(11)}y_{t-p-w}^{(1m)}\right] \\ E\left[y_{t-p-r}^{(12)}y_{t-p-w}^{(11)}\right] & E\left[y_{t-p-r}^{(12)}y_{t-p-w}^{(12)}\right] & \cdots & E\left[y_{t-p-r}^{(12)}y_{t-p-w}^{(1m)}\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[y_{t-p-r}^{(1m)}y_{t-p-w}^{(11)}\right] & E\left[y_{t-p-r}^{(m1)}y_{t-p-w}^{(12)}\right] & \cdots & E\left[y_{t-p-r}^{(1m)}y_{t-p-w}^{(1m)}\right] \end{pmatrix}$$

$$= \begin{pmatrix} \gamma_{|r-w|}^{(11)} + (\mu_{11}^{(y)})^2 & \gamma_{|r-w|}^{(11,12)} + \mu_{11}^{(y)}\mu_{12}^{(y)} & \cdots & \gamma_{|r-w|}^{(11,1m)} + \mu_{11}^{(y)}\mu_{1m}^{(y)} \\ \gamma_{|r-w|}^{(12,11)} + \mu_{12}^{(y)}\mu_{11}^{(y)} & \gamma_{|r-w|}^{(12)} + (\mu_{12}^{(y)})^2 & \cdots & \gamma_{|r-w|}^{(12,1m)} + \mu_{12}^{(y)}\mu_{1m}^{(y)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{|r-w|}^{(1m,11)} + \mu_{1m}^{(y)}\mu_{11}^{(y)} & \gamma_{|r-w|}^{(1m,12)} + \mu_{1m}^{(y)}\mu_{12}^{(y)} & \cdots & \gamma_{|r-w|}^{(1m)} + (\mu_{1m}^{(y)})^2 \end{pmatrix}$$

Therefore, a typical scalar element of $\mathbf{Y}_1^{*'}\mathbf{Y}_1^*$ has a non-zero probability limit of the general form $\text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)}y_{t-p-w}^{(1j)}\right) = \gamma_{|r-w|}^{(1i,1j)} + \mu_{1i}^{(y)}\mu_{1j}^{(y)}$.

Now, the second block element of $\mathbf{Z}'\mathbf{Z}$ is given by

$$\mathbf{Y}_1^{*\prime}\mathbf{Y}_2^{*} = \begin{pmatrix} \mathbf{y}_{p-1}^{(1j)\prime} \\ \mathbf{y}_{p-2}^{(1j)\prime} \\ \vdots \\ \mathbf{y}_{p-a}^{(1j)\prime} \end{pmatrix} \begin{pmatrix} \mathbf{y}_{-1}^{(2j)} & \mathbf{y}_{-2}^{(2j)} & \cdots & \mathbf{y}_{-p-b}^{(2j)} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{-1}^{(2j)} & \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{-2}^{(2j)} & \cdots & \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{-p-b}^{(2j)} \\ \mathbf{y}_{p-2}^{(1j)\prime}\mathbf{y}_{-1}^{(2j)} & \mathbf{y}_{p-2}^{(1j)\prime}\mathbf{y}_{-2}^{(2j)} & \cdots & \mathbf{y}_{p-1}^{(1j)\prime}\mathbf{y}_{-p-b}^{(2j)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{p-a}^{(1j)\prime}\mathbf{y}_{-1}^{(2j)} & \mathbf{y}_{p-a}^{(1j)\prime}\mathbf{y}_{-2}^{(2j)} & \cdots & \mathbf{y}_{p-a}^{(1j)\prime}\mathbf{y}_{-p-b}^{(2j)} \end{pmatrix}$$

Similar to the above derivation, we can write the typical scalar element of $\mathbf{Y}_1^{*\prime}\mathbf{Y}_2^{*}$ as $\sum_{t=1}^{T} y_{t-p-r}^{(1i)} y_{t-w}^{(2j)}$, and therefore $\text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} y_{t-w}^{(2j)}\right) = \gamma_{|p-r-w|}^{(1i,2j)} + \mu_{1i}^{(y)}\mu_{2j}^{(y)}$.

Following the same principle and applying again Lemma (5.2.2.3), we can write typical elements of each of the blocks of the matrix $\mathbf{Z}'\mathbf{Z}$ as

$$\begin{array}{ll}
\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} y_{t-p-w}^{(1j)}\right) = \gamma_{|r-w|}^{(1i,1j)} + \mu_{1i}^{(y)}\mu_{1j}^{(y)} \\
\mathbf{Y}_1^{*\prime}\mathbf{Y}_2^{*} \,\&\mathbf{Y}_2^{*\prime}\mathbf{Y}_1^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} y_{t-w}^{(2j)}\right) = \gamma_{|p-r-w|}^{(1i,2j)} + \mu_{1i}^{(y)}\mu_{2j}^{(y)} \\
\mathbf{Y}_2^{*\prime}\mathbf{Y}_2^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)} y_{t-w}^{(2j)}\right) = \gamma_{|r-w|}^{(2i,2j)} + \mu_{2i}^{(y)}\mu_{2j}^{(y)} \\
\mathbf{Y}_1^{*\prime}\mathbf{X}_1^{*} \,\&\mathbf{X}_1^{*\prime}\mathbf{Y}_1^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} x_{t-q-w}^{(1j)}\right) = \psi_{|p-r-q-w|}^{(1i,1j)} + \mu_{1i}^{(y)}\mu_{1j}^{(x)} \\
\mathbf{Y}_2^{*\prime}\mathbf{X}_1^{*} \,\&\mathbf{X}_1^{*\prime}\mathbf{Y}_2^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)} x_{t-q-w}^{(1j)}\right) = \psi_{|r-q-w|}^{(2i,1j)} + \mu_{2i}^{(y)}\mu_{1j}^{(x)} \\
\mathbf{Y}_2^{*\prime}\mathbf{X}_2^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)} x_{t-w}^{(2j)}\right) = \psi_{|r-w|}^{(2i,2j)} + \mu_{2i}^{(y)}\mu_{2j}^{(x)} \\
\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-q-r}^{(1i)} x_{t-q-w}^{(1j)}\right) = \delta_{|r-w|}^{(1i,1j)} + \mu_{1i}^{(x)}\mu_{1j}^{(x)} \\
\mathbf{X}_1^{*\prime}\mathbf{X}_2^{*} \,\&\mathbf{X}_2^{*\prime}\mathbf{X}_1^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-q-r}^{(1i)} x_{t-w}^{(2j)}\right) = \delta_{|r-q-w|}^{(1i,2j)} + \mu_{1i}^{(x)}\mu_{2j}^{(x)} \\
\mathbf{X}_2^{*\prime}\mathbf{X}_2^{*}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-r}^{(2i)} x_{t-w}^{(2j)}\right) = \delta_{|r-w|}^{(2i,2j)} + \mu_{2i}^{(x)}\mu_{2j}^{(x)}
\end{array}$$

This proves (i) by showing that $\mathbf{\Sigma}_{ZZ} \neq \mathbf{0}$.

To show that (ii) holds, define $\mathbf{X} = (\boldsymbol{\iota}, \mathbf{Y}_{1j}, \mathbf{X}_{1j})$ where $\mathbf{Y}_{1j} = (\mathbf{Y}_{1jt}, \mathbf{Y}_{1jt-k})$ and $\mathbf{X}_{1j} = (\mathbf{X}_{1jt}, \mathbf{X}_{1jt-k})$. Let $\mathbf{Z}$ be defined as above. Therefore, the matrix $\mathbf{Z}'\mathbf{X}$ is given by

$$\mathbf{Z}'\mathbf{X} = \begin{pmatrix} \mathbf{Y}_1^{*\prime}\boldsymbol{\iota} & \mathbf{Y}_1^{*\prime}\mathbf{Y}_{1jt} & \mathbf{Y}_1^{*\prime}\mathbf{Y}_{1jt-k} & \mathbf{Y}_1^{*\prime}\mathbf{X}_{1jt} & \mathbf{Y}_1^{*\prime}\mathbf{X}_{1jt-k} \\ \mathbf{Y}_2^{*\prime}\boldsymbol{\iota} & \mathbf{Y}_2^{*\prime}\mathbf{Y}_{1jt} & \mathbf{Y}_2^{*\prime}\mathbf{Y}_{1jt-k} & \mathbf{Y}_2^{*\prime}\mathbf{X}_{1jt} & \mathbf{Y}_2^{*\prime}\mathbf{X}_{1jt-k} \\ \mathbf{X}_1^{*\prime}\boldsymbol{\iota} & \mathbf{X}_1^{*\prime}\mathbf{Y}_{1jt} & \mathbf{X}_1^{*\prime}\mathbf{Y}_{1jt-k} & \mathbf{X}_1^{*\prime}\mathbf{X}_{1jt} & \mathbf{X}_1^{*\prime}\mathbf{X}_{1jt-k} \\ \mathbf{X}_2^{*\prime}\boldsymbol{\iota} & \mathbf{X}_2^{*\prime}\mathbf{Y}_{1jt} & \mathbf{X}_2^{*\prime}\mathbf{Y}_{1jt-k} & \mathbf{X}_2^{*\prime}\mathbf{X}_{1jt} & \mathbf{X}_2^{*\prime}\mathbf{X}_{1jt-k} \end{pmatrix} \tag{8.46}$$

Using the same technique and applying Lemma (5.2.2.3) we derive typical elements of the blocks of specified in (8.46) and show that they have non-zero probability limits. Namely, we have

$$\begin{array}{ll}
\mathbf{Y}_1^{*\prime}\boldsymbol{\iota}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)}\right) = \mu_{1i}^{(y)} \\
\mathbf{Y}_2^{*\prime}\boldsymbol{\iota}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)}\right) = \mu_{2i}^{(y)} \\
\mathbf{X}_1^{*\prime}\boldsymbol{\iota}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-q-r}^{(1i)}\right) = \mu_{1i}^{(x)} \\
\mathbf{X}_2^{*\prime}\boldsymbol{\iota}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-r}^{(2i)}\right) = \mu_{2i}^{(x)} \\
\mathbf{Y}_1^{*\prime}\mathbf{Y}_{1jt}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} y_t^{(1j)}\right) = \gamma_{|p-r|}^{(1i,1j)} + \mu_{1i}^{(y)}\mu_{1j}^{(y)} \\
\mathbf{Y}_2^{*\prime}\mathbf{Y}_{1jt}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)} y_t^{(1j)}\right) = \gamma_r^{(2i,1j)} + \mu_{2i}^{(y)}\mu_{1j}^{(y)} \\
\mathbf{X}_1^{*\prime}\mathbf{Y}_{1jt}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-q-r}^{(1i)} y_t^{(1j)}\right) = \psi_{|q-r|}^{(1i,1j)} + \mu_{1i}^{(x)}\mu_{1j}^{(y)} \\
\mathbf{X}_2^{*\prime}\mathbf{Y}_{1jt}: & \text{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-r}^{(2i)} y_t^{(1j)}\right) = \psi_r^{(2i,1j)} + \mu_{2i}^{(x)}\mu_{1j}^{(y)}
\end{array}$$

$$
\begin{array}{ll}
\mathbf{Y}_1^{*\prime}\mathbf{Y}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} y_{t-w}^{(1j)}\right) = \gamma_{|p-r-w|}^{(1i,1j)} + \mu_{1i}^{(y)}\mu_{1j}^{(y)} \\
\mathbf{Y}_2^{*\prime}\mathbf{Y}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)} y_{t-w}^{(1j)}\right) = \gamma_{|r-w|}^{(2i,1j)} + \mu_{2i}^{(y)}\mu_{1j}^{(y)} \\
\mathbf{X}_1^{*\prime}\mathbf{Y}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-q-r}^{(1i)} y_{t-w}^{(1j)}\right) = \psi_{|q-r-w|}^{(1i,1j)} + \mu_{1i}^{(x)}\mu_{1j}^{(y)} \\
\mathbf{X}_2^{*\prime}\mathbf{Y}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-r}^{(2i)} y_{t-w}^{(1j)}\right) = \psi_{|r-w|}^{(2i,1j)} + \mu_{2i}^{(x)}\mu_{1j}^{(y)} \\
\mathbf{Y}_1^{*\prime}\mathbf{X}_{1jt}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} x_{t}^{(1j)}\right) = \psi_{|p-r|}^{(1i,1j)} + \mu_{1i}^{(y)}\mu_{1j}^{(x)} \\
\mathbf{Y}_2^{*\prime}\mathbf{X}_{1jt}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)} x_{t}^{(1j)}\right) = \psi_{r}^{(2i,1j)} + \mu_{2i}^{(y)}\mu_{1j}^{(x)} \\
\mathbf{X}_1^{*\prime}\mathbf{X}_{1jt}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-q-r}^{(1i)} x_{t}^{(1j)}\right) = \delta_{|q-r|}^{(1i,1j)} + \mu_{1i}^{(x)}\mu_{1j}^{(x)} \\
\mathbf{X}_2^{*\prime}\mathbf{X}_{1jt}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-r}^{(2i)} x_{t}^{(1j)}\right) = \delta_{r}^{(2i,1j)} + \mu_{2i}^{(x)}\mu_{1j}^{(x)} \\
\mathbf{Y}_1^{*\prime}\mathbf{X}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-p-r}^{(1i)} x_{t-w}^{(1j)}\right) = \psi_{|p-r-w|}^{(1i,1j)} + \mu_{1i}^{(y)}\mu_{1j}^{(y)} \\
\mathbf{Y}_2^{*\prime}\mathbf{X}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} y_{t-r}^{(2i)} x_{t-w}^{(1j)}\right) = \psi_{|r-w|}^{(2i,1j)} + \mu_{2i}^{(y)}\mu_{1j}^{(y)} \\
\mathbf{X}_1^{*\prime}\mathbf{X}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-q-r}^{(1i)} x_{t-w}^{(1j)}\right) = \delta_{|q-r-w|}^{(1i,1j)} + \mu_{1i}^{(x)}\mu_{1j}^{(x)} \\
\mathbf{X}_2^{*\prime}\mathbf{X}_{1jt-k}: & \operatorname{plim}\left(T^{-1}\sum_{t=1}^{T} x_{t-r}^{(2i)} x_{t-w}^{(1j)}\right) = \delta_{|r-w|}^{(2i,1j)} + \mu_{2i}^{(x)}\mu_{1j}^{(x)}
\end{array}
$$

This proves (ii) by showing that $\boldsymbol{\Sigma}_{ZX} \neq \mathbf{0}$, as required.

Finally, to prove (iii), let $\mathbf{u}_j = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, using the notation from (5.14)–(5.16). We require that $E[\mathbf{Z}'\mathbf{u}_1] = E[\mathbf{Z}'\mathbf{u}_2] = E[\mathbf{Z}'\mathbf{u}_3] = \mathbf{0}$. From (5.11), (5.12), and (5.13) this implies that

$$
\begin{aligned}
E[\mathbf{Z}'\mathbf{u}_1] &= E\left[\mathbf{Z}'\left(\boldsymbol{\zeta}_t + \boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j}\right)\right] \\
E[\mathbf{Z}'\mathbf{u}_2] &= E\left[\mathbf{Z}'\left(\boldsymbol{\varepsilon}_{2t} - \boldsymbol{\Lambda}_2^{(y)}\boldsymbol{\varepsilon}_{1t}\right)\right] \\
E[\mathbf{Z}'\mathbf{u}_3] &= E\left[\mathbf{Z}'\left(\boldsymbol{\delta}_{2t} - \boldsymbol{\Lambda}_2^{(x)}\boldsymbol{\delta}_{1t}\right)\right]
\end{aligned}
$$

Using the definition of $\mathbf{Z}$ we have

$$
\begin{aligned}
E[\mathbf{Z}'\mathbf{u}_1] &= E\left\{\begin{bmatrix}\mathbf{Y}_1^{*\prime}\\ \mathbf{Y}_2^{*\prime}\\ \mathbf{X}_1^{*\prime}\\ \mathbf{X}_2^{*\prime}\end{bmatrix}\left(\boldsymbol{\zeta}_t + \boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j}\right)\right\} \\
&= E\begin{bmatrix}
\mathbf{Y}_1^{*\prime}\boldsymbol{\zeta}_t + \mathbf{Y}_1^{*\prime}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{Y}_1^{*\prime}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{Y}_1^{*\prime}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j} \\
\mathbf{Y}_2^{*\prime}\boldsymbol{\zeta}_t + \mathbf{Y}_2^{*\prime}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{Y}_2^{*\prime}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{Y}_2^{*\prime}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j} \\
\mathbf{X}_1^{*\prime}\boldsymbol{\zeta}_t + \mathbf{X}_1^{*\prime}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{X}_1^{*\prime}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{X}_1^{*\prime}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j} \\
\mathbf{X}_2^{*\prime}\boldsymbol{\zeta}_t + \mathbf{X}_2^{*\prime}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{X}_2^{*\prime}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{X}_2^{*\prime}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j}
\end{bmatrix} \\
&= E\begin{bmatrix}
\mathbf{Y}'_{1t-p-r}\boldsymbol{\zeta}_t + \mathbf{Y}'_{1t-p-r}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{Y}'_{1t-p-r}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{Y}'_{1t-p-r}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j} \\
\mathbf{Y}'_{2t-r}\boldsymbol{\zeta}_t + \mathbf{Y}'_{2t-r}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{Y}'_{2t-r}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{Y}'_{2t-r}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j} \\
\mathbf{X}'_{1t-q-r}\boldsymbol{\zeta}_t + \mathbf{X}'_{1t-q-r}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{X}'_{1t-q-r}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{X}'_{1t-q-r}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j} \\
\mathbf{X}'_{2t-r}\boldsymbol{\zeta}_t + \mathbf{X}'_{2t-r}\boldsymbol{\varepsilon}_{1t} - \sum_{j=0}^{p}\mathbf{X}'_{2t-r}\mathbf{B}_j\boldsymbol{\varepsilon}_{1t-j} - \sum_{j=0}^{q}\mathbf{X}'_{2t-r}\boldsymbol{\Gamma}_j\boldsymbol{\delta}_{1t-j}
\end{bmatrix} \\
&= \mathbf{0}.
\end{aligned}
$$

Similarly, it easily seen that $E[\mathbf{Z}'\mathbf{u}_2] = E[\mathbf{Z}'\mathbf{u}_3] = \mathbf{0}$, as the lagged variables are uncorrelated with the contemporaneous latent errors.

Q.E.D.

Appendix §5C

Proof of Proposition 5.4.0.1 We show that a general DSEM model can be written in the form $\mathbf{VK} + \mathbf{WG} = \mathbf{u}$. Then the result follows since rank $[\mathbf{R}_j (\mathbf{K}' : \mathbf{G}')'] \geq M - 1$,[1] where $\mathbf{M}$ is the number of the left-hand side variables, i.e., number of the equations in the system and where $\mathbf{R}_j$ is the selection matrix for the j^{th} equation, therefore $M = m + n + h$.

Note that the DSEM model (5.14)–(5.16) can be written as

$$\begin{aligned} \mathbf{y}_{1t} &= \boldsymbol{\alpha}_\eta + \mathbf{B}_0\mathbf{y}_{1t} + \mathbf{B}_1\mathbf{y}_{1t-1} + \ldots + \mathbf{B}_p\mathbf{y}_{1t-p} + \boldsymbol{\Gamma}_0\mathbf{x}_{1t} \\ &\qquad + \boldsymbol{\Gamma}_1\mathbf{x}_{1t-1} + \ldots + \boldsymbol{\Gamma}_q\mathbf{x}_{1t-q} + \mathbf{u}_{1t} \\ \mathbf{y}_{2t} &= \boldsymbol{\alpha}_2^{(y)} + \boldsymbol{\Lambda}_2^{(y)}\mathbf{y}_{1t} + \mathbf{u}_{2t} \\ \mathbf{x}_{2t} &= \boldsymbol{\alpha}_2^{(x)} + \boldsymbol{\Lambda}_2^{(x)}\mathbf{x}_{1t} + \mathbf{u}_{3t} \end{aligned} \tag{8.47}$$

By rearranging by moving the endogenous variables on the left hand side (8.47) can be written as

$$\begin{aligned} &\begin{pmatrix} (\mathbf{I} - \mathbf{B}_0) & \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\Lambda}_2^{(y)} & \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_h \end{pmatrix} \begin{pmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \\ \mathbf{x}_{2t} \end{pmatrix} \\ &\quad - \begin{pmatrix} -\boldsymbol{\alpha}_\eta & -\mathbf{B}_1 & -\mathbf{B}_2 & \cdots & -\mathbf{B}_p & -\boldsymbol{\Gamma}_0 & -\boldsymbol{\Gamma}_1 & \cdots & -\boldsymbol{\Gamma}_q \\ -\boldsymbol{\alpha}_2^{(y)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\boldsymbol{\alpha}_2^{(x)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Lambda}_2^{(x)} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \\ &\times \begin{pmatrix} \boldsymbol{\iota} & \mathbf{y}'_{1t-1} & \mathbf{y}'_{1t-2} & \cdots & \mathbf{y}'_{1t-p} & \mathbf{x}'_{1t} & \mathbf{x}'_{1t-1} & \cdots & \mathbf{x}'_{1t-q} \end{pmatrix}' = \begin{pmatrix} \mathbf{u}_{1t} \\ \mathbf{u}_{2t} \\ \mathbf{u}_{3t} \end{pmatrix} \end{aligned}$$

Transposing the entire system we get

$$\begin{aligned} &(\mathbf{y}_{1t} : \mathbf{y}_{2t} : \mathbf{x}_{2t}) \begin{pmatrix} (\mathbf{I} - \mathbf{B}'_0) & -\boldsymbol{\Lambda}_2^{(y)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_h \end{pmatrix} + \\ &(\boldsymbol{\iota}, \mathbf{y}_{1t-1}, \mathbf{y}_{1t-2}, \ldots, \mathbf{y}_{1t-p}, \mathbf{x}_{1t}, \mathbf{x}_{1t-1}, \ldots, \mathbf{x}_{1t-q}) \begin{pmatrix} -\boldsymbol{\alpha}_\eta & -\boldsymbol{\alpha}_2^{(y)} & -\boldsymbol{\alpha}_2^{(x)} \\ -\mathbf{B}'_1 & \mathbf{0} & \mathbf{0} \\ -\mathbf{B}'_2 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\mathbf{B}'_p & \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\Gamma}'_0 & \mathbf{0} & -\boldsymbol{\Lambda}_2^{(x)'} \\ -\boldsymbol{\Gamma}'_1 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\boldsymbol{\Gamma}'_q & \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= (\mathbf{u}_{1t} : \mathbf{u}_{2t} : \mathbf{u}_{3t}), \end{aligned} \tag{8.48}$$

which can be written in full-sample notation

[1] See e.g. Judge *et al.* (1985)

$$
(\mathbf{y}_{1j},\ \mathbf{y}_{2j},\ \mathbf{x}_{2j}) \begin{pmatrix} (\mathbf{I}-\mathbf{B}'_0) & -\boldsymbol{\Lambda}_2^{(y)\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_h \end{pmatrix}
- (\boldsymbol{\iota},\ \mathbf{Y}_{1jt-k},\ \mathbf{X}_{1jt},\ \mathbf{X}_{1jt-k}) \begin{pmatrix} -\boldsymbol{\alpha}_\eta & -\boldsymbol{\alpha}_2^{(y)} & -\boldsymbol{\alpha}_2^{(x)} \\ -\mathbf{B}'_1 & \mathbf{0} & \mathbf{0} \\ -\mathbf{B}'_2 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\mathbf{B}'_p & \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\Gamma}'_0 & \mathbf{0} & -\boldsymbol{\Lambda}_2^{(x)\prime} \\ -\boldsymbol{\Gamma}'_1 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\boldsymbol{\Gamma}'_q & \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{u},
$$

where $\mathbf{u} \equiv (\mathbf{u}_{1t},\ \mathbf{u}_{2t},\ \mathbf{u}_{3t})$. Now define $\mathbf{V} \equiv (\mathbf{y}_{1j},\ \mathbf{y}_{2j},\ \mathbf{x}_{2j})$ and $\mathbf{W} \equiv (\boldsymbol{\iota},\ \mathbf{Y}_{1jt},\ \mathbf{X}_{1jt},\ \mathbf{X}_{1jt-k})$,

$$
\mathbf{K} \equiv \begin{pmatrix} (\mathbf{I}-\mathbf{B}'_0) & -\boldsymbol{\Lambda}_2^{(y)\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_h \end{pmatrix}, \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} \boldsymbol{\alpha}_\eta & \boldsymbol{\alpha}_2^{(y)} & \boldsymbol{\alpha}_2^{(x)} \\ -\mathbf{B}'_1 & \mathbf{0} & \mathbf{0} \\ -\mathbf{B}'_2 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\mathbf{B}'_p & \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\Gamma}'_0 & \mathbf{0} & -\boldsymbol{\Lambda}_2^{(x)\prime} \\ -\boldsymbol{\Gamma}'_1 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ -\boldsymbol{\Gamma}'_q & \mathbf{0} & \mathbf{0} \end{pmatrix},
$$

therefore the system (8.48) can be written as $\mathbf{VK} + \mathbf{WG} = \mathbf{u}$, as required.

Q.E.D.

8.4 Chapter §6 appendices: Estimation code

Appendix §6A: S matrix functions

Lag operator

Note that for pure time series data ($N = 1$) the `power.shift()` function can be used in its own right for computing lags. For example,

```
>    x_1 <- power.shift(120,1)%*%x
```

computes the first lag of variable x, thus having the effect of a simple lag operator. The p-period lag for data in the "long panel" format can be computed using the vectorised S+ function `lag.panel()` defined as

```
"lag.panel" <- function(x,p,n,t)
{
    lag <- power.shift(n*t,p)%*%x
    eliminator <- kronecker(rep(1,n), c(rep(NA,p), rep(1,t-p)))
    lag*eliminator
}
```

where p denotes the lag, n is the cross-section sample size and t is the time-dimension of the panel and x which needs to be lagged. Thus, if x is a vector of $N = 200$ individuals observed over $T = 10$ time periods, stacked in the "long panel" format, the first lag can be computed by executing the command

```
>    x_1 <- lag.panel(x,1,200,10)
```

For very large panels, the vectorised function `panel.lag()` might require large RAM memory and tends to be very slow if virtual memory is used via paging files. Therefore, with large panels it might be necessary to use a looping version `lag.lp()` defined as

```
"lagp.lp" <- function(x, p, n, t)
{
    m <- n-1
    v <- lag.df(x,p)

    for(i in 0:m)
        {
                  for(j in 1:p)
            {
              v[t * i + j] <- 0
                }
            }
    v
}
```

which makes use of the function `lag.df()` that returns the lagged vector of the same length as the original vector,

```
"lag.df" <- function(x, p)
{
    g <- length(x)
    v <- x
    for(j in 1:p)
        {
        v[j] <- 0
        }
    v
       q <- p+1
       for(i in q:g)
       {
         v[i] <- x[i-p]
       }
    v
}
```

Instrumental variable operator S_{IV}^j

We define an S function that has the effect of applying the operator S_{IV}^j, defined in (5.20), on the variable x, where j denotes the lag, n is the cross-section sample size and t is the time-dimension of the panel,

```
"instrument" <- function(x, j, n, t)
{
    d1 <- power.shift(t, j) + rbind(cbind(0, diag(j),
    + matrix(0, j, t - j - 1)), matrix(0, t - j, t))
    d2 <- kronecker(diag(n), d1)
    d2 % * % x
}
```

The instrument() function makes use of the power.shift() function defined above. As an illustration, suppose we wish to create a 2-period lagged instrument for the variable "income" and name it as "IVincome2" in a panel with $N = 200$, $T = 10$. This can be achieved by writing

```
>   IVincome2 <- instrument(income,2,200,10)
```

from the command line in S-PLUS. The above function uses matrix operators, namely shifting matrices and is hence a fully vectorised function. For very large panels, it might be necessary to use loops and apply the following S+ function

```
"instrument.lp" <- function(x, p, n, t)
{
      m <- n-1
      v <- lag.df(x,p)
        for(i in 0:m)
         {
                  for(j in 1:p)
                {
                 v[t*i+j] <- 0
                }
         }
    v
}
```

Note that the functions `instrument()` and `instrument.lp()` have the same arguments and will return the same result.

Appendix §6B: Estimation code

GiveWin batch syntax for FD model

The GIVE estimation code for the two structural equations of the FD model (in GiveWin batch syntax), which can be run with GiveWin 1.30 or later. The equation for g_t using the IV_2 set can be estimated with

```
module("PcGive");
usedata("FDpanel.xls");
system
{
    Y = g(t), l(t);
    Z = Constant, g(t-1), i(t);
    A = IVb_2, IVl_2, IVp_2;
}
estsystem("IVE", 1, 1, 180, 1);
```

while the code for the second structural equation (l_t) using IV_2 that includes lags of g_t and i_t is given by

```
module("PcGive");
usedata("FDpanel.xls");
system
{
    Y = l(t), g(t-1), i(t);
    Z = Constant;
    A = IVb_2, IVl_2, IVp_2, IVg_2, IVi_2;
}
estsystem("IVE", 1, 1, 180, 1);
```

The measurement model (equations for b_t and p_t), using l_t as the unit-loading proxy for latent financial development, can be estimated with the following code

```
module("PcGive");
usedata("FDpanel.xls");
system
{
    Y = b(t), l(t);
    Z = Constant;
    A = IVb_3, IVl_3, IVp_3;
}
estsystem("IVE", 1, 1, 180, 1);
```

where IV_3 is used as the set of instruments. The second measurement equation uses IV_2 set of instruments and the code is

```
module("PcGive");
usedata("FDpanel.xls");
system
```

```
{
    Y = p(t), l(t);
    Z = Constant;
    A = IVb_2, IVl_2, IVp_2;
}
estsystem("IVE", 1, 1, 180, 1);
```

The FIVE (3SLS) estimation code that estimates the structural model jointly with the measurement model is given by

```
module("PcFiml");
usedata("FDpanel.xls");
system
{
    Y = g(t), b(t), l(t), p(t);
    Z = i(t), g(t-1), Constant,
        IVb_2, IVl_2, IVp_2, IVg_2, IVi_2,
        IVb_3, IVl_3, IVp_3;
}
estsystem("OLS", 1, 1, 180, 1);
model
{
    g(t) =  l(t), i(t), g(t-1);
    l(t) =  i(t), g(t-1), Constant;
    b(t) =  l(t), Constant;
    p(t) =  l(t), Constant;
}
estmodel("3SLS", 0);
```

where IV_2 and IV_3 are used as instrument sets.

LISREL estimation syntax for the FD model

The following LISREL syntax uses the covariance matrix saved in the file wgpanel.CM and does not specify user-defined starting values,

```
TI
DA NI=40 NO=20 NG=1 MA=CM
CM=wgpanel.CM
SE
6 7 8 9 10 16 17 18 19 20 21 22 23 24 25 26 27 28
29 30 36 37 38 39 40 /
MO NX=5 NY=20 NK=5 NE=10 LY=FU,FI LX=FU,FI
MO BE=FU,FI GA=FU,FI PH=SY,FR PS=DI,FR TE=DI,FR TD=DI,FI
FR LY(6,1) LY(7,2) LY(8,3) LY(9,4) LY(10,5)
EQ LY(6,1) LY(7,2) LY(8,3) LY(9,4) LY(10,5)
FR LY(11,1) LY(12,2) LY(13,3) LY(14,4) LY(15,5)
EQ LY(11,1) LY(12,2) LY(13,3) LY(14,4) LY(15,5)
VA 1.00 LY(16,1) LY(17,2) LY(18,3) LY(19,4) LY(20,5)
VA 1.00 LY(1,6) LY(2,7) LY(3,8) LY(4,9) LY(5,10)
VA 1.00 LX(1,1) LX(2,2) LX(3,3) LX(4,4) LX(5,5)
FR BE(6,1) BE(7,2) BE(8,3) BE(9,4) BE(10,5)
```

```
EQ BE(6,1) BE(7,2) BE(8,3) BE(9,4) BE(10,5)
FR BE(2,6) BE(3,7) BE(4,8) BE(5,9)
EQ BE(2,6) BE(3,7) BE(4,8) BE(5,9)
FR BE(7,6) BE(8,7) BE(9,8) BE(10,9)
EQ BE(7,6) BE(8,7) BE(9,8) BE(10,9)
FR GA(1,1) GA(2,2) GA(3,3) GA(4,4) GA(5,5)
EQ GA(1,1) GA(2,2) GA(3,3) GA(4,4) GA(5,5)
FR GA(6,1) GA(7,2) GA(8,3) GA(9,4) GA(10,5)
EQ GA(6,1) GA(7,2) GA(8,3) GA(9,4) GA(10,5)
FI PH(5,1)
FI PH(5,2) PH(4,1)
EQ PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
EQ PH(2,1) PH(3,2) PH(4,3) PH(5,4)
EQ PH(3,1) PH(4,2) PH(5,3)
EQ PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(5,5)
EQ PS(6,6) PS(7,7) PS(8,8) PS(9,9) PS(10,10)
FI TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5)
EQ TE(6,6) TE(7,7) TE(8,8) TE(9,9) TE(10,10)
EQ TE(11,11) TE(12,12) TE(13,13) TE(14,14) TE(15,15)
EQ TE(16,16) TE(17,17) TE(18,18) TE(19,19) TE(20,20)
OU ME=ML ND=5
```

Starting values based on the FIVE (3SLS) estimates are given below. The ST command lines should be include anywhere before the OU command in the above LISREL syntax.

```
ST  0.1818  LY(6,1)   LY(7,2)   LY(8,3)   LY(9,4)   LY(10,5)
ST  0.5246  LY(11,1)  LY(12,2)  LY(13,3)  LY(14,4)  LY(15,5)
ST -0.0089  BE(6,1)   BE(7,2)   BE(8,3)   BE(9,4)   BE(10,5)
ST  0.4001  BE(2,6)   BE(3,7)   BE(4,8)   BE(5,9)
ST  0.0000  BE(7,6)   BE(8,7)   BE(9,8)   BE(10,9)
ST -0.8658  GA(1,1)   GA(2,2)   GA(3,3)   GA(4,4)   GA(5,5)
ST -0.0560  GA(6,1)   GA(7,2)   GA(8,3)   GA(9,4)   GA(10,5)
ST  0.0000  PH(1,1)   PH(2,2)   PH(3,3)   PH(4,4)   PH(5,5)
ST  0.0000  PH(2,1)   PH(3,2)   PH(4,3)   PH(5,4)
ST  0.0000  PH(3,1)   PH(4,2)   PH(5,3)
ST  0.0189  PS(1,1)   PS(2,2)   PS(3,3)   PS(4,4)   PS(5,5)
ST  0.0004  PS(6,6)   PS(7,7)   PS(8,8)   PS(9,9)   PS(10,10)
ST  0.0000  TE(6,6)   TE(7,7)   TE(8,8)   TE(9,9)   TE(10,10)
ST  0.0000  TE(11,11) TE(12,12) TE(13,13) TE(14,14) TE(15,15)
ST  0.0000  TE(16,16) TE(17,17) TE(18,18) TE(19,19) TE(20,20)
```

IV GiveWin batch code for the BHPS model

Measurement equations:

```
module("PcGive");
usedata("bhps.long");
system
{
    Y = h, f;
    A = fiv4,fiv5;
}
```

```
estsystem("IVE", 1, 1, 66976, 1);

module("PcGive");
usedata("bhps.long");
system
{
    Y = i, l;
    Z = Constant;
    A = liv5,liv6;
}
estsystem("IVE", 1, 1, 66976, 1);

module("PcGive");
usedata("bhps.long");
system
{
    Y = r, s;
    Z = Constant;
    A = siv5,siv6;
}
estsystem("IVE", 1, 1, 66976, 1);


\noindent Structural equations:

module("PcGive");
usedata("bhps.long");
system
{
    Y = f,f1,f2,f4,l,l1,l2,s,s1,s2;
    Z = Constant;
    A = fiv5,fiv6,fiv7,liv4,liv5,liv6,liv7,siv4,siv5,siv6,siv7;
}
estsystem("IVE", 1, 1, 66976, 1);

module("PcGive");
usedata("bhps.long");
system
{
    Y = s,s1,s2,s3,s4,f1,f2,l,l1,l2,l3;
    Z = Constant;
    A = fiv5,fiv6,fiv7,liv4,liv5,liv6,liv7,siv4,siv5,siv6,siv7;
}
estsystem("IVE", 1, 1, 66976, 1);

module("PcGive");
usedata("bhps.long");
system
{
    Y = l,l1,l2,l3,l4,l5;
```

```
    Z = Constant,e;
    A = fiv5,fiv6,fiv7,fiv8,iiv6,iiv7,iiv8;
}
estsystem("IVE", 1, 1, 66976, 1);
```

FIML code:

```
module("PcFiml");
usedata("bhps.long");
system
{
    Y = h,i,r;
    Z = Constant,f,l,s,fiv4,fiv5,liv5,fiv6,siv5,siv6;
}
estsystem("OLS", 1, 1, 66976, 1);
model
{
    h = f,Constant;
    i = l,Constant;
    r = s,Constant;
}
estmodel("3SLS", 0);

module("PcFiml");
usedata("bhps.long");
system
{
    Y = f,s,l;
    Z = Constant,f1,f2,f4,l1,l2,l3,l4,l5,s1,s2,s3,s4,e,
                          fiv5,fiv6,fiv7,fiv8,
                     liv4,liv5,liv6,liv7,
                     siv4,siv5,siv6,siv7,
                               iiv6,iiv7,iiv8;
}
estsystem("OLS", 1, 1, 66976, 1);
model
{
    f = f1,f2,f4,l,l1,l2,s,s1,s2;
    s = s1,s2,s3,s4,f1,f2,l,l1,l2,l3,Constant;
    l = l1,l2,l3,l4,l5,e,Constant;
}
estmodel("3SLS", 0);

module("PcFiml");
usedata("bhps.long");
system
{
    Y = h,i,r,f,s,l;
    Z = Constant,e,f1,f2,f4,l1,l2,l3,l4,l5,s1,s2,s3,s4,
                   fiv4,fiv5,fiv6,fiv7,fiv8,
                   liv4,liv5,liv6,liv7,
```

```
                  siv4,siv5,siv6,siv7,
                            iiv6,iiv7,iiv8;
}
estsystem("OLS", 1, 1, 66976, 1);
model
{
   h = f;
   i = l;
   r = s;
   f = f1,f2,f4,l,l1,l2,s,s1,s2;
   s = s1,s2,s3,s4,f1,f2,l,l1,l2,l3;
   l = l1,l2,l3,l4,l5,e;
}
estmodel("3SLS", 0);
```

LISREL syntax for the BHPS model

```
TI
DA NI=104 NO=3347 NG=1 MA=CM
KM=bhps.KC
LA
AFOOD AHOUSING ALABOUR ANLABOUR AINVEST ASAVINGS AED ACREDIT
BFOOD BHOUSING BLABOUR BNLABOUR BINVEST BSAVINGS BED BCREDIT
CFOOD CHOUSING CLABOUR CNLABOUR CINVEST CSAVINGS CED CCREDIT
DFOOD DHOUSING DLABOUR DNLABOUR DINVEST DSAVINGS DED DCREDIT
EFOOD EHOUSING ELABOUR ENLABOUR EINVEST ESAVINGS EED ECREDIT
FFOOD FHOUSING FLABOUR FNLABOUR FINVEST FSAVINGS FED FCREDIT
GFOOD GHOUSING GLABOUR GNLABOUR GINVEST GSAVINGS GED GCREDIT
HFOOD HHOUSING HLABOUR HNLABOUR HINVEST HSAVINGS HED HCREDIT
IFOOD IHOUSING ILABOUR INLABOUR IINVEST ISAVINGS IED ICREDIT
JFOOD JHOUSING JLABOUR JNLABOUR JINVEST JSAVINGS JED JCREDIT
KFOOD KHOUSING KLABOUR KNLABOUR KINVEST KSAVINGS KED KCREDIT
LFOOD LHOUSING LLABOUR LNLABOUR LINVEST LSAVINGS LED LCREDIT
MFOOD MHOUSING MLABOUR MNLABOUR MINVEST MSAVINGS MED MCREDIT
SE
1 9 17 25 33 41 49 57 65 73 81 89 97
2 10 18 26 34 42 50 58 66 74 82 90 98
3 11 19 27 35 43 51 59 67 75 83 91 99
4 12 20 28 36 44 52 60 68 76 84 92 100
5 13 21 29 37 45 53 61 69 77 85 93 101
6 14 22 30 38 46 54 62 70 78 86 94 102
8 16 24 32 40 48 56 64 72 80 88 96 104
7 15 23 31 39 47 55 63 71 79 87 95 103/
MO NX=13 NY=78 NK=13 NE=39 LY=FU,FI LX=FU,FI BE=FU,FI GA=FU,FI PH=SY,FI PS=DI,FR
MO TE=SY,FI TD=DI,FR
LE
aCons bCons cCons dCons eCons fCons gCons hCons iCons jCons kCons lCons mCons
aIncome bIncome cIncome dIncome eIncome fIncome gIncome hIncome iIncome jIncome
kIncome lIncome mIncome
aLQ bLQ cLQ dLQ eLQ fLQ gLQ hLQ iLQ jLQ kLQ lLQ mLQ
LK
```

```
aEd bEd cEd dEd eEd fEd gEd hEd iEd jEd kEd lEd mEd
VA 1 LY(1,1) LY(2,2) LY(3,3) LY(4,4) LY(5,5) LY(6,6) LY(7,7) LY(8,8) LY(9,9)
VA 1 LY(10,10) LY(11,11) LY(12,12) LY(13,13)
FR LY(14,1) LY(15,2) LY(16,3) LY(17,4) LY(18,5) LY(19,6) LY(20,7) LY(21,8)
FR LY(22,9) LY(23,10) LY(24,11) LY(25,12) LY(26,13)
EQ LY(14,1) LY(15,2) LY(16,3) LY(17,4) LY(18,5) LY(19,6) LY(20,7) LY(21,8)
EQ LY(21,8) LY(22,9) LY(23,10) LY(24,11) LY(25,12) LY(26,13)
VA 1 LY(27,14) LY(28,15) LY(29,16) LY(30,17) LY(31,18) LY(32,19) LY(33,20)
VA 1 LY(34,21) LY(35,22) LY(36,23) LY(37,24)  LY(38,25) LY(39,26)
FR LY(40,14) LY(41,15) LY(42,16) LY(43,17) LY(44,18) LY(45,19) LY(46,20)
FR LY(47,21) LY(48,22) LY(49,23) LY(50,24) LY(51,25) LY(52,26)
EQ LY(40,14) LY(41,15) LY(42,16) LY(43,17) LY(44,18) LY(45,19) LY(46,20)
EQ LY(46,20) LY(47,21) LY(48,22) LY(49,23) LY(50,24) LY(51,25) LY(52,26)
VA 1 LY(53,27) LY(54,28) LY(55,29) LY(56,30) LY(57,31) LY(58,32) LY(59,33)
VA 1 LY(60,34) LY(61,35) LY(62,36) LY(63,37) LY(64,38) LY(65,39)
FR LY(66,27) LY(67,28) LY(68,29) LY(69,30) LY(70,31) LY(71,32) LY(72,33)
FR LY(73,34) LY(74,35) LY(75,36) LY(76,37) LY(77,38) LY(78,39)
EQ LY(66,27) LY(67,28) LY(68,29) LY(69,30) LY(70,31) LY(71,32) LY(72,33)
EQ LY(72,33) LY(73,34) LY(74,35) LY(75,36) LY(76,37) LY(77,38) LY(78,39)
FR BE(2,1) BE(3,2) BE(4,3) BE(5,4) BE(6,5) BE(7,6) BE(8,7) BE(9,8) BE(10,9)
FR BE(11,10) BE(12,11) BE(13,12)
EQ BE(2,1) BE(3,2) BE(4,3) BE(5,4) BE(6,5) BE(7,6) BE(8,7) BE(9,8) BE(10,9)
EQ BE(10,9) BE(11,10) BE(12,11) BE(13,12)
FR BE(3,1) BE(4,2) BE(5,3) BE(6,4) BE(7,5) BE(8,6) BE(9,7) BE(10,8) BE(11,9)
FR BE(12,10) BE(13,11)
EQ BE(3,1) BE(4,2) BE(5,3) BE(6,4) BE(7,5) BE(8,6) BE(9,7) BE(10,8) BE(11,9)
EQ BE(11,9) BE(12,10) BE(13,11)
FR BE(4,1) BE(5,2) BE(6,3) BE(7,4) BE(8,5) BE(9,6) BE(10,7) BE(11,8) BE(12,9)
FR BE(13,10)
EQ BE(4,1) BE(5,2) BE(6,3) BE(7,4) BE(8,5) BE(9,6) BE(10,7) BE(11,8) BE(12,9)
EQ BE(12,9) BE(13,10)
FR BE(5,1) BE(6,2) BE(7,3) BE(8,4) BE(9,5) BE(10,6) BE(11,7) BE(12,8) BE(13,9)
EQ BE(5,1) BE(6,2) BE(7,3) BE(8,4) BE(9,5) BE(10,6) BE(11,7) BE(12,8) BE(13,9)
FR BE(6,1) BE(7,2) BE(8,3) BE(9,4) BE(10,5) BE(11,6) BE(12,7) BE(13,8)
EQ BE(6,1) BE(7,2) BE(8,3) BE(9,4) BE(10,5) BE(11,6) BE(12,7) BE(13,8)
FR BE(1,14) BE(2,15)  BE(3,16)  BE(4,17) BE(5,18) BE(6,19) BE(7,20) BE(8,21)
FR BE(9,22) BE(10,23) BE(11,24) BE(12,25) BE(13,26)
EQ BE(1,14) BE(2,15)  BE(3,16)  BE(4,17) BE(5,18) BE(6,19) BE(7,20) BE(8,21)
EQ BE(8,21) BE(9,22) BE(10,23) BE(11,24) BE(12,25) BE(13,26)
FR BE(2,14) BE(3,15)  BE(4,16) BE(5,17) BE(6,18) BE(7,19) BE(8,20) BE(9,21)
FR BE(10,22) BE(11,23) BE(12,24) BE(13,25)
EQ BE(2,14) BE(3,15)  BE(4,16) BE(5,17) BE(6,18) BE(7,19) BE(8,20) BE(9,21)
EQ BE(9,21) BE(10,22) BE(11,23) BE(12,24) BE(13,25)
FR BE(3,14) BE(4,15)  BE(5,16) BE(6,17) BE(7,18) BE(8,19) BE(9,20) BE(10,21)
FR BE(11,22) BE(12,23) BE(13,24)
EQ BE(3,14) BE(4,15)  BE(5,16) BE(6,17) BE(7,18) BE(8,19) BE(9,20) BE(10,21)
EQ BE(10,21) BE(11,22) BE(12,23) BE(13,24)
FR BE(1,27) BE(2,28) BE(3,29) BE(4,30) BE(5,31) BE(6,32) BE(7,33) BE(8,34)
FR BE(9,35) BE(10,36) BE(11,37) BE(12,38) BE(13,39)
EQ BE(1,27) BE(2,28) BE(3,29) BE(4,30) BE(5,31) BE(6,32) BE(7,33) BE(8,34)
EQ BE(8,34) BE(9,35) BE(10,36) BE(11,37) BE(12,38) BE(13,39)
```

```
FR BE(2,27) BE(3,28) BE(4,29) BE(5,30) BE(6,31) BE(7,32) BE(8,33) BE(9,34)
FR BE(10,35) BE(11,36) BE(12,37) BE(13,38)
EQ BE(2,27) BE(3,28) BE(4,29) BE(5,30) BE(6,31) BE(7,32) BE(8,33) BE(9,34)
EQ BE(9,34) BE(10,35) BE(11,36) BE(12,37) BE(13,38)
FR BE(3,27) BE(4,28) BE(5,29) BE(6,30) BE(7,31) BE(8,32) BE(9,33) BE(10,34)
FR BE(11,35) BE(12,36) BE(13,37)
EQ BE(3,27) BE(4,28) BE(5,29) BE(6,30) BE(7,31) BE(8,32) BE(9,33) BE(10,34)
EQ BE(10,34) BE(11,35) BE(12,36) BE(13,37)
FR BE(4,27) BE(5,28) BE(6,29) BE(7,30) BE(8,31) BE(9,32) BE(10,33) BE(11,34)
FR BE(12,35) BE(13,36)
EQ BE(4,27) BE(5,28) BE(6,29) BE(7,30) BE(8,31) BE(9,32) BE(10,33) BE(11,34)
EQ BE(11,34) BE(12,35) BE(13,36)
FR BE(15,14) BE(16,15) BE(17,16) BE(18,17) BE(19,18) BE(20,19) BE(21,20)
FR BE(22,21) BE(23,22) BE(24,23) BE(25,24) BE(26,25)
EQ BE(15,14) BE(16,15) BE(17,16) BE(18,17) BE(19,18) BE(20,19) BE(21,20)
EQ BE(21,20) BE(22,21) BE(23,22) BE(24,23) BE(25,24) BE(26,25)
FR BE(16,14) BE(17,15) BE(18,16) BE(19,17) BE(20,18) BE(21,19) BE(22,20)
FR BE(23,21) BE(24,22) BE(25,23) BE(26,24)
EQ BE(16,14) BE(17,15) BE(18,16) BE(19,17) BE(20,18) BE(21,19) BE(22,20)
EQ BE(22,20) BE(23,21) BE(24,22) BE(25,23) BE(26,24)
FR BE(17,14) BE(18,15) BE(19,16) BE(20,17) BE(21,18) BE(22,19) BE(23,20)
FR BE(24,21) BE(25,22) BE(26,23)
EQ BE(17,14) BE(18,15) BE(19,16) BE(20,17) BE(21,18) BE(22,19) BE(23,20)
EQ BE(23,20) BE(24,21) BE(25,22) BE(26,23)
FR BE(18,14) BE(19,15) BE(20,16) BE(21,17) BE(22,18) BE(23,19) BE(24,20)
FR BE(25,21) BE(26,22)
EQ BE(18,14) BE(19,15) BE(20,16) BE(21,17) BE(22,18) BE(23,19) BE(24,20)
EQ BE(24,20) BE(25,21) BE(26,22)
FR BE(19,14) BE(20,15) BE(21,16) BE(22,17) BE(23,18) BE(24,19) BE(25,20)
FR BE(26,21)
EQ BE(19,14) BE(20,15) BE(21,16) BE(22,17) BE(23,18) BE(24,19) BE(25,20)
EQ BE(25,20) BE(26,21)
FR GA(14,1) GA(15,2) GA(16,3) GA(17,4) GA(18,5) GA(19,6) GA(20,7) GA(21,8)
FR GA(22,9) GA(23,10) GA(24,11) GA(25,12) GA(26,13)
EQ GA(14,1) GA(15,2) GA(16,3) GA(17,4) GA(18,5) GA(19,6) GA(20,7) GA(21,8)
EQ GA(21,8) GA(22,9) GA(23,10) GA(24,11) GA(25,12) GA(26,13)
FR BE(28,27) BE(29,28) BE(30,29) BE(31,30) BE(32,31) BE(33,32) BE(34,33)
FR BE(35,34) BE(36,35) BE(37,36) BE(38,37) BE(39,38)
EQ BE(28,27) BE(29,28) BE(30,29) BE(31,30) BE(32,31) BE(33,32) BE(34,33)
EQ BE(34,33) BE(35,34) BE(36,35) BE(37,36) BE(38,37) BE(39,38)
FR BE(29,27) BE(30,28) BE(31,29) BE(32,30) BE(33,31) BE(34,32) BE(35,33)
FR BE(36,34) BE(37,35) BE(38,36) BE(39,37)
EQ BE(29,27) BE(30,28) BE(31,29) BE(32,30) BE(33,31) BE(34,32) BE(35,33)
EQ BE(35,33) BE(36,34) BE(37,35) BE(38,36) BE(39,37)
FR BE(30,27) BE(31,28) BE(32,29) BE(33,30) BE(34,31) BE(35,32) BE(36,33)
FR BE(37,34) BE(38,35) BE(39,36)
EQ BE(30,27) BE(31,28) BE(32,29) BE(33,30) BE(34,31) BE(35,32) BE(36,33)
EQ BE(36,33) BE(37,34) BE(38,35) BE(39,36)
FR BE(31,27) BE(32,28) BE(33,29) BE(34,30) BE(35,31) BE(36,32) BE(37,33)
FR BE(38,34) BE(39,35)
EQ BE(31,27) BE(32,28) BE(33,29) BE(34,30) BE(35,31) BE(36,32) BE(37,33)
```

```
EQ BE(37,33) BE(38,34) BE(39,35)
FR BE(28,1) BE(29,2) BE(30,3) BE(31,4) BE(32,5) BE(33,6) BE(34,7) BE(35,8)
FR BE(36,9) BE(37,10) BE(38,11) BE(39,12)
EQ BE(28,1) BE(29,2) BE(30,3) BE(31,4) BE(32,5) BE(33,6) BE(34,7) BE(35,8)
EQ BE(35,8) BE(36,9) BE(37,10) BE(38,11) BE(39,12)
FR BE(29,1) BE(30,2) BE(31,3) BE(32,4) BE(33,5) BE(34,6) BE(35,7) BE(36,8)
FR BE(37,9) BE(38,10) BE(39,11)
EQ BE(29,1) BE(30,2) BE(31,3) BE(32,4) BE(33,5) BE(34,6) BE(35,7) BE(36,8)
EQ BE(36,8) BE(37,9) BE(38,10) BE(39,11)
FR BE(30,1) BE(31,2) BE(32,3) BE(33,4) BE(34,5) BE(35,6) BE(36,7) BE(37,8)
FR BE(38,9) BE(39,10)
EQ BE(30,1) BE(31,2) BE(32,3) BE(33,4) BE(34,5) BE(35,6) BE(36,7) BE(37,8)
EQ BE(37,8) BE(38,9) BE(39,10)
FR BE(27,14) BE(28,15) BE(29,16) BE(30,17) BE(31,18) BE(32,19) BE(33,20)
FR BE(34,21) BE(35,22) BE(36,23) BE(37,24) BE(38,25) BE(39,26)
EQ BE(27,14) BE(28,15) BE(29,16) BE(30,17) BE(31,18) BE(32,19) BE(33,20)
EQ BE(33,20) BE(34,21) BE(35,22) BE(36,23) BE(37,24) BE(38,25) BE(39,26)
FR BE(28,14) BE(29,15) BE(30,16) BE(31,17) BE(32,18) BE(33,19) BE(34,20)
FR BE(35,21) BE(36,22) BE(37,23) BE(38,24) BE(39,25)
EQ BE(28,14) BE(29,15) BE(30,16) BE(31,17) BE(32,18) BE(33,19) BE(34,20)
EQ BE(34,20) BE(35,21) BE(36,22) BE(37,23) BE(38,24) BE(39,25)
FR BE(29,14) BE(30,15) BE(31,16) BE(32,17) BE(33,18) BE(34,19) BE(35,20)
FR BE(36,21) BE(37,22) BE(38,23) BE(39,24)
EQ BE(29,14) BE(30,15) BE(31,16) BE(32,17) BE(33,18) BE(34,19) BE(35,20)
EQ BE(35,20) BE(36,21) BE(37,22) BE(38,23) BE(39,24)
FR BE(30,14) BE(31,15) BE(32,16) BE(33,17) BE(34,18) BE(35,19) BE(36,20)
FR BE(37,21) BE(38,22) BE(39,23)
EQ BE(30,14) BE(31,15) BE(32,16) BE(33,17) BE(34,18) BE(35,19) BE(36,20)
EQ BE(36,20) BE(37,21) BE(38,22) BE(39,23)
FR BE(31,14) BE(32,15) BE(33,16) BE(34,17) BE(35,18) BE(36,19) BE(37,20)
FR BE(38,21) BE(39,22)
EQ BE(31,14) BE(32,15) BE(33,16) BE(34,17) BE(35,18) BE(36,19) BE(37,20)
EQ BE(37,20) BE(38,21) BE(39,22)
VA 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4)  PH(5,5) PH(6,6) PH(7,7) PH(8,8)
VA 1 PH(9,9)  PH(10,10) PH(11,11) PH(12,12) PH(13,13)
VA 1 PH(2,1) PH(3,2) PH(4,3) PH(5,4)  PH(6,5) PH(7,6) PH(8,7) PH(9,8)
VA 1 PH(10,9) PH(11,10) PH(12,11) PH(13,12)
VA 1 PH(3,1) PH(4,2) PH(5,3) PH(6,4)  PH(7,5) PH(8,6) PH(9,7) PH(10,8)
VA 1 PH(11,9) PH(12,10) PH(13,11)
VA 1 PH(4,1) PH(5,2) PH(6,3) PH(7,4) PH(8,5) PH(9,6) PH(10,7) PH(11,8)
VA 1 PH(12,9) PH(13,10)
VA 1 PH(5,1) PH(6,2) PH(7,3) PH(8,4) PH(9,5) PH(10,6) PH(11,7) PH(12,8)
VA 1 PH(13,9)
VA 1 PH(6,1) PH(7,2) PH(8,3)  PH(9,4) PH(10,5) PH(11,6) PH(12,7) PH(13,8)
VA 1 PH(7,1) PH(8,2) PH(9,3) PH(10,4) PH(11,5) PH(12,6) PH(13,7)
VA 1 PH(8,1) PH(9,2) PH(10,3) PH(11,4) PH(12,5) PH(13,6)
VA 1 PH(9,1) PH(10,2) PH(11,3) PH(12,4) PH(13,5)
VA 1 PH(10,1) PH(11,2) PH(12,3) PH(13,4)
VA 1 PH(11,1) PH(12,2) PH(13,3)
VA 1 PH(12,1) PH(13,2)
VA 1 PH(13,1)
```

```
FR PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(5,5) PS(6,6) PS(7,7) PS(8,8) PS(9,9)
FR PS(10,10) PS(11,11) PS(12,12) PS(13,13)
EQ PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(5,5) PS(6,6) PS(7,7) PS(8,8) PS(9,9)
EQ PS(9,9) PS(10,10) PS(11,11) PS(12,12) PS(13,13)
FR PS(14,14) PS(15,15) PS(16,16) PS(17,17) PS(18,18) PS(19,19) PS(20,20)
FR PS(21,21) PS(22,22) PS(23,23) PS(24,24) PS(25,25) PS(26,26)
EQ PS(14,14) PS(15,15) PS(16,16) PS(17,17) PS(18,18) PS(19,19) PS(20,20)
EQ PS(20,20) PS(21,21) PS(22,22) PS(23,23) PS(24,24) PS(25,25) PS(26,26)
FR PS(27,27) PS(28,28) PS(29,29) PS(30,30) PS(31,31) PS(32,32) PS(33,33)
FR PS(34,34) PS(35,35) PS(36,36) PS(37,37) PS(38,38) PS(39,39)
EQ PS(27,27) PS(28,28) PS(29,29) PS(30,30) PS(31,31) PS(32,32) PS(33,33)
EQ PS(33,33) PS(34,34) PS(35,35) PS(36,36) PS(37,37) PS(38,38) PS(39,39)
FR TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6) TE(7,7) TE(8,8) TE(9,9)
FR TE(10,10) TE(11,11) TE(12,12) TE(13,13)
EQ TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6) TE(7,7) TE(8,8) TE(9,9)
EQ TE(9,9) TE(10,10) TE(11,11) TE(12,12) TE(13,13)
FR TE(14,14) TE(15,15) TE(16,16) TE(17,17) TE(18,18) TE(19,19) TE(20,20)
FR TE(21,21) TE(22,22) TE(23,23) TE(24,24) TE(25,25) TE(26,26)
EQ TE(14,14) TE(15,15) TE(16,16) TE(17,17) TE(18,18) TE(19,19) TE(20,20)
EQ TE(20,20) TE(21,21) TE(22,22) TE(23,23) TE(24,24) TE(25,25) TE(26,26)
FR TE(27,27) TE(28,28) TE(29,29) TE(30,30) TE(31,31) TE(32,32) TE(33,33)
FR TE(34,34) TE(35,35) TE(36,36) TE(37,37) TE(38,38) TE(39,39)
EQ TE(27,27) TE(28,28) TE(29,29) TE(30,30) TE(31,31) TE(32,32) TE(33,33)
EQ TE(33,33) TE(34,34) TE(35,35) TE(36,36) TE(37,37) TE(38,38) TE(39,39)
FR TE(40,40) TE(41,41) TE(42,42) TE(43,43) TE(44,44) TE(45,45) TE(46,46)
FR TE(47,47) TE(48,48) TE(49,49) TE(50,50) TE(51,51) TE(52,52)
EQ TE(40,40) TE(41,41) TE(42,42) TE(43,43) TE(44,44) TE(45,45) TE(46,46)
EQ TE(46,46) TE(47,47) TE(48,48) TE(49,49) TE(50,50) TE(51,51) TE(52,52)
FR TE(53,53) TE(54,54) TE(55,55) TE(56,56) TE(57,57) TE(58,58) TE(59,59)
FR TE(60,60) TE(61,61) TE(62,62) TE(63,63) TE(64,64) TE(65,65)
EQ TE(53,53) TE(54,54) TE(55,55) TE(56,56) TE(57,57) TE(58,58) TE(59,59)
EQ TE(59,59) TE(60,60) TE(61,61) TE(62,62) TE(63,63) TE(64,64) TE(65,65)
FR TE(66,66) TE(67,67) TE(68,68) TE(69,69) TE(70,70) TE(71,71) TE(72,72)
FR TE(73,73) TE(74,74) TE(75,75) TE(76,76) TE(77,77) TE(78,78)
EQ TE(66,66) TE(67,67) TE(68,68) TE(69,69) TE(70,70) TE(71,71) TE(72,72)
EQ TE(72,72) TE(73,73) TE(74,74) TE(75,75) TE(76,76) TE(77,77) TE(78,78)
FR TD(1,1) TD(2,2) TD(3,3) TD(4,4) TD(5,5) TD(6,6) TD(7,7) TD(8,8) TD(9,9)
FR TD(10,10) TD(11,11) TD(12,12) TD(13,13)
EQ TD(1,1) TD(2,2) TD(3,3) TD(4,4) TD(5,5) TD(6,6) TD(7,7) TD(8,8) TD(9,9)
EQ TD(9,9) TD(10,10) TD(11,11) TD(12,12) TD(13,13)
OU ME=ML IT=55 ND=4
```

Bibliography

Aasness, J., Biørn, E., and Skjerpen, T. (1993), Engel functions, panel data, and latent variables, *Econometrica*, **61**, 1395–422.

Aasness J., Biørn, E. and Skjerpen, T. (1995), Distribution of Preferences and Measurement Errors in a Desegregated Expenditure System, Discussion Paper No. 149. Statistics Norway.

Aigner, D.J. and Goldberger, A.S., Eds. (1977), *Latent Variables in Socio-economic Models.* Amsterdam: North-Holland.

Aigner, D.J., Hsiao, C., Kapteyn, A., and Wansbeek, T. (1984), Latent variable models in econometrics. In: Griliches, Z. and Intriligator, M., Eds. *Handbook of Econometrics.* Amsterdam: North-Holland.

Aiyagari, R. (1994), Uninsured Idiosyncratic Risk and Aggregate Savings. *Quarterly Journal of Economics*, **109**, 659–684.

Amemiya, Y. and Anderson, T.W. (1990), Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics*, **18**(3), 1453–1463.

Anderson, T.W. (1989), Asymptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics*, **1**, 135–141.

Anderson, T.W. (1984), Estimating linear statistical relationships. *Annals of Statistics*, **12**(1), 1–45.

Anderson, T.W. (1971), *An Introduction to Multivariate Statistical Analysis.* 2nd ed. New York: Wiley.

Anderson, T.W. (1989), Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91–119.

Anderson, T.W. (1971), *The Statistical Analysis of Time Series.* New York: Wiley.

Anderson, T.W. and Amemiya, Y. (1988), The asymptotic normal distribution of estimators in factor analysis under general conditions. *Annals of Statistics*, **16**(2), 759–771.

Anderson, T.W. (2002), Reduced rank regression in cointegrated models. *Journal of Econometrics*, **106**, 203–216.

Arellano, M. (2003), *Panel Data Econometrics.* Oxford University Press: Oxford.

Arellano, M. and Bover, O. (1995), Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, **68**, 29–51.

Arminger, G. (1986), Linear differential equation models for panel data with unobserved variables. *Sociological Methodology*, **16**, 187–212.

Bai, J. and Ng. S. (2002), Determining the number of factors in approximate factor models. *Econometrica*, **70**(1), 191–221.

Bai, J. (2003), Inferential theory for factor models of large dimensions. *Econometrica*, **71**(1), 135–171.

Bai, J. (2004), A PANIC attack on unit roots and cointegration. *Econometrica*, forthcoming.

Bartholomew, D.J. (1984), The foundations of factor analysis. *Biometrika*, **71**(2), 221–232.

Bartholomew, D.J. and Knott, M. (1999), *Latent Variable Models and Factor Analysis*. 2^{nd} ed. London: Arnold.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2000), A New Database on Financial Development and Structure. *World Bank Economic Review*, **14**, 597–605.

Beck, T., Levine, R., and Loayza, N. (2000), Finance and the sources of growth. *Journal of Financial Economics*, **58**(1), 261–300.

Beck, T. and Levine, R. (2002), Industry Growth and Capital Allocation: Does Having a Market- or Bank-Based System Matter? *Journal of Financial Economics*, **64**, 147–180.

Beck, T. and Levine, R. (2003), Stock Markets, Banks and Growth: Panel Evidence. *Journal of Banking and Finance*, **28**(3), 423–442.

Bencivenga, V.R. and Smith, B.D. (1991), Financial intermediation and endogenous growth. *Review of Economic Studies*, **58**, 195–209.

Bencivenga, V.R., Smith, B.D., and Starr, R.M. (1995), Transaction costs, technological choice, and endogenous growth. *Journal of Economic Theory*, **67**(1), 53–117.

Berkson, J. (1950), Are there two regressions? *Journal of the American Statistical Association*, **45**, 164–180.

Blanchard, O. and Fischer, S. (1989), *Lectures on Macroeconomics*. Cambridge: MIT Press.

Bloch, A.M. (1989), Identification and estimation of dynamic errors-in-variables models, *Journal of econometrics*, **41**, 145–158.

Bollen, K.A. (1995), Structural equation models that are nonlinear in latent variables: A least-squares estimator. *Sociological Methodology*, **25**, 223–251.

Bollen, K.A. (1996), An alternative two stage least squares (2SLS) estimator for latent variable equations, *Psychometrika*, **61**, 109–121.

Bollen, K.A. (1996), A limited-information estimator for LISREL models with or without heteroscedastic errors. In: Marcoulides, G.A. and Schumacker, R.E. (Eds.), *Advanced Structural Equation Modeling: Issues and Techniques*, Lawrence Erlbaum Associates: Mahwah, New Jersey, pp. 227–241.

Bollen, K.A. (2001), Two-stage least squares and latent variable models: simultaneous estimation and robustness to misspecification, In: Cudeck, R., Du Toit, S. and Sörbom, D. (Eds.), *Structural Equation Modeling: Present and Future.* Chicago: Scientific Software International, 119–138.

Bollen, K.A. (2002), A Note on a Two-Stage Least Squares Estimator for Higher-Order Factor Analysis, *Sociological Methods and Research*, **30**, 568–579.

Boswijk, H.P. (1995), Efficient inference on cointegration parameters in structural error correction models, *Journal of Econometrics*, **69**, 133–158.

Browne, M.W. (1984), Asymptotic distribution free methods in the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.

Brown, M.W. and Arminger, G. (1995), Specification and estimation of mean- and covariance structure models. In: Arminger, G., Clogg, CC., and Sobel, M.E. (Eds.), *Handbook of Statistical Modelling for the Social and Behavioral Sciences.* New-York: Plenum Press, 185–249.

Brown, R.G. and Hwang, P.Y.C. (1992), *Introduction to Random Signals and Applied Kalman Filtering.* New York: John Wiley.

Browning, M. and Lusardi, A. (1996), Household saving: Micro theories and macro facts. *Journal of Economic Literature*, **34**, 1797–1855.

Burmaister, E. and Wall, K.D. (1982), Kalman filtering estimation of unobserved rational expectations with an application to the German hyperinflation. *Journal of Econometrics*, **20**, 255–284.

Campbell, J.Y. and Mankiw, N.G. (1989), Consumption, Income and Interests Rates: Reinterpreting the Time Series Evidence. In: in Blanchard and Fisher (Eds.), *NBER Macroeconomics Annual 1989.* Cambridge: MIT Press, 185-216.

Campbell, J.Y. and Mankiw, N.G. (1990), Permanent income, current income and consumption. *Journal of Business and Economic Statistics*, **8**, 269–279.

Campbell, J.Y. and Mankiw, N.G. (1991), The response of consumption to income: A Cross-country investigation. *European Economic Review*, **35**, 723–67.

Carroll, C.D. (1997), Buffer stock saving and the life cycle/permanent Income hypothesis. *Quarterly Journal of Economics*, **107**, 1–56.

Chamberlain, G. (1990), Distinguished fellow: Arthur S. Goldberger and latent variables in economics. *Journal of Economic Perspectives*, **4**(4), 125–152.

Chamberlain, G. and Rothschild, M. (1983), Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, **51**, 1305–1324.

Chen, C. (1981), The EM approach to the multiple indicator and multiple cause model via the estimation of the latent variable. *Journal of the American Statistical Association*, **76**, 704–708.

Cheng, C.L. and Van Ness, J.W. (1999), *Statistical Regression with Measurement Error.* Arnold: London.

Choi, I. and Phillips, P.C.B. (1992), Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics*, **51**, 113–150.

Connor, G. and Korajzcyk, R. (1993), A test for the number of factors in an approximate factor model, *Journal of Finance*, **48**, 1263–1291.

Cragg, J. and Donald, S. (1997), Inferring the rank of a matrix. *Journal of Econometrics*, **76**, 223–250.

Curran, P.J. and Bollen, K.A. (2001), The best of both worlds: Combining autoregressive and latent curve models. In Collins, L. and Sayer, A. (Eds.), *New Methods for the Analysis of Change.* Washingtong, D.C.: American Psychological Association.

Cziráky, D., Filipan, T., and Pisarović, A. (2003a), Multivariate modelling in experimental designs. *Developments in Applied Statistics*, **19**, 138–155.

Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S. and Polić, M. (2002), Covariance Structure Analysis of Regional Development Data: An Application to Municipality Development Assessment. In: Bade, V., Heinrich, H., Brauser, H.L. (Eds.), *From Industry to Advanced Services: Perspectives of European Metropolitan Region.* Dortmund: European Regional Science Association.

Cziráky, D., Sambt, J., Rovan, J. Puljiz, P., Maleković, S. and Polić, M. (2003b), Regional development assessment using parametric and non-parametric methods: A comparative analysis of Slovenia and Croatia. Proceedings from the 43rd European Regional Science Association (ERSA) Congress, Jyväskyklä: ERSA (R-session papers).

Cziráky, D. (2004), LISREL 8.54: A programme for structural equation modelling with latent variables, *Journal of Applied Econometrics*, **19**, 135–141.

Cziráky, D. (2004d), Estimation of a dynamic structural equation model with latent variables. *Developments in Applied Statistics*, **1**(1), 185–204.

Cziráky, D. and Gillman, M. (2003a), The long-run money demand and the relationship between money and output growth in Croatia: A multivariate cointegration analysis. Proceedings from the ICES 2003 International Conference. Sarajevo: Faculty of Economics.

Cziráky, D., Tišma, S. and Pisarović, A. (2004), Determinants of the low SME loan approval rate in Croatia: A latent variable structural equation approach. *Small Business Economics Journal* (in press).

Dampster, A.P., Laird, N.M., and Rubin, D.B. (1977), Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Deaton, A.S. (1991), Saving and liquidity constraints. *Econometrica*, **59**, 1221–1248.

Deaton, A.S. (1992), *Understanding Consumption.* Oxford: Oxford University Press.

Deistler, M. and Anderson, B.D.O. (1989), Linear dynamic errors-in-variables models: Some structure theory. *Journal of Econometrics*, **41**, 39–63.

Demetriades, P. and Hussein, K. (1996), Does Financial Development Cause Economic Growth? Time Series Evidence from 16 Countries. *Journal of Development Economics*, **51**, 387–411.

Demirgüç-Kunt, A. and R. Levine (2001a), Financial Structure and Economic Growth: Perspectives and Lessons, In: Demirgüç-Knut, A. and Levine, R. Eds., *Financial Structure and Economic Growth: A Cross Country Comparison of Banks, Markets, and Development.* Cambridge, MA: MIT Press, pp. 3–14.

Demirgüç-Kunt, A. and R. Levine (2001b), *Financial Structures and Economic Growth: A Cross-Country Comparison of Banks, Markets, and Development.* Cambridge, MA: MIT Press.

Dhrymes, P.J., Friend, I., and Glutekin, N.B. (1984), A critical re-examination of the empirical evidence on the arbitrage pricing theory, *Journal of Finance*, **39**, 323–346.

Donald, S. (1997), Inference concerning the number of factors in a multivariate nonparametric relationship. *Econometrica*, **65**, 103–132.

Doornik, J.A. and Hansen, H. (1994), A practical test for univariate and multivariate normality. Discussion paper, Nuffield College, University of Oxford.

Dunson (2003), Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, **98**(463), 555–563.

Dupacŏvá, J. and Wold, H. (1982), On some identification problems in ML modeling of systems with indirect observation. In: Jöreskog, K.G. and Wold, H. (Eds.), *Systems under Indirect Observation: Causality, Structure, Prediction*, Part II. Amsterdam: North-Holland, 293–315.

Durbin, J. and Koopman, S.J. (2001), *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Du Toit, S.H.C. and Brown, M.W. (2001), The covariance structure of a vector ARMA time series. In Cudeck, R., Du Toit, S.H.C., and Sörbom, D. (Eds.), *Structural Equation Modeling: Present and Future.* Lincolnwood: Scientific Software International, 279–314.

Du Toit, M. and Du Toit, S. (2001), *Interactive LISREL: User's Guide.* Lincolnwood, IL: Scientific Software International.

Dupuis, M. and Ryan, L.M. (1996), Latent variable models with fixed effects. *Biometrics*, **52**(2), 650–663.

Engle, R.F., Hendry, D.F, and Richard, J.F. (1983), Exogeneity. *Econometrica*, **51**, 277–304.

Engle, R.F., Lilien, D.M., and Watson, M. (1985), A DYMIMIC model of housing price determination. *Journal of Econometrics*, **28**, 307–326.

Engle, R.F. and Watson, M. (1981), A one-factor multivariate time-series model of metropolitan wage rates. *Journal of the American Statistical Association*, **76**(376), 774–781.

Ericsson, N.R. and Irons, J.S. (1994), *Testing Exogeneity.* Oxford: Oxford University Press.

Favara, G. (2003), An empirical reassessment of the relationship between finance and growth. IMF Working Paper 123.

Flavin, M. (1981), The adjustment of consumption to changing expectations about future income. *Journal of Political Economy*, **89**, 974–1009.

Fletcher, R. and Powell, M.J.D. (1963), A rapidly convergent descent method for minimization. *Computer Journal*, **6**, 163–168.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, **82**, 540–554.

Forni, M. and Reichlin, L. (1998), Let's get real: A factor-analytic approach to disaggregated business cycle dynamics. *Review of Economic Studies*, **65**, 453–473.

Giannini, C. (1992), *Topics in Structural VAR Econometrics.* New York: Springer-Verlag.

Gillman, M. and Cziráky, D. (2003b), Inflation and endogenous growth in underground economies. Research Report, WIIW/GDN-SEE, research field II Enterprise Development, Informal Economy and Labour Markets. Vienna: The Vienna Institute for International Economic Studies.

Garratt, A. and Hall, S.G. (1996), Measuring underlying economic activity. *Journal of Applied Econometrics*, **11**(2), 135–151.

Geraci, V.J. (1983), Errors in Variables and the individual structural equation. *International Economic Review*, **24**(1), 217–236.

Geweke, J. (1977), The dynamic factor analysis of economic time series. In: Aigner, D.J. and Goldberger, A.S. (Eds.) *Latent Variables in Socio Economic Models.* Amsterdam: North Holland.

Geweke, J.F. and Singleton, K.J. (1981), Maximum likelihood "confirmatory" factor analysis of economic time series. *International Economic Review*, **22**(1), 37–54.

Ghosh, D. (1989), Maximum likelihood estimation of the dynamic shock-error model. *Journal of Econometrics*, **41**, 121–143.

Gill, P.E., Murray, W., Saunders, M.A., and Wright, M. (1998), *User's Guide for NPSOL 5.0: A Fortran Package for Nonlinear Programming.* Technical Report SOL 86-1, Stanford, CA: Stanford University, Department of Operations Research.

Goldberger, A. (1972a), Maximum likelihood estimation of regressions containing unobservable independent variables. *International Economic Review*, **13**(1), 1–15.

Goldberger, A.S. (1972b), Structural equation methods in the social sciences. *Econometrica*, **40**(6), 979–1001.

Gourinchas, P.O. and Parker, J.A. (2002), Consumption Over the Life Cycle. *Econometrica*, **70**, 47–89.

Griliches, Z. and Hausman, J.A. (1986), Errors in variables in panel data. *Journal of Econometrics*, **31**, 93–118.

Gruvaeus, G.T. and Jöreskog, K.G. (1970), A computer program for minimizing a function of several variables. *Educational Testing Service Research Bulletin*, **70**(14).

Hall, A.R. (2005), *Generalized Method of Moments.* Oxford: Oxford University Press.

Hahm, J.H. and Steigerwarld, D.G. (1999), Consumption adustment under time-varying income uncertainty. *Review of Economics and Statistics*, **81**, 32–40.

Ellen L. Hamaker, Conor V. Dolan, and Peter C. M. Molenaar (2002), On the Nature of SEM Estimates of ARMA Parameters. *Structural Equation Modelling*, **9**(3), 347-368.

Ellen L. Hamaker, Conor V. Dolan, and Peter C. M. Molenaar (2002), ARMA-Based SEM When the Number of Time Points T Exceeds the Number of Cases N. *Structural Equation Modelling*, **10**(3), 352-379.

Hali, J.E., Levine, R., Ricci, L., and Sløk, T. (2002), International financial integration and economic growth. IMF Working Paper 145.

Hamilton, J. (1986), A standard error for the estimated state vector of a state-space model. *Journal of Econometrics*, **33**, 387–397.

Hamilton, J. (1994), *Time Series Analysis.* New Jersey: Princeton University Press.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Hayashi, F. (1982), The permanent income hypothesis: Estimation and testing by instrumental variables. *Journal of Political Economy*, **90**, 895–916.

Hershberger, S.L., Corneal, S.E., and Molenaar, P.C.M. (1994), Dynamic factor analysis: An application to emotional response patterns underlying daughter/father and stepdaughter/stepfather relationships. *Structural Equation Modeling*, **2**, 31–52.

Hershberger, S.L., Corneal, S.E., and Molenaar, P.C.M. (1996), A hierarchy of univariate and multivariate structural time series models. In Marcoulides, G.A. and Schumacker, R.E. (Eds.), *Advanced structural equation modeling: Issues and techniques.* Mahwah, NJ: Lawrence Erlbaum Associates, 159-194.

Hsiao, C. (1997), Cointegration and dynamic simultaneous equations model. *Econometrica*, **65**(3), 647–670.

Hsiao, C. (2003), *Analysis of Panel Data.* 2nd Ed. Cambridge: Cambridge University Press.

Hubbard, R.G. and Judd, K.J. (1986), Liquidity constraints, fiscal policy and consumption. *Brookings Papers on Economics Activity*, **1**, 1–59.

Jansen, R.A.R.G., and Oud, J.H.L. (1995), Longitudinal LISREL model estimation from incomplete panel data using the EM algorithm and the Kalman smoother. *Statistica Neerlandica*, **49**, 362-377.

Jappelli, T. and Pagano, M. (1989), Consumption and capital market imperfections: An international comparison. *American Economic Review*, **79**, 1088–1105.

Jöreskog, K.G. (1969), A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34**, 183–202.

Jöreskog, K.G. (1970), A general method for analysis of covariance structures. *Biometrika*, **57**(2), 239–251.

Jöreskog, K.G. (1973), A general method for estimating a linear structural equation system. In Goldberger, A.S. and Duncan, O.D. (Eds.), *Structural Equation Models in the Social Sciences.* Chicago: Academic Press, 85–112.

Jöreskog, K.G. (1977), Structural equation models in the social sciences: Specification, estimation and testing. In: Krishnaiah, P.R. (Ed.) *Application of Statistics*, Amsterdam: North-Holland, 265–287.

Jöreskog, K.G. and Goldberger, A. (1975), Factor analysis by generalized least squares, *Psychometrika*, **37**(35), 243–260.

Jöreskog, K.G. and Goldberger, A. (1975), Estimation of a model with multiple indicators and multiple causes of a single latent variable, *Journal of the American Statistical Association*, **70**(35), 631–639.

Jöreskog, K.G., Gruvaeus, G.T., and van Thillo, M. (1970), ACOVS: A general computer program for analysis of covariance structures. *Educational Testing Service Research Bulletin*, **70**(15).

Jöreskog, K.G. and Sörbom, D. (1977), Statistical methods for analysis of longitudinal data. In: Aigner, D.J. and Goldberger, A.S. (Eds.) *Latent Variables in Socioeconomic Models.* Amsterdam: North-Holland, 285–325.

Jöreskog, K.G. and Sörbom, D. (1996a), *PRELIS 2 User's Reference Guide: A Program for Multivariate Data Screening and Data Summarization; A Preprocessor for LISREL.* Chicago: Scientific Software International.

Jöreskog, K.G. and Sörbom, D. (1996b), *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language.* Chicago: Scientific Software International.

Jöreskog, K.G. and Sörbom, D. (1996c), *LISREL 8 User's Reference Guide.* Chicago: Scientific Software International.

Jöreskog, K. G. and D. Sörbom, S. Du Toit and M. Du Toit (2001), *LISREL 8: New Statistical Features.* Lincolnwood, IL: Scientific Software International.

Jöreskog, K.G. (1977a), Factor analysis by least-squares and maximum-likelihood methods. In Enslein, K., Ralston, A., and Wilf, H.S. (Eds.), *Statistical Methods for Digital Computers.* New York: Wiley.

Jöreskog, K.G. (1977b), Structural equation models in the social sciences: Specification, estimation and testing. In Krishnaiah, P.R. (Ed.), *Applications of statistics.* Amsterdam: North-Holland, 265–286.

Jöreskog, K.G. (1979), Basic ideas of factor and component analysis. In Jöreskog, K.G. and Sp̈rbom, D. (Eds.), *Advances in Factor Analysis and Structural Equation Models.* Cambridge, MA: Abt Books, 5–20.

Jöreskog, K.G. (1981), Analysis of covariance structures. *Scandinavian Journal of Statistics*, **8**, 65–92.

Jöreskog, K.G. (2003), Factor analysis by MINRES. Mimeo (http://www.ssicentral.com).

Judge, G.G., Hill, R.C., Griffiths, Lütkepohl, H., and Lee, C.-T. (1988), *Introduction to the Theory and Practice of Econometrics.* New York: John Wiley.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME Series D*, **82**, 35-45.

Kano, Y. (1986), Conditions on consistency of estimators in covariance structure model, *Journal of the Japan Statistical Society*, **16**, 75–80.

Kay, J.A., Keen, M.J. and Morris, C.N. (1984), Estimating consumption from expenditure data. *Journal of Public Economics*, **23**, 169–181.

Keesling, J.W. (1972), *Maximum Likelihood Approaches to Causal Flow Analysis.* University of Chicago: School of Education PhD dissertation.

King, R.G. and Levine, R. (1993a), Finance and Growth: Schumpeter Might Be Right", *Quarterly Journal of Economics*, **108**, 717–738.

King, R.G. and Levine, R. (1993b), Finance, entrepreneurship, and growth: Theory and evidence. *Journal of Monetary Economics*, **32**, 513–542.

Koopman, S.J., Shephard, N., and Doornik, J.A. (1999), Statistical alghorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, **2**, 113–166.

Larsen, E.R. (2002), Estimating latent total consumption in a household. Discussion Papers No. 324, Statistics Norway.

Lee, S.-Y. and Jennrich, R.I. (1979), A study of algorithms for covariance structure analysis with specific comparison using factor analysis. *Psychometrika*, **44**(1), 99–113.

Levine, R. (1997), Financial development and growth: Views and agenda. *Journal of Economic Literature*, **35**, 688–726.

Levine, R. (1999), Law, Finance, and Economic Growth. *Journal of Financial Intermediation*, **8**, 36–67.

Levine, R., Loayza, N., and Beck, T. (2000), Financial Intermediation and Growth: Causality and Causes. *Journal of Monetary Economics*, **46**, 31–77.

Levine, R. (2003), Finance and growth: Theory, evidence, and mechanisms, forthcoming in Handbook of Economic Growth.

Levine, R., Loayza, N., and Beck, T. (2000), Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, **46**(1), 31–77.

Levine, R., Beck, T., and Demirguc-Knut, A. (2001), A new database on the structure and development of the financial sector. *World Bank Economic Review*. **14**(3), 597–605.

Levine, R. and Zervos, S. (1996), Stock markets development and long-run growth. *World Bank Economic Review*, **10**, 323–339.

Levine, R. and Zervos, S. (1998), Stock markets, banks, and economic growth. *American Economic Review*, **88**, 537–558.

Lewbel, A. (1991), The rank of demand systems: Theory and nonparametric estimation. *Econometrica*, **59**, 711–730.

Liu, C. and Rubin D.B (1994), The $ECME$ algorithm: A simple extension of EM and ECM with faster monotome convergence. *Biometrika*, **81**(4), 633–648.

Liu, C. and Rubin D.B, and Wu, Y.N. (1998), Parameter expansion to accelerate EM: The $PX - EM$ algorithm. *Biometrika*, **81**(4), 633–648.

Lucas, R.E. (1988), On the mechanics of economic development. *Journal of Monetary Economics*, **22**(1), 3-42.

Lucchetti, R. (1999), Analytic score for multivariate GARCH models. Manuscript.

Ludvigson, S. and Paxson, C. (2001), Approximation bias in linearized Euler equations. *Review of Economics and Statistics*, **83**, 242–256.

Madsen, J.B. and McAleer, M. (2000), Direct tests of the permanent income hypothesis under uncertainty, inflationary expectations, and liquidity constraints. *Journal of Macroeconomics*, **22**, 229–252.

Madsen, J.B. and McAleer, M. (2001), Consumption, liquidity constraints, uncertainty and temptation: An international comparison. *Journal of Economic Psychology*, **22**, 61–89.

MacCallum, R. and Ashby, F.G. (1986), Relationship between linear systems theory and covariance structure modeling. *Journal of Mathematical Psychology*, **30**, 1–27.

Magnus, J.R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Chichester: Wiley.

Mallows, C.L. (1973), Some comments on C_p. *Technometrics*, **15**, 661–675.

Mardia, K.V. (1970), Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**(3), 519–530.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979), *Multivariate Analysis.* London: Academic Press.

Maravall, A. (1979), *Identification in Dynamic Shock-error Models.* New York: Springer-Verlag.

Maravall, A. and Aigner, D.J. (1977), Identification of the dynamic shock-error model: The case of dynamic regression. In: Aigner, D.J. and Goldberger, A.S., Eds. *Latent Variables in Socio-economic Models.* Amsterdam: North-Holland.

McArdle, J.J. (1988), Dynamic but structural equation modeling of repeated measures data. In Nesselroade, J.R. and Cattell, R.B. (Eds.), *Handbook of Multivariate Experimental Psychology.* New York: Plenum, 561-614.

McArdle, J.J. (2001), A latent difference score approach to longitudinal dynamic structural analysis. In Cudeck, R., Du Toit, S.H.C., and Sörbom, D. (Eds.), *Structural Equation Modeling: Present and Future.* Lincolnwood: Scientific Software International, 341–380.

Meinhold, R.J. and Singpurwalla, N.D. (1983), Understanding the Kalman filter. *American Statistician*, **37**(2), 123–127.

Meng, X.-L. and Rubin, D.B. (1991), Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Biometrika*, **80**(2), 276–278.

Meng, X.-L. and Rubin, D.B. (1993), Maximum likelihood estimation via the ECM algorithm: A general approach. *Biometrika*, **80**(2), 276–278.

Molenaar, P.C.M. (1985), A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, **50**, 181–202.

Molenaar, P.C.M., de Gooijer, J.G., and Schmitz, B. (1992), Dynamic factor analysis of non-stationary multivariate time series. *Psychometrika*, **57**, 333–349.

Molenaar, P.C.M. (1999), Comment on fitting MA time series by structural equation models. *Psychometrika.* **64**, 91-94.

Muellbauer, J. and Lattimore, R. (1995), The consumption function: A theoretical and empirical review. In: Pasaran, M.H. and Wickens, M.R. (eds.), *Handbook of Applied Econometrics: Macroeconomics.* Oxford: Blackwell, 221–311.

Neale, M.C., Boker, S.M., Xie, G., and Maes, H.H. (2004), *Mx: Statistical Modeling.* Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Department of Psychiatry.

Neudecker, H. and Satorra, A. (1991), Linear structural relations: Gradient and Hessian of the fitting function. *Statistics & Probability Letters*, **11**, 57–61.

Neudecker, H. and Satorra, A. (1994), On the asymptotic optimality of alternative minimum-distance estimators in linear latent-variable models. *Econometric Theory*, **10**, 867–883.

Neuser, K. and Kugler, M. (1998), Manufacturing growth and financial development: Evidence from OECD countries. *Review of Economics and Statistics*, **80**, 636–646.

Norrbin, S.C. and Schlagenhauf, D.E. (1995), The role of international factors in the business cycle: A multi-country study. *Journal of International Economics*, **40**, 85–104.

Oud, J.H.L. (2001), Quasi-longitudinal designs in SEM state space modelling. *Statistica Neerlandica*, **55**(2), 200–220.

Oud, J.H.L. (2002), Continuous time modeling of the cross-lagged panel design. *Kwantitatieve Methoden*, **69**, 1–26.

Oud, J.H.L. (2005), Continuous time modeling of reciprocal relationships in the cross-lagged panel design. In Boker, S.M. and Wenger, M.J. (Eds.), Data analytic techniques for dynamical systems in the social and behavioral sciences. Mahwah, NJ: Lawrence Erlbaum Associates.

Oud, J.H.L. (2004), SEM state space modeling of panel data in discrete and continuous time and its relationship to traditional state space modeling. In van Montfort, K., Oud, J.H.L. and Satorra, A. (Eds.), *Recent developments on structural equations models: Theory and applications.* Dordrecht: Kluwer Academic Publishers, 13–40.

Oud, J.H.L., and Jansen, R.A.R.G. (2000), Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, **65**, 199–215.

Oud, J.H.L., Jansen, R.A.R.G., and Haughton, D.M.A. (1999), Small samples in structural equation state space modeling. In R. Hoyle (Ed.), *Statistical strategies for small sample research.* Thousand Oakes: Sage, 288–305.

Oud, J.H.L., and Jansen, R.A.R.G. (1996), Nonstationary longitudinal LISREL model estimation from incomplete panel data using EM and the Kalman smoother. In U. Engel and J. Reinecke (Eds.), *Analysis of change: advanced techniques in panel data analysis.* Berlin: Walter de Gruyter, 135–159.

Oud, J.H.L., and Jansen, R.A.R.G. (1995). An ARMA extension of the longitudinal LISREL model for LISKAL. In Partchev, I. (Ed.), *Multivariate analysis in the behavioral sciences: philosophic to technical.* Sofia: "Prof. Marin Drinov" Academic Publishing House, 49–69.

Oud, J.H.L., van Leeuwe, J.F.J., and Jansen, R.A.R.G. (1993), Kalman filtering in discrete and continuous time based on longitudinal LISREL models. In Oud, J.H.L. and van Blokland-Vogelesang, A.W. (Eds.), *Advances in longitudinal and multivariate analysis in the behavioral sciences*, Nijmegen: ITS, 3–26.

Pagan, A. (1973), Efficient estimation of models with composite disturbance terms. *Journal of Econometrics*, **1**, 329–340.

Pena, D. and Box, G.E.P. (1987), Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, **82**(399), 836–843.

Pesaran, M.H., Smith, R.P. and Yeo, J.S. (1999), An autoregressive distributed lag modelling approach to cointegration analysis. In: Strom, S. (Ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, Cambridge: Cambridge University Press.

Rousseau, P.L. and Wachtel, P. (2000), Equity markets and growth: Cross-country evidence on timing and outcomes, 1980–1995. *Journal of Business and Finance*, **24**, 1933-1957.

Sargan, J.D. (1958), The estimation of economic relationships using instrumental variables. *Econometrica*, **26**, 393–415.

Sargan, J.D. (1959), The Estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society*, **21**, 91–105.

Sargan, J.D. (1988), Wages and prices in United Kingdom: A study in econometric methodology. In Sargan, J.D. (Ed.), *Contributions to Econometrics*, Vol. 1. Cambridge: Cambridge University Press, 124–169.

Sargent, T. and Sims, C. (1977), Business cycle modelling without pretending to have too much a priori economic theory. In: Sims, C. (Ed.), *New Methods in Business Cycle Research.* Minneapolis: Federal Reserve Bank of Minneapolis.

Satorra, A. (1989), Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, **54**(1), 131–151–.

Satorra, A. (1990), Robustness issues in structural equatino modeling: A review of recent developments. *Quality & Quantity*, **24**, 367–386.

Satorra, A. (1992a), Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures. *Sociological Methodology*, **22**, 249–278.

Satorra, A. (1992b), The variance matrix of sample second-order moments in multivariate linear relations. *Statistics & Probability Letters*, **15**, 63–69.

Satorra, A., van Montfort, K., and Oud, H. (2004), *Recent Developments on structural equations models: Theory and Applications.* Kluwer Academic Publishers.

Scheinkman, J.A. and Weiss, L. (1986), Borrowing constraints and aggregate economic activity. *Econometrica*, **54**, 23–46.

Scherer, B. and Douglas, M. (2005), Introduction to Modern Portfolio Optimization with NuOPT, S-Plus, and S+Bayes. New York: Springer.

Shapiro, A. (1983), Asymptotic distribution theory in the analysis of covariance structures (a unified approach), *South African Statistical Journal*, **17**, 33–81.

Shapiro, A. (1984), A note on the consistency of estimators in the analysis of moment structures, *British Journal of Mathematical and Statistical Psychology*, **37**, 84–88.

Shapiro, A. and Brown, M.W. (1987), Analysis of covariance structures under elliptical distributions, *Journal of the American Statistical Association*, **82**(400), 1092–1097.

Singleton, K.J. (1980), A latent time series model of the cyclical behaviour of interest rates. *International Economic Review*, **21**(3), 559–575.

Skrondal, A. and Rabe-Hesketh, S. (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Chapman and Hall/CRC.

Stock, J.H. and Watson, M. (1989), New indexes of coincident and leading economic indications. In: Blanchard, O.J. and Fisher, S. (Eds.), *NBER Macroeconomic Annual 1989.* Cambridge, MA.: M.I.T. Press.

Stock, J.H. and Watson, M. (1998), Diffusion indexes. NBER Working Paper 6702.

Stock, J.H. and Watson, M. (1999), Forecasting inflation. *Journal of Monetary Economics*, **44**, 293–335.

Taylor, M.F. (Ed.) with Bruce J., Nick, B. and Prentice-Lane, E. (2005), *British Household Panel Survey User Manual, Volume A: Introduction, Technical Report and Appendices.* Colchester, University of Essex.

Terceiro Lomba, J. (1990), *Estimation of Dynamic Econometric Models with Errors in Variables.* Berlin: Springer-Verlag.

Turkington, D.A. (1998), Efficient estimation in the linear simultaneous equations model with vector autoregressive disturbances. *Journal of Econometrics*, **85**, 51–74.

Turkington, D. (2002), *Matrix Calculus and Zero-One Matrices: Statistical and Econometric Applications.* Cambridge: Cambridge University Press.

Ventura, E. and Satorra, A. (1998), Life-cycle effects on household expenditure: A latent-variable approach. Universitat Pompeu Fabra.

Wansbeek, T. (2001), GMM estimation in panel data models with measurement error. *Journal of Econometrics*, **104**, 259–268.

Wansbeek, T. and Meijer, E. (2000), *Measurement Error and Latent Variables in Econometrics.* Amsterdam: North Holland.

Watson, M.W. and Engle, R.F. (1983), Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *Journal of Econometrics*, **23**, 385–400.

Watson, M. and Kraft, D.F. (1984), Testing the interpretation of indices ini a macroeconomic index model. *Journal of Monetary Economics*, **13**, 165–181.

Zellner, A. (1970), Estimation of regression relationships containing unobservable independent variables. *International Economic Review,* **11**(3), 441–454.

Zivot, E., Wang, J., and Koopman, S.J. (2002), State Space Modeling in Macroeconomics and Finance Using *SsfPack* in `S+FinMetrics`. Paper prepare for the Academy Colloquium Masterclass on State Space and Unobserved Components Models in honour of Professor J. Durbin.

Zivot, E. and Wang, J. (2006), *Modeling Financial Time Series with S-PLUS.* Springer.

Van Buuren, S. (1997), Fitting ARMA time series by structural equation models. *Psychometrika,* **62**, 215-236.

Venables, W.N. and Ripley, B.D. (2000), S Programming. New York: Springer

Venables, W.N. and Ripley, B.D. (2000), Modern Applied Statistics with S. New York: Springer

Wiley, D.E. (1973), The identification problem for structural equation models with unmeasured variables. In Goldberger, A.S. and Duncan, O.D. (Eds.), *Structural Equation Models in the Social Sciences.* Chicago: Academic Press, 69–83.

Zeldes, S.P. (1989), Consumption and liquidity constrains: An empirical investigation. *Journal of Political Economy*, **97**, 305–46.

Zivot, E. and Wang, J. (2005), *Modeling Financial Time Series with S-PLUS.* 2nd Ed. New York: Springer.

Zivot, E., Wang, J., and Koopman, S.J. (2004), State space models in economics and finance using SsfPack in S+FinMetrics. In: Harvey, A., Koopman, S.J., and Shephard, N. (Eds.), *Unobserved Components Models.* Cambridge: Cambridge University Press.

Yamashita, H. (1992), A globally convergent primal-dual interior point method for constrained optimization, Technical Report, Mathematical Systems Institute In c., Tokyo, Japan.

Yamashita, H. and H. Yabe, H. (1993), Superlinear and quadratic convergence of primaldual interior point methods for constrained optimization Technical Report, Mathematical Systems Institute Inc., Tokyo, Japan.

Yamashita, H. and Tanabe, T. (1993). A primal-dual interior point trust region method for large scale constrained optimization, Technical Report, Mathematical Systems Institute Inc., Tokyo, Japan.