London School of Economics and Political Science

Essays in Empirical Asset Pricing

Svetlana Bryzgalova

Thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of

Doctor of Philosophy

July 2015

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Statement of conjoint work
I confirm that chapter 3 is jointly co-authored with Christian Julliard and I contributed 50 % of this work.

I declare that my thesis consists of 44, 821 words.

There were big adventures and small adventures, Mr Bunnsy knew. You didnt get told what size they were going to be before you started. Sometimes you could have a big adventure even when you were standing still.

Terry Pratchett, *The Amazing Maurice and His Educated Rodents*

Acknowledgements

My deepest gratitude goes to my advisors, Christian Julliard and Peter Robinson, for their generous, invaluable support, guidance, patience and trust. You were the best advisors I could ever wish for.

I am very grateful to fellow researchers. Being a 'child' of both Economics Department and Financial Markets Group, I was particularly lucky to be surrounded by so many talented, insightful and dedicated people. I would like to thank Oliver Linton, Andrea Tamoni, Tatiana Komarova, Philippe Mueller, Marcia Schafgans, Daniel Ferreira, Myung Hwan Seo, Ian Martin, Dong Lou, Javier Hidalgo, Christopher Polk, Kathy Yuan, Taisuke Otsu, Ulf Axelson, and late Sudipto Bhattacharya for all the help and feedback over the years.

I will miss the inspiring atmosphere of the Financial Markets Group, thanks to my fellow PhD students: Huazhi Chen, Sean Lew, Pedro Pinto, John Kuong, Abhimanyu Gupta, Yiqing Lu, Sergey Glebkin, and Olga Obizhaeva. You have been there with me through highs and lows.

I would also like to acknowledge the financial support from the Economics Department at LSE.

I owe a lot to my parents, who were the source of everlasting love and support over the years. Without them, I would never have gotten to where I am today. And very special thanks goes to my partner, Andrey, who was there every step of the way, always ready to give that extra push.

Abstract

In this thesis, I study asset pricing models of stock and bond returns, and the role of macroeconomic factors in explaining and forecasting their dynamics.

The first chapter is devoted to the identification and measurement of risk premia in the cross-section of stocks, when some of the risk factors are only weakly related to asset returns and, as a result, spurious inference problems are likely to arise. I develop a new estimator for cross-sectional asset pricing models that, simultaneously, provides model diagnostic and parameter estimates. This novel approach removes the impact of spurious factors and restores consistency and asymptotic normality of the parameter estimates. Empirically, I identify both robust factors and those that instead suffer from severe identification problems that render the standard assessment of their pricing performance unreliable (e.g. consumption growth, human capital proxies and others).

The second chapter extends the shrinkage-based estimation approach to the class of affine factor models of the term structure of interest rates, where many macroeconomic factors are known to improve the yield forecasts, while at the same time being unspanned by the cross-section of bond returns.

In the last chapter (with Christian Julliard), we propose a simple macro model for the co-pricing of stocks and bonds. We show that aggregate consumption growth reacts slowly, but significantly, to bond and stock return innovations. As a consequence, slow consumption adjustment (SCA) risk, measured by the reaction of consumption growth cumulated over many quarters following a return, can explain most of the cross-sectional variation of expected bond and stock returns. Moreover, SCA shocks explain about a quarter of the time series variation of consumption growth, a large part of the time series variation of stock returns, and a significant (but small) fraction of the time series variation of bond returns, and have substantial predictive power for future consumption growth.

Contents

List of Tables

List of Figures

Chapter 1

Spurious Factors in Linear Asset Pricing Models

1.1 Introduction

Sharpe (1964) and Lintner (1965) CAPM pioneered the class of linear factor models in asset pricing. Now, decades later, what started as an elegant framework has turned into a well-established and successful tradition in finance. Linear models, thanks to their inherent simplicity and ease of interpretation, are widely used as a reference point in much of the empirical work, having been applied to nearly all kinds of financial assets[1]. In retrospect, however, such heavy use produced a rather puzzling outcome: Harvey, Liu, and Zhu (2013) list over 300 factors proposed in the literature, all of which have been claimed as important (and significant) drivers of the cross-sectional variation in stock returns[2].

One of the reasons for such a wide range of apparently significant risk factors is perhaps a

[1] Notable examples are the 3-factor model of Fama and French (1992), Fama and French (1993); the conditional CAPM of Jagannathan and Wang (1996); the conditional CCAPM of Lettau and Ludvigson (2001b), the Q1-Q4 consumption growth of Jagannathan and Wang (2007), the durable/nondurable consumption CAPM of Yogo (2006); the ultimate consumption risk of Parker and Julliard (2005); the pricing of currency portfolios in Lustig and Verdelhan (2007) and Lustig, Roussanov, and Verdelhan (2011); and the regression-based approach to the term structure of interest rates in Adrian, Crump, and Moench (2013)

[2] In the context of predictive regressions, Novy-Marx (2014) recently demonstrated that many unconventional factors, such as the party of the U.S. President, sunspots, the weather in Manhattan, planet location and the El-Nino phenomenon have a statistically significant power for the performance of many popular trading strategies, such as those based on market capitalisation, momentum, gross profitability, earnings surprises and others.

simple lack of model identification, and consequently, an invalid inference about risk premia parameters. As pointed out in a growing number of papers (see e.g. Jagannathan and Wang (1998), Kan and Zhang (1999b), Kleibergen (2009), Kleibergen and Zhan (2013), Burnside (2010), Gospodinov, Kan, and Robotti (2014a)), in the presence of factors that only weakly correlate with assets (or do not correlate at all), all the risk premia parameters are no longer strongly identified and standard estimation and inference techniques become unreliable. As a result, identification failure often leads to the erroneous conclusion that such factors are important, although they are totally *spurious* by nature. The impact of the true factors could, in turn, be crowded out from the model.

The shrinkage-based estimators that I propose (Pen-FM and Pen-GMM, from the *penalised* version of the Fama-MacBeth procedure or GMM, accordingly), not only allow to detect the overall problem of rank-deficiency caused by irrelevant factors, but also indicate which particular variables are causing it, and recover the impact of strong risk factors without compromising any of its properties (e.g. consistency, asymptotic normality, etc).

My estimator can bypass the identification problem because, in the case of useless (or weak) factors, we know that it stems from the low correlation between these variables and asset returns. This, consequently, is reflected in the regression-based estimates of *betas*, asset exposures to the corresponding sources of risk. Therefore, one can use the L_1-norm of the vector of $\hat{\beta}$'s (or related quantities, such as correlations) to assess the overall factor strength for a given cross-section of returns, and successfully isolate the cases when it is close to zero. Therefore, I modify the second stage of the Fama-MacBeth procedure[1] (or the GMM objective function) to include a penalty that is inversely proportional to the factor strength, measured by the L_1-norm of the vector $\hat{\beta}$.

One of the main advantages of this penalty type is its ability to simultaneously recognise the presence of both useless and weak factors[2], allowing Pen-FM(GMM) to detect the problem of both under- and weak identification. On the contrary, the critical values for the tests often used in practice[3] are all derived under the assumption of *strictly zero* correlation between the factor and returns. As a result, faced with a *weak* factor, such tests tend to

[1]The problem of identification is not a consequence of having several stages in the estimation. It is well known that the two-pass procedure gives exactly the same point estimates as GMM with the identity weight matrix under a particular moment normalisation.

[2]If the time series estimates of beta have the standard asymptotic behaviour, then for both useless ($\beta = \mathbf{0_n}$) and weak ($\beta = \frac{B}{\sqrt{T}}$) factors L_1-norm of $\hat{\beta}$ is of the order $\frac{1}{\sqrt{T}}$.

[3]Wald test for the joint spread of betas or more general rank deficiency tests, such as Cragg and Donald (1997), Kleibergen and Paap (2006)

reject the null hypothesis of betas being jointly zero; however, risk premia parameters still have a nonstandard asymptotic distribution, should the researcher proceed with the standard inference techniques[1].

Combining model selection and estimation in one step is another advantage of Pen-FM(GMM), because it makes the model less prone to the problem of *pretesting*, when the outcome of the initial statistical procedure and decision of whether to keep or exclude some factors from the model further distort parameter estimation and inference[2].

Eliminating the influence of irrelevant factors is one objective of the estimator; however, it should also reflect the pricing ability of other variables in the model. I construct the penalty in such a way that does not prevent recovering the impact of strong factors. In fact, I show that Pen-FM(GMM) provide consistent and asymptotically normal estimates of the strong factors risk premia that have *exactly* the same asymptotic distribution as if the irrelevant factors had been known and excluded from the model *ex ante*. Further, I illustrate, with various simulations, that my estimation approach also demonstrates good finite sample performance even for a relatively small sample of 50-150 observations. It is successful in a) eliminating spurious factors from the model, b) retaining the valid ones, c) estimating their pricing impact, and d) recovering the overall quality of fit.

I revisit some of the widely used linear factor models and confirm that many tradable risk factors seem to have substantial covariance with asset returns. This allows researchers to rely on either standard or shrinkage-based estimation procedures, since both deliver identical point estimates and confidence bounds (e.g. the three-factor model of Fama and French (1993), or a four-factor model that additionally includes the quality-minus-junk factor of Asness, Frazzini, and Pedersen (2014)).

There are cases, however, when some of the factors are particularly weak for a given cross-section of assets, and their presence in the model only masks the impact of the true sources of risk. The new estimator proposed in this chapter allows then to uncover this relationship and identify the actual pricing impact of the strong factors. This is the case, for example,

[1] A proper test for the strength of the factor should be derived under the null of weak identification, similar to the critical value of 10 for the first stage F-statistics in the case of a single endogenous variable and 1 instrument in the IV estimation, or more generally the critical values suggested in Stock and Yogo (2005)

[2] See, e.g. simulation designs in Breiman (1996) highlighting the model selection problem in the context of linear regressions and the choice of variables, Guggenberger (2010) for the impact of Hausman pretest in the context of panel data, and Berk, Brown, Buja, Zhang, and Zhao (2013) for recent advances in constructing confidence bounds, robust to prior model selection

of the q-factor model of Hou, Xue, and Zhang (2014) and the otherwise 'hidden' impact of the profitability factor, which I find to be a major driving force behind the cross-sectional variation in momentum-sorted portfolios.

Several papers have recently proposed[1] asset pricing models that highlight, among other things, the role of investment and profitability factors, and argue that these variables should be important drivers of the cross-sectional variation in returns, explaining a large number of asset pricing puzzles[2]. However, when I apply the q-factor model (Hou, Xue, and Zhang (2014)) to the momentum-sorted cross-section of portfolios using the Fama-MacBeth procedure, none of the variables seem to command a significant risk premium, although the model produces an impressive R^2 of 93%. Using Pen-FM on the same dataset eliminates the impact of two out of four potential risk drivers, and highlights a significant pricing ability of the profitability factor (measured by ROE), largely responsible for 90% of the cross-sectional variation in portfolio returns. Point estimates of the risk premia (for both market return and ROE), produced by Pen-FM in this case are also closer to the average return generated by a tradable factor, providing further support for the role of the firm's performance in explaining the momentum effect, as demonstrated in Hou, Xue, and Zhang (2014). The importance of this factor in explaining various characteristics of stocks is also consistent with the findings of Novy-Marx (2013), who proposes an alternative proxy for expected profitability and argues that it is crucial in predicting the cross-sectional differences of stock returns.

While specifications with tradable factors seem to be occasionally contaminated by the problem of useless factors, the situation seems to be much worse when a nontradable source of risk enters into the model. For example, I find that specifications including such factors as durable consumption growth or human capital proxies are not strongly identified[3] and Pen-FM shrinks their risk premia towards zero. Since conventional measures of fit, such as the crossectional R^2, are often inflated in the presence of spurious factors (Kleibergen and Zhan (2013), Gospodinov, Kan, and Robotti (2014b)), their high in-sample values only mask a poorly identified model.

[1]E.g. Fama and French (2015) and Hou, Xue, and Zhang (2014)

[2]There is vast empirical support for shocks to a firm's profitability and investment to be closely related to the company's stock performance, e.g. Ball and Brown (1968), Bernand and Thomas (1990), Chan, Jegadeesh, and Lakonishok (1996), Haugen and Baker (1996), Fairfield, Whisenant, and Yohn (2003), Titman, Wei, and Xie (2004), Fama and French (2006), Cooper, Gulen, and Schill (2008), Xing (2008), Polk and Sapienza (2009), Fama and French (2015)

[3]This finding is consistent with the results of identification tests in Zhiang and Zhan (2013) and Burnside (2010)

It is worth noting, however, that when a particular risk driver is identified as weak (or useless), it does not necessarily render the model containing it invalid. The finding merely highlights the impossibility of assessing the size of the risk premia paremeters, significance of their pricing impact and the resulting quality of fit, based on the standard estimation techniques. The method that I propose allows to recover identification and quality of fit only for strong risk factors (which is contaminated otherwise), but stays silent regarding the impact of the weak ones. Furthermore, since I focus on the multiple-beta representation, the risk premia reflect the *partial* pricing impact of a factor. Therefore, it is also plausible to have a model with a factor being *priced* within a linear SDF setting, but not contributing anything on its own, that is *conditional on other factors in the model*. When estimated by the Fama-MacBeth procedure, its risk premium is no longer identified. Although the focus of this chapter is on the models that admit multivariate beta-representation, nothing precludes extending shrinkage-based estimators to a linear SDF setting to assess the *aggregate* factor impact as well.

Why does identification have such a profound impact on parameter estimates? The reason is simple: virtually any estimation technique relies on the existence of a unique set of true parameter values that satisfies the model's moment conditions or minimises a loss function. Therefore, violations of this requirement in general deliver estimates that are inconsistent, have non-standard distribution, and require (when available) specifically tuned inference techniques for hypothesis testing. Since the true, population values of the β's on an irrelevant factor are zero for all the assets, the risk premia in the second stage are no longer identified, and the entire inference is distorted. Kan and Zhang (1999b) show that even a small degree of model misspecification would be enough to inflate the useless factor t-statistic, creating an illusion of its pricing importance. Kleibergen (2009) further demonstrates that the presence of such factors has a drastic impact on the consistency and asymptotic distribution of the estimates even if the model is correctly specified and the true β's are zero only asymptotically ($\beta = \frac{B}{\sqrt{T}}$).

When the model is not identified, obtaining consistent parameter estimates is generally hard, if not impossible. There is, however, an extensive literature on inference, originating from the problem of weak instruments (see, e.g. Stock, Watson, and Yogo (2002)). Kleibergen (2009) develops identification-robust tests for the two-step procedure of Fama and MacBeth, and demonstrates how to build confidence bounds for the risk premia and test hypotheses of interest in the presence of spurious or weak factors. Unfortunately, the more severe is the identification problem, the less information can be extracted from the data. Therefore, it comes as no surprise that in many empirical applications robust confidence bounds can be unbounded at least from one side, and sometimes even coincide with the whole real line (as in the case of conditional Consumption-CAPM of Lettau and Ludvigson (2001b)), making it impossible to draw any conclusions either in favour of or against a particular hypothesis. In contrast, my approach consists in recovering a subset of parameters that are strongly identified from the data, resulting in their consistent, asymptotically normal estimates and usual confidence bounds. I prove that when the model is estimated by Pen-FM, standard bootstrap techniques can be used to construct valid confidence bounds for the strong factors risk premia even in the presence of useless factors. This is due to the fact that my penalty depends the nature of the second stage regressor (strong or useless), which remaines the same in bootstrap and allows the shrinkage term to eliminate the impact of the useless factors. As a result, bootstrap remains consistent and does not require additional modifications (e.g. Andrews and Guggenberger (2009), Chatterjee and Lahiri (2011)).

Using various types of penalty to modify the properties of the original estimation procedure has a long and celebrated history in econometrics, with my estimator belonging to the class of Least Absolute Selection and Shrinkage Operator (i.e. lasso, Tibshirani (1996))[1]. The structure of the penalty, however, is new, for it is designed not to choose significant parameters in the otherwise fully identified model, but rather select a subset of parameters that can be strongly identified and recovered from the data. The difference is subtle, but empirically rather striking. Simulations confirm that whereas Pen-FM successfully captures the distinction between strong and weak factors even for a very small sample size, the estimates produced, for instance, by the adaptive lasso (Zou (2006)), display an erratic behaviour[2].

The chapter also conributes to a recent strand of literature that examines the properties of conventional asset pricing estimation techniques. Lewellen, Nagel, and Shanken (2010) demonstrate that when a set of assets exhibits a strong factor structure, any variable correlated with those unobserved risk drivers may be indentified as a significant determinant

[1]Various versions of shrinkage techniques have been applied to a very wide class of models, related to variable selection, e.g. adaptive lasso (Zou (2006)) for variable selection in a linear model, bridge estimator for GMM (Caner (2009)), adaptive shrinkage for parameter and moment selection (Liao (2013)), or instrument selection (Caner and Fan (2014))

[2]This finding is expected, since the adaptive lasso, like all other similar estimators, requires identification of the original model parameters used either as part of the usual loss function, or the penalty imposed on it. Should this condition fail, the properties of the estimator will be substantially affected. This does not, however, undermine any results for the correctly identified model

of the cross-section of returns. They assume that model parameters are identified, and propose a number of remedies to the problem, such as increasing the asset span by including portfolios, constructed on other sorting mechanisms, or reporting alternative measures of fit and confidence bounds for them. These remedies, however, do not necessarily lead to better identification.

Burnside (2010) highlights the importance of using different SDF normalisations, their effect on the resulting identification conditions and their relation to the useless factor problem. He further suggests using the Kleibergen and Paap (2006) test for rank deficiency as a model selection tool. Gospodinov, Kan, and Robotti (2014a) also consider the SDF-based estimation of a potentially misspecified asset pricing model, contaminated by the presence of irrelevant factors. They propose a sequential elimination procedure that successfully identifies spurious factors and those that are not priced in the cross-section of returns, and eliminates them *simultaneously* from the candidate model. In contrast, the focus of my chapter is on the models with β-representation, which reflect the *partial pricing impact* of different risk factors[1]. Further, I use the simulation design from Gospodinov, Kan, and Robotti (2014a) to compare and contrast the finite sample performance of two approaches when the useless factors are assumed to have zero true covariance with asset returns. Pen-FM(GMM) seems to be less conservative by correctly preserving the strongly identified risk factors even in case of a relatively small sample size, when it is notoriously hard to reliably assess the pricing impact of the factor. This could be particularly important for empirical applications that use quarterly or yearly data, where the available sample is naturally quite small.

The rest of the chapter is organised as follows. I first discuss the structure of a linear factor model and summarise the consequences of identification failure established in the prior literature. Section 1.4 introduces Pen-FM and Pen-GMM estimators. I then discuss their asymptotic properties (Section 1.5) and simulation results (Section 2.8). Section 1.7 presents empirical applications, and Section 1.8 concludes.

[1] In addition, the two-step procedure could also be used in the applications that rely on the separate datasets used in the estimation of betas and risk premia. For example, Bandi and Tamoni (2015) and Boons and Tamoni (2014) estimate betas from long-horizon regressions and use them to price the cross-section of returns observed at a higher frequency, which would be impossible to do using a standard linear SDF-based approach

1.2 Linear factor model

I consider a standard linear factor framework for the cross-section of asset returns, where the risk premia for n portfolios are explained through their exposure to k factors, that is

$$\begin{aligned} \mathbb{E}\left[R_t^e\right] &= i_n\lambda_{0,c} + \beta_f\lambda_{0,f}, \\ cov(R_t^e, F_t) &= \beta_f var(F_t), \\ \mathbb{E}\left[F_t\right] &= \mu_f, \end{aligned} \tag{1.1}$$

where $t = 1...T$ is the time index of the observations, R_t^e is the $n \times 1$ vector of excess portfolios returns, F_t is the $k \times 1$ vector of factors, $\lambda_{0,c}$ is the intercept (zero-beta excess return), $\lambda_{0,f}$ is the $k \times 1$ vector of the risk premia on the factors, β_f is the $n \times k$ matrix of portfolio betas with respect to the factors, and μ_f is the $k \times 1$ vector of the factors means. Although many theoretical models imply that the common intercept should be equal to 0, it is often included in empirical applications to proxy the imperfect measurement of the risk-free rate, and hence is a common level factor in excess returns.

Model (1.1) can also be written equivalently as follows

$$\begin{aligned} R_t^e &= i_n\lambda_{0,c} + \beta_f\lambda_{0,f} + \beta_f v_t + u_t, \qquad (1.2) \\ F_t &= \mu_f + v_t, \end{aligned}$$

where u_t and v_t are $n \times 1$ and $k \times 1$ vectors of disturbances.

After demeaning the variables and eliminating μ_f, the model becomes:

$$\begin{aligned} R_t^e &= i_n\lambda_{0,c} + \beta_f(\overline{F}_t + \lambda_{0,f}) + \epsilon_t = \beta\lambda_0 + \beta_f\overline{F}_t + \epsilon_t, \qquad (1.3) \\ F_t &= \mu_f + v_t, \end{aligned}$$

where $\epsilon_t = u_t + \beta_f\bar{v}$, $\bar{v} = \frac{1}{T}\sum_{t=1}^{T} v_t$, $\overline{F}_t = F_t - \overline{F}$, $\overline{F} = \frac{1}{T}\sum_{t=1}^{T} F_t$, $\beta = (i_n \quad \beta_F)$ is a $n \times (k+1)$ matrix, stacking both the $n \times 1$ unit vector and asset betas, and $\lambda_0 = (\lambda_{0,c}, \lambda_{0,f}')'$ is a $(k+1) \times 1$ vector of the common intercept and risk premia parameters.

Assuming ϵ_t and v_t are asymptotically uncorrelated, our main focus is on estimating the parameters from the first equation in (1.3). A typical approach would be to use the Fama-MacBeth procedure, which decomposes the parameter estimation in two steps, focusing separately on time series and cross-sectional dimensions.

The first stage consists in time series regressions of excess returns on factors, to get the estimates of β_f:

$$\hat{\beta}_f = \sum_{t=1}^{T} \bar{R}_t^e \bar{F}_t' \left[\sum_{j=1}^{T} \bar{F}_j \bar{F}_j' \right]^{-1},$$

where $\hat{\beta}_f$ is an $n \times k$ matrix, $\overline{R}_t^e$ is a $n \times 1$ vector of demeaned asset returns, $\overline{R}_t^e = R_t^e - \frac{1}{T}\sum_{t=1}^{T} R_t^e$.

While the time series beta reveals how a particular factor correlates with the asset excess returns over time, it does not indicate whether this correlation is *priced* and could be used to explain the differences between required rates of return on various securities. The second stage of the Fama-MacBeth procedure aims to check whether asset holders demand a premium for being exposed to this source of risk ($\beta_j,\ j = 1..k$), and consists in using a single OLS or GLS cross-sectional regression of the average excess returns on their risk exposures.

$$\begin{aligned} \hat{\lambda}_{OLS} &= \left[\hat{\beta}'\hat{\beta}\right]^{-1} \hat{\beta}'\bar{R}^e, \\ \hat{\lambda}_{GLS} &= \left[\hat{\beta}'\hat{\Omega}^{-1}\hat{\beta}\right]^{-1} \hat{\beta}'\hat{\Omega}^{-1}\bar{R}^e, \end{aligned} \tag{1.4}$$

where $\hat{\beta} = [i_n \ \ \hat{\beta}_f]$ is the extended $n \times (k+1)$ matrix of $\hat{\beta}$'s, $\hat{\lambda} = [\hat{\lambda}_c \ \ \hat{\lambda}_f']'$ is a $(k+1) \times 1$ vector of the risk premia estimates, $\bar{R}^e = \frac{1}{T}\sum_{t=1}^{T} R_t^e$ is a $n \times 1$ vector of the average cross-sectional excess returns, and $\hat{\Omega}$ is a consistent estimate of the disturbance variance-covariance matrix, e.g. $\hat{\Omega} = \frac{1}{T-k-1}\sum_{t=1}^{T}(\bar{R}_t^e - \hat{\beta}_f t)(\bar{R}_t^e - \hat{\beta}_f \bar{F}_t)'$.

If the model is identified, that is, if the matrix of β has full rank, the Fama-MacBeth procedure delivers risk premia estimates that are consistent and asymptotically normal, allowing one to construct confidence bounds and test hypotheses of interest in the usual way (e.g. using t-statistics). In the presence of a useless or weak factor ($\beta_j = \mathbf{0}_n$ or more generally $\beta_j = \frac{B}{\sqrt{T}}$, where B is an $n \times 1$ vector), however, this condition is violated, thus leading to substantial distortions in parameter inference.

Although the problem of risk premia identification in the cross-section of assets is particularly clear when considering the case of the two-stage procedure, the same issue arises when trying to jointly estimate time series and cross-sectional parameters by GMM, using the following set of moment conditions:

$$
\begin{aligned}
\mathbb{E}\left[R_t^e - i_n\lambda_{0,c} - \beta_f(\lambda_{0,f} - \mu_f + F_t)\right] &= \mathbf{0}_n, \\
\mathbb{E}\left[(R_t^e - i_n\lambda_{0,c} - \beta_f(\lambda_{0,f} - \mu_f + F_t))F_t'\right] &= \mathbf{0}_{n\times k}, \\
\mathbb{E}[F_t - \mu_f] &= \mathbf{0}_k.
\end{aligned} \tag{1.5}
$$

Assuming the true values of model parameters $\theta_0 = \{vec(\beta_f); \lambda_{0,c}; \lambda_{0,f}; \mu_f\}$ belong to the interior of a compact set $S \in \mathbb{R}^{nk+k+k+1}$, one could then proceed to estimate them jointly by minimizing the following objective function:

$$
\hat{\theta} = \arg\min_{\theta \in S} \left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\right]' W_T(\theta) \left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\right], \tag{1.6}
$$

where $W_T(\theta)$ is a positive definite weight $(n + nk + k) \times (n + nk + k)$ matrix, and

$$
g_t(\theta) = \begin{bmatrix} R_t^e - i_n\lambda_c - \beta_f(\lambda_f - \mu + F_t) \\ vec\left(\left[R_t^e - i_n\lambda_c - \beta_f(\lambda_f - \mu + F_t)\right] F_t'\right) \\ F_t - \mu \end{bmatrix} \tag{1.7}
$$

is a sample moment of dimension $(n + nk + k) \times 1$.

In the presence of a useless factor the model is no longer identified, since the matrix of first derivatives $G(\theta_0) = E[G_t(\theta_0)] = E\left[\frac{dg_t(\theta_0)}{d\theta}\right]$ has a reduced column rank if at least one of the vectors in β_f is $\mathbf{0}_{n\times 1}$ or $\frac{B}{\sqrt{T}}$, making the estimates from eq.(1.6) generally inconsistent and having a nonstandard asymptotic distribution, since

$$
\frac{dg_t(\theta_0)}{d\theta'} = \begin{bmatrix} [\lambda_{0,f} - \mu_f + F_t]' \otimes I_n & -i_n & \beta_f & \beta_f \\ (F_t \otimes I_n)\left[(\lambda_{0,f} - \mu_f + F_t)' \otimes I_n\right] & -vec(i_n F_t') & -(F_t \otimes I_n)\beta_f & (F_t \otimes I_n)\beta_f \\ \mathbf{0}_{k\times nk} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} & -I_k \end{bmatrix}, \tag{1.8}
$$

where $\otimes$ denotes the Kronecker product and I_n is the identity matrix of size n. Note that the presence of useless factors affects only the risk premia parameters, since as long as the mean and the variance-covariance matrix of the factors are well-defined, the first moment conditions in eq. (1.5) would be satisfied for any λ_f as long as $\beta_f(\lambda_f - \lambda_{0,f}) = 0$. Therefore, identification problem relates only to the risk premia, but not the factor exposures, betas.

Throughout the paper, I consider the linear asset pricing framework, potentially contaminated by the presence of useless/weak factors, whether correctly specified or not. I call the model *correctly specified* if it includes all the true risk factors and eq.(1.3) holds. The model under estimation, however, could also include a useless/weak risk driver that is not priced in the cross-section of asset returns.

The model is called *misspecified* if eq.(1.3) does not hold. This could be caused by either omitting some of the risk factors necessary for explaining the cross-section of asset returns, or if the model is actually a non-linear one. The easiest way to model a misspecification would be to assume the true data-generating process including individual fixed effects for the securities in the cross-sectional equation:

$$\mathbb{E}\left[R_t^e\right] = \lambda_{0,i} + \beta_f \lambda_{0,f}$$

where $\lambda_{0,i}$ is a $n \times 1$ vector of individual intercepts. In the simulations I consider the case of a misspecified model, where the source of misspecification comes from the omitted risk factors. Therefore, it contaminates the estimation of both betas and risk premia.

1.3 Identification and what if it's not there

Depending on the nature of the particular identification failure and the rest of the model features, conventional risk premia estimators generally lose most of their properties: consistency, asymptotic normality, not to mention the validity of standard errors and confidence interval coverage for *all* the factors in the model. Further, numerical optimisation techniques may have convergence issues, faced with a relatively flat region of the objective function, leading to unstable point estimates.

Kan and Zhang (1999a) are the first to notice the problem generated by including a factor uncorrelated with asset returns in the GMM estimation framework of a linear stochastic discount factor model. They show that if the initial model is misspecified, the Wald test for the risk premia overrejects the null hypothesis of a factor having zero risk premium, and hence a researcher will probably conclude that it indeed explains the systematic differences in portfolio returns. The likelihood of finding significance in the impact of a useless factor increases with the number of test assets; hence, expanding the set of assets (e.g. combining 25 Fama-French with 19 industry portfolios) may even exacerbate the issue (Gospodinov, Kan, and Robotti (2014a)). Further, if the model is not identified, tests for model misspecification have relatively low power, thus making it even more difficult to detect the problem.

Gospodinov, Kan, and Robotti (2014a) consider a linear SDF model that includes both strong and useless factors, and the effect of misspecificaion-robust standard errors. Their estimator is based on minimizing the Hansen-Jagannathan distance (Hansen and Jagannathan (1997)) between the set of SDF pricing the cross-section of asset returns, and the ones implied by a given linear factor structure. This setting allows to construct misspecification-robust standard errors, because the value of the objective function can be used to assess the degree of model misspecification. They demonstrate that the risk premia estimates of the useless factors converge to a bounded random variable, and are inconsistent. Under correct model specificatation, strong factors risk premia estimates are consistent; however, they are no longer asymptotically normal. Further, if the model is misspecified, risk premia estimates for the strong factors are inconsistent and their pricing impact could be crowded out by the influence of the useless ones. Useless factors t-statistics, in turn, are inflated and asymptotically tend to infinity.

Kan and Zhang (1999b) study the properties of the Fama-MacBeth two-pass procedure with a single useless risk factor ($\beta = \mathbf{0}_n$), and demonstrate the same outcome. Thus, faced with a finite sample, a researcher is likely to conclude that such a factor explains the cross-sectional differences in asset returns. Kleibergen (2009) also considers the properties of the OLS/GLS two-pass procedure, if the model if weakly identified ($\beta = \frac{B}{\sqrt{T}}$). The paper proposes several statistics that are robust to identification failure and thus could be used to construct confidence sets for the risk premia parameters without pretesting.

Cross-sectional measures of fit are also influenced by the presence of irrelevant factors. Kan and Zhang (1999b) conjecture that in this case cross-sectional OLS-based R^2 tends to be substantially inflated, while its GLS counterpart appears to be less affected. This was later proved by Kleibergen and Zhan (2013), who derive the asymptotic distribution of R^2 and GLS-R^2 statistics and confirm that, although both are affected by the presence of useless factors, the OLS-based measure suffers substantially more. Gospodinov, Kan, and Robotti (2014b) consider cross-sectional measures of fit for the families of invariant (i.e. MLE, CUE-GMM, GLS) and non-invariant estimators in both SDF and beta-based frameworks and show that the invariant estimators and their fit are particularly affected by the presence of useless factors and model misspecification.

1.4 Pen-FM Estimator

Assuming the true values of risk premia parameters $\lambda_0 = (\lambda_{0,c}, \lambda_{0,F})$ lie in the interior of the compact parameter space $\Theta \in \mathbb{R}^k$, consider the following penalised version of the second stage in the Fama-MacBeth procedure:

$$\hat{\lambda}_{pen} = \arg\min_{\lambda \in \Theta} \left[\bar{R}^e - \hat{\beta}\lambda\right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda\right] + \eta_T \sum_{j=1}^{k} \frac{1}{||\hat{\beta}_j||_1^d} |\lambda_j|, \tag{1.9}$$

where $d > 0$ and $\eta_T > 0$ are tuning parameters, and $||\cdot||_1$ stands for the L_1 norm of the vector, $||\hat{\beta}_j||_1 = \sum_{i=1}^{n} |\hat{\beta}_{i,j}|$.

The objective function in Equation (1.9) is composed of two parts: the first term is the usual loss function, that typically delivers the OLS or GLS estimates of the risk premia parameters in the cross-sectional regression, depending on the type of the weight matrix, W_T. The second term introduces the penalty that is inversely proportional to the strength of the factors, and is used to eliminate the irrelevant ones from the model.

Equation (1.9) defines an estimator in the spirit of the lasso, Least Absolute Selection and Shrinkage Estimator of Tibshirani (1996) or the adaptive lasso of Zou (2006)[1]. The modification here, however, ensures that the driving force for the shrinkage term is not the value of the risk premium or its prior regression-based estimates (which are contaminated by the identification failure), but the nature of the betas. In particular, in the case of the adaptive lasso, the second stage estimates for the risk premia would have the penalty weights inversely proportional to their prior estimates:

$$\hat{\lambda}_{A.Lasso} = \arg\min_{\lambda \in \Theta} \left[\bar{R}^e - \hat{\beta}\lambda\right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda\right] + \eta_T \sum_{j=1}^{k} \frac{1}{|\hat{\lambda}_{j,ols}|^d} |\lambda_j|, \tag{1.10}$$

where $\hat{\lambda}_j$ is the OLS-based estimate of the factor j risk premium. Since these weights are derived from inconsistent estimates, with those for useless factors likely to be inflated under model misspecification, the adaptive lasso will no longer be able to correctly identify strong risk factors in the model. Simulations in Section 2.8 further confirm this distinction.

The reason for using the L_1 norm of the vector $\hat{\beta}_j$, however, is clear from the asymptotic

[1]Similar shrinkage-based estimators were later employed in various contexts of parameter estimation and variable selection. For a recent survey of the shrinkage-related techniques, see, e.g. Liao (2013).

behaviour of the latter:

$$vec(\hat{\beta}_j) = vec(\beta_j) + \frac{1}{\sqrt{T}} N\left(0, \Sigma_{\beta_j}\right) + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where $vec(\cdot)$ is the vectorisation operator, stacking the columns of a matrix into a single vector, $N\left(0, \Sigma_{\beta_j}\right)$ is the asymptotic distribution of the estimates of betas, a normal vector with mean 0 and variance-covariance matrix Σ_{β_j}, and $o_p\left(\frac{1}{\sqrt{T}}\right)$ contains the higher-order terms from the asymptotic expansion that do not influence the estimates $\sqrt{T}$ asymptotics. If a factor is strong, there is at least one portfolio that has true non-zero exposure to it; hence the L_1 norm of $\hat{\beta}$ converges to a positive number, different from 0 ($\left\|\hat{\beta}_j\right\|_1 = O_p(1)$). However, if a factor is useless and does not correlate with any of the portfolios in the cross-section, $\beta_j = \mathbf{0_{n\times 1}}$, therefore the L_1 norm of $\hat{\beta}$ converges to $\left\|\hat{\beta}_j\right\|_1 = O_p(\frac{1}{\sqrt{T}})$. This allows to clearly distinguish the estimation of their corresponding risk premia, imposing a higher penalty on the risk premium for a factor that has small absolute betas. Note that in the case of local-to-zero asymptotics in weak identification ($\beta_{sp} = \frac{1}{\sqrt{T}} B_{sp}$), again $\left\|\hat{\beta}_j\right\|_1 = O_p(\frac{1}{\sqrt{T}})$, the same penalty would be able to pick up its scale and shrink the risk premium at the second pass, eliminating its effect.

What is the driving mechanism for such an estimator? It is instructive to show its main features with an example of a single risk factor and no intercept at the second stage.

$$\begin{aligned}
\hat{\lambda}_{pen} &= \arg\min_{\lambda\in\Theta} \left[\bar{R}^e - \hat{\beta}\lambda\right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda\right] + \eta_T \frac{1}{||\hat{\beta}||_1^d}|\lambda| \\
&= \arg\min_{\lambda\in\Theta} \left[\lambda - \hat{\lambda}_{WLS}\right]' \hat{\beta}' W_T \hat{\beta}(\lambda - \hat{\lambda}_{WLS}) + \eta_T \frac{1}{||\hat{\beta}||_1^d}|\lambda|,
\end{aligned}$$

where $\lambda_{WLS} = \left[\hat{\beta}' W_T \hat{\beta}\right]^{-1} \hat{\beta}' W_T \bar{R}^e$ is the weighted least squares estimate of the risk premium (which corrresponds to either the OLS or GLS cross-sectional regressions).

The solution to this problem can easily be seen as a soft-thresholding function:

$$\hat{\lambda}_{pen} = sign\left(\hat{\lambda}_{WLS}\right)\left(|\hat{\lambda}_{WLS}| - \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d}\right)_+ \tag{1.11}$$

$$= \begin{cases} \hat{\lambda}_{WLS} - \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} & \text{if } \hat{\lambda}_{WLS} \geq 0 \text{ and } \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} < |\hat{\lambda}_{WLS}| \\ \hat{\lambda}_{WLS} + \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} & \text{if } \hat{\lambda}_{WLS} < 0 \text{ and } \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} < |\hat{\lambda}_{WLS}| \\ 0 & \text{if } \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} \geq |\hat{\lambda}_{WLS}| \end{cases}$$

Equation (1.11) illustrates the whole idea behind the modified lasso technique: if the penalty associated with the factor betas is high enough, the weight of the shrinkage term will asymptotically tend to infinity, setting the estimate directly to 0. At the same time, I set the tuning parameters (d and η_T) to such value that the threshold component does not affect either consistency or the asymptotic distribution for the strong factors (for more details, see Section 1.5).

If there is more than one regressor at the second stage, there is no analytical solution to the minimization problem of Pen-FM; however, it can be easily derived numerically through a sequence of 1-dimensional optimizations on the partial residuals, which are easy to solve. This is the so-called pathwise coordinate descent algorithm, where, at each point in time only one parameter estimate is updated. The algorithm goes as follows:

Step 1. Pick a factor $i \in [1..k]$ and write the overall objective function as

$$L = \left[\bar{R}^e - \hat{\beta}_i\lambda_i - \underset{j\neq i}{\hat{\beta}_j\widetilde{\lambda}_j}\right]' W_T \left[\bar{R}^e - \hat{\beta}_i\lambda_i - \underset{j\neq i}{\hat{\beta}_j\widetilde{\lambda}_j}\right] + \eta_T \left(\sum_{j=1, j\neq i}^{k} \frac{1}{\left\|\hat{\beta}_j\right\|_1^d}\left|\widetilde{\lambda}_j\right| + \frac{1}{\left\|\hat{\beta}_i\right\|_1^d}\left|\lambda_i\right|\right)$$

where all the values of λ_j, except for the one related to factor i, are fixed at certain levels $\widetilde{\lambda}_{j,j\neq i}$.

Step 2. Optimise L w.r.t λ_i. Note that this is a univariate lasso-style problem, where the residual pricing errors are explained only by the chosen factor i.

Step 3. Repeat the coordinate update for all the other components of λ.

Step 4. Repeat the procedure in Steps 1-3 until convergence is reached.

The convergence of the algorithm above to the actual solution of Pen-FM estimator problem follows from the general results of Tseng (1988, 2001), who studies the coordinate descent in a general framework. The only requirement for the algorithm to work is that the penalty function is convex and additively separable in the parameters, which is clearly satisfied in the case of Pen-FM. Pathwise-coordinate descent has the same level of computational complexity as OLS (or GLS), and therefore works very fast. It has been applied before to various types of shrinkage estimators, as in Friedman, Hastie, Hoffling, and Tibshirani (2007), and has been shown to be very efficient and numerically stable. It is also robust to potentially high correlations between the vectors of beta, since each iteration relies only on the residuals from the pricing errors.

As in the two-stage procedure, I define the shrinkage-based estimator for GMM (Pen-

GMM) as follows:

$$\hat{\theta}_{pen} = \arg\min_{\theta \in S} \left[\frac{1}{T} \sum_{t=1}^{T} g_t(\theta) \right]' W_T(\theta) \left[\frac{1}{T} \sum_{t=1}^{T} g_t(\theta) \right] + \eta_T \sum_{j=1}^{k} \frac{1}{||\hat{\beta}||_1^d} |\lambda_j|, \tag{1.12}$$

where S is a compact set in $\mathbb{R}^{nk+k+k+1}$.

The rationale for constructing such a penalty is the same as before, since one can use the properties of the $\hat{\beta}$s to automatically distinguish the strong factors from the weak ones on the basis of some prior estimates of the latter (OLS or GMM based).

It is important to note that the penalty proposed here does not necessarily need to be based on $||\hat{\beta}_j||_1$. In fact, the proofs can easily be modified to rely on any other variable that has the same asymptotic properties, i.e. being $O_p\left(\frac{1}{\sqrt{T}}\right)$ for the useless factors and $O_p(1)$ for the strong ones. Different scaled versions of the estimates of β, such as partial correlations or their Fischer transformation all share this property. Partial correlations, unlike betas, are invariant to linear transformation of the data, while Fisher transformation ($f(\hat{\rho}) = \frac{1}{2} ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right)$) provides a map of partial correlations from $[-1, 1]$ to $\mathbb{R}$.

1.5 Asymptotic Results

Similar to most of the related literature, I rely on the following high-level assumptions regarding the behaviour of the disturbance term ϵ_t:

Assumption 1 *(Kleibergen (2009)). As $T \to \infty$,*

1.

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\left(\begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes \left(R_t - i_n \lambda_{0,c} - \beta_f (\overline{f}_t + \lambda_{0,f}) \right) \right) \right] \xrightarrow{d} \begin{bmatrix} \varphi_R \\ \varphi_\beta \end{bmatrix}$$

where φ_R is $n \times 1$, φ_β is $nk \times 1$, where n is the number of portfolios and k is the number of factors. Further, $(\varphi_R', \varphi_\beta')' \sim N(0, V)$, where $V = Q \otimes \Omega$, and

$$\underset{(k+1)\times(k+1)}{Q} = \begin{pmatrix} 1 & \mu_f' \\ \mu_f & V_{ff} + \mu_f \mu_f' \end{pmatrix} = E\left[\begin{pmatrix} 1 \\ F_t \end{pmatrix} \begin{pmatrix} 1 \\ F_t \end{pmatrix}' \right], \quad \underset{n \times n}{\Omega} = var(\varepsilon_t), \quad \underset{k \times k}{V_{ff}} = var(F_t)$$

2.

$$\underset{T\to\infty}{plim} \frac{1}{T} \sum_{j=1}^{T} \bar{R}_j \bar{F}_j' = Q_{ff}, \quad \underset{T\to\infty}{plim}\, \bar{F} = \mu_f,$$

where Q_{ff} has full rank.

Assumption 1 provides the conditions required for the regression-based estimates of β_f to be easily computed using conventional methods, i.e. the data should conform to certain CLT and LLN, resulting in the standard $\sqrt{T}$ convergence. This assumption is not at all restrictive, and can be derived from various sets of low-level conditions, depending on the data generating process in mind for the behaviour of the disturbance term and its interaction with the factors, e.g. as in Shanken (1992) or Jagannathan and Wang (1998)[1].

Lemma 1.1 *Under Assumption 1, average cross-sectional returns and OLS estimator $\hat{\beta}$ have a joint large sample distribution:*

$$\sqrt{T}\begin{pmatrix} \bar{R}-\beta\lambda_f \\ vec(\hat{\beta}-\beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \psi_\beta \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega & 0 \\ 0 & V_{ff}^{-1}\otimes\Omega \end{pmatrix},\right]$$

where $\psi_R = \varphi_R$ is independent of $\psi_\beta = (V_{ff}^{-1}\otimes I_n)(\varphi_\beta - (\mu_f \otimes I_n)\varphi_R)$

Proof. See Kleibergen (2009), Lemma 1. ■

1.5.1 Correctly Specified Model

Having intuitively discussed the driving force behind the proposed shrinkage-based approach, I now turn to its asymptotic properties. The following propositions describe the estimator's behaviour in the presence of irrelevant factors: $\beta = (\beta_{ns}, \beta_{sp})$, where β_{ns} is an $n \times k_1$ matrix of the set of betas associated with k_1 non-spurious factors (including a unit vector) and β_{sp} denotes the matrix of the true value of betas for useless ($\beta_{sp} = \mathbf{0}_{n\times(k+1-k_1)}$) or weak ($\beta_{sp} = \frac{B_{sp}}{\sqrt{T}}$) factors.

Proposition 1.1 *Under Assumption 1, if $W_T \xrightarrow{p} W$, W is a positive definite $n \times n$ matrix, $\eta_T = \eta T^{-d/2}$ with a finite constant $\eta > 0$, $d > 0$ and $\beta_{ns}'\beta_{ns}$ having full rank, $\hat{\lambda}_{ns} \xrightarrow{p} \lambda_{0,ns}$ and $\hat{\lambda}_{sp} \xrightarrow{p} 0$*

[1] For example, Shanken (1992) uses the following assumptions, which easily result in Assumption 1:

1. The vector ϵ_t is independently and identically distributed over time, conditional on (the time series values for) F, with $E[\epsilon_t|F] = 0$ and $Var(\epsilon_t|F) = \Omega$ (rank N)
2. F_t is generated by a stationary process such that the first and second sample moments converge in probability, as $T \to \infty$ to the true moments which are finite. Also, $\bar{F}$ is asymptotically normally distributed.

Jagannathan and Wang (1998) provide low level conditions for a process with conditional heteroscedasticity.

Further, if $d > 2$

$$\sqrt{T}\begin{pmatrix}\hat{\lambda}_{ns}-\lambda_{0,ns}\\ \hat{\lambda}_{sp}\end{pmatrix} \xrightarrow{d} \begin{pmatrix}\left[\beta_{ns}' W\beta_{ns}\right]^{-1}\beta_{ns}' W\Psi_{\beta,ns}\lambda_{0,ns}+\left(\beta_{ns}' W\beta_{ns}\right)^{-1}\beta_{ns}' W\psi_R\\ 0\end{pmatrix}$$

Proof. See Appendix B.1. ■

The intuition behind the proof for consistency is clear: the tuning parameter η_T is set in such a way that the overall effect of the penalty, η_T, disappears with the sample size, and therefore does not affect the consistency of the parameter estimation, unless some of its shrinkage components are inflated by the presence of irrelevant factors. If a factor is useless, the L_1 norm of $\hat{\beta}_j$ tends to 0 at the $\sqrt{T}$ rate, and the penalty converges to a positive constant in front of the corresponding $|\lambda_j|$. Further, since $\hat{\beta}_j \to \mathbf{0_{n\times 1}}$, λ_j disappears from the usual loss function, $\left[\bar{R}^e-\hat{\beta}\lambda\right]' W_T\left[\bar{R}^e-\hat{\beta}\lambda\right]$, and it is the penalty component that determines its asymptotic behaviour, shrinking the estimate towards 0. At the same time, other parameter estimates are not affected, and their behaviour is fully described by standard arguments.

The shrinkage-based second pass estimator has the so-called *oracle* property for the non-spurious factors: the estimates of their risk premia have the same asymptotic distribution as if we had not included the useless factors in advance. Risk premia estimates are asymptotically normal, with two driving sources of the error component: estimation error from the first pass β's (and the resulting error-in-variables problem), and the disturbance term effect from the second pass.

The risk premia for the useless factors are driven towards 0 even at the level of the asymptotic distribution to ensure that they do not affect the estimation of other parameters. It should be emphasized, that the effect of the penalty does not depend on the actual value of the risk premium. Unlike the usual lasso or related procedures, the mechanism of the shrinkage here is driven by the strength of $\hat{\beta}$, regressors in the second pass. Therefore, there is no parameter discontinuity in the vicinity of 0, and bootstrap methods can be applied to approximate the distribution and build the confidence bounds.

One could argue that the assumption of $\beta = 0$ is quite restrictive, and a more realistic approximation of local-to-zero asymptotics should be used. Following the literature on weak instruments, I model this situation by assuming that $\beta_{sp} = \frac{B_{sp}}{\sqrt{T}}$. This situation could arise when a factor has some finite-sample correlation with the assets that eventually disappears asymptotically. As with the case of useless factors, I present the asymptotic results properties of the Pen-FM estimator, when there are weak factors in the model.

Proposition 1.2 *Under Assumption 1, if* $\beta_{sp} = \frac{B_{sp}}{\sqrt{T}}$, $W_T \xrightarrow{p} W$, W *is a positive definite* $n \times n$ *matrix,* $\eta_T = \eta T^{-d/2}$ *with a finite constant* $\eta > 0$, $d > 0$ *and* $\beta'_{ns}\beta_{ns}$ *having full rank,* $\hat{\lambda}_{ns} \xrightarrow{p} \lambda_{0,ns}$ *and* $\hat{\lambda}_{sp} \xrightarrow{p} 0$

Further, if $d > 2$

$$\sqrt{T}\begin{pmatrix} \hat{\lambda}_{ns} - \lambda_{0,ns} \\ \hat{\lambda}_{sp} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (\beta'_{ns}W^{-1}\beta_{ns}]^{-1}\beta'_{ns}W^{-1}B_{sp}\lambda_{0,sp} + [\beta'_{ns}W^{-1}\beta_{ns})^{-1}\beta'_{ns}W^{-1}(\psi_R + \Psi_{\beta,ns}\lambda_{0,ns}) \\ 0 \end{pmatrix}$$

Proof. See Appendix B.2. ■

The logic behind the proof is exactly the same as in the previous case. Recall that even in the case of weak identification again $\left\|\hat{\beta}_j\right\|_1 = O_p(\frac{1}{\sqrt{T}})$. Therefore, the penalty function recognises its impact, shrinking the corresponding risk premia towards 0, while leaving the other parameters intact.

The situation with weak factors is slightly different from that with purely irrelevant ones. While excluding such factors does not influence consistency of the strong factors risk premia estimates, it affects their asymptotic distribution, as their influence does not disappear fast enough (it is of the rate $\frac{1}{\sqrt{T}}$, the same as the asymptotic convergence rate), and hence we get an asymptotic bias apart from the usual components of the distribution. Note, that any procedure eliminating the impact of weak factors from the model (e.g. Gospodinov, Kan, and Robotti (2014a), Burnside (2010)), results in the same effect. In small sample it could influence the risk premia estimates; however, the size of this effect depends on several factors, and in general is likely to be quite small, especially compared to the usual error component.

Note that the $\frac{1}{\sqrt{T}}$ bias arises only if the omitted risk premium is non-zero. This requires a factor that asymptotically is not related to the cross-section of returns, but is nevertheless priced. Though unlikely, one cannot rule out such a case ex ante. If the factor is tradable, the risk premium on it should be equal to the corresponding excess return; hence one can use this property to recover a reliable estimate of the risk premium, and argue about the possible size of the bias or try to correct for it.

1.5.2 Misspecified Model

Model misspecification severely exacerbates many consequences of the identification failure[1]; however, its particular influence depends on the degree and nature of such misspecification.

The easiest case to consider is mean-misspecification, when factor betas are properly estimated, but the residual average returns on the second stage are non-zero. One might draw an analogy here with panel data, where the presence of individual fixed effects would imply that the pooled OLS regression is no longer applicable. The case of mean-misspecification is also easy to analyse, because it allows us to isolate the issue of the correct estimation of β from the one of recovering the factor risk premia. For example, one can model the return generation process as follows:

$$\bar{R} = c + \beta\lambda_0 + \frac{1}{\sqrt{T}}\psi_R + o_p\left(\frac{1}{\sqrt{T}}\right),$$
$$vec(\hat{\beta}) = vec(\beta) + \frac{1}{\sqrt{T}}\psi_\beta + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where c is a $n \times 1$ vector of the constants. It is well known that both OLS and GLS, applied to the second pass, result in diverging estimates for the spurious factors risk premia and t-statistics asymptotically tending to infinity. Simulations confirm the poor coverage of the standard confidence intervals and the fact that the spurious factor is often found to be significant even in relatively small samples. However, the shrinkage-based second pass I propose successfully recognises the spurious nature of the factor. Since the first-pass estimates of β's are consistent and asymptotically normal, the penalty term behaves in the same way as in the correctly specified model, shrinking the risk premia for spurious factors to 0 and estimating the remaining parameters as if the spurious factor had been omitted from the model. Of course, since the initial model is misspecified to begin with, risk premia estimates would suffer from inconsistency, but it would not stem from the lack of model identification.

A more general case of model misspecification would involve an omitted variable bias (or the nonlinear nature of the factor effects). This would in general lead to the inconsistent estimates of betas (e.g. if the included factors are correlated with the omitted ones), invalidating the inference in both stages of the estimation. However, as long as the problem of

[1] See, e.g. Kan and Zhang (1999a), Jagannathan and Wang (1998), Kleibergen (2009) and Gospodinov, Kan, and Robotti (2014a)

rank deficiency caused by the useless factors remains, the asymptotic distribution of Pen-FM estimator will continue to share that of the standard Fama-MacBeth regressions without the impact of spurious factors. A similar result can easily be demonstrated for Pen-GMM.

1.5.3 Bootstrap

While the asymptotic distribution gives a valid description of the pointwise convergence, a different procedure is required to construct valid confidence bounds. Although traditional shrinkage-based estimators are often used in conjunction with bootstrap techniques, it has been demonstrated that even in the simplest case of a linear regression with independent factors and i.i.d. disturbances, such inferences will be invalid (Chatterjee and Lahiri (2010)). Intuitively this happens because the classical lasso-related estimators incorporate the penalty function, which behaviour depends on the true parameter values (in particular, whether they are 0 or not). This in turn requires the bootstrap analogue to correctly identify the sign of parameters in the ε-neighborhood of zero, which is quite difficult. Some modifications to the residual bootstrap scheme have been proposed to deal with this feature of the lasso estimator (Chatterjee and Lahiri (2011, 2013)).

Fortunately, the problem explained above is not relevant for the estimator that I propose, because the driving force of the penalty function comes only from the nature of the regressors, and hence there is no discontinuity, depending on the true value of the risk premium. Further, in the baseline scenario I work with a 2-step procedure, where shrinkage is used only in the second stage, leaving the time series estimates of betas and average returns unchanged. All of the asymptotic properties discussed in the previous section result from the first order asymptotic expansions of the time series regressions. Therefore, it can be demonstrated that once a consistent bootstrap procedure for time series regressions is established (be it pairwise bootstrap, blocked or any other technique appropriate to the data generating process in mind), one can easily modify the second stage so that the bootstrap risk premia have proper asymptotic distributions.

Consider any bootstrap procedure (pairwise, residual or block bootstrap) that remains consistent for the first stage estimates, that is

$$\hat{\beta}^* = \hat{\beta} + \frac{1}{\sqrt{T}}\Psi_\beta + o_p\left(\frac{1}{\sqrt{T}}\right)$$
$$\bar{R}^{e*} = \bar{R}^e + \frac{1}{\sqrt{T}}\Psi_R + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where $\hat{\beta}^*$ and $\bar{R}^*$ are the the bootstrap analogues of $\hat{\beta}$ and $\bar{R}$.

Then

$$\hat{\lambda}^*_{pen} = \arg\min_{\lambda\in\Theta} \left[\bar{R}^* - \hat{\beta}^*\lambda\right]' W_T \left[\bar{R}^* - \hat{\beta}^*\lambda\right] + \eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}^*_j\right\|_1^d} |\lambda_j| \tag{1.13}$$

is the bootstrap analogue of $\hat{\lambda}_{pen}$.

Let $\hat{H}_n(\cdot)$ denote the conditional cdf of the bootstrap version $B^*_T = \sqrt{T}(\hat{\lambda}^*_{pen} - \hat{\lambda}_{pen})$ of the centred and scaled Pen-FM estimator of the risk premia $B_T = \sqrt{T}(\hat{\lambda}_{pen} - \lambda_0)$.

Proposition 1.3 *Under conditions of Proposition 1.1,*

$$\rho(\hat{H}^*_T, \hat{H}_T) \to 0, \quad as\ T \to \infty,$$

where $\hat{H}_T = P(B_T \leq x), x \in R$ and ρ denotes weak convergence in distribution on the set of all probability measures on $(\mathbb{R}^{(k+1)}, B(\mathbb{R}^{(k+1)}))$

Proof. See Appendix B.3 ■

Proposition 1.3 implies that the bootstrap analogue of Pen-FM can be used as an approximation for the distribution of the risk premia estimates. This result is similar to the properties of the adaptive lasso, that naturally incorporates soft thresholding with regard to the optimisation solution, and unlike the usual lasso of Tibshirani (1996), does not require aditional corrections (e.g. Chatterjee and Lahiri (2010)).

Let $b_T(\alpha)$ denote the α-quantile of $||B_T||$, $\alpha \in (0,1)$. Define

$$I_{T,\alpha} = b \in R^k : ||b - \hat{\lambda}_{pen}|| \leq T^{-1/2} b_T(\alpha)$$

the level-α confidence set for λ.

Proposition 1.4 *Let $\alpha \in (0,1)$ be such that $P(||B|| \leq t(\alpha) + \nu) > \alpha$ for all $\nu > 0$. Then under the conditions of Proposition 1.1*

$$P(\lambda_0 \in I_{T,\alpha}) \to \alpha \quad as\ T \to \infty$$

This holds if there is at least 1 non-spurious factor, or an intercept in the second stage.

Proof. See Appendix B.4 ■

In other words, the above proposition states that having a sample of bootstrap analogues for $\hat{\lambda}_{pen}$, one can construct valid percentile-based confidence bounds for strongly identified parameters.

1.5.4 Generalised Method of Moments

One can modify the objective function in Equation (1.6) to include a penalty based on the initial OLS estimates of the β_F parameters. Similar to the two-step procedure, this would shrink the risk premia coefficients for the spurious factors to 0, while providing consistent estimates for all the other parameters in the model.

The following set of assumptions provides quite general high level conditions for deriving the asymptotic properties of the estimator in the GMM case.

Assumption 2 1. *For all* $1 \leq t \leq T$, $T \geq 1$ *and* $\theta \in S$

a) $g_t(\theta)$ *is m-dependent*

b) $|g_t(\theta_1) - g_t(\theta_2)| \leq M_t|\theta_1 - \theta_2|$,

with $lim_{T\to\infty} \sum_{t-1}^{T} EM_t^p < \infty$, *for some* $p > 2$;

c) $\sup_{\theta \in S} E|g_t(\theta)|^p < \infty$, *for some* $p > 2$

2. *Define* $E\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) = g_{1T}(\theta)$

   a) Assume that $g_{1T}(\theta) \to g_1(\theta)$ *uniformly over* S, *and* $g_{1T}(\theta)$ *is continuously differentiable in* θ;

   b) $g_1(\theta_{0,ns}, \lambda_{sp} = \mathbf{0}_{k_2}) = 0$, *and* $g_1(\theta_{ns}, \lambda_{sp} = \mathbf{0}_{k_2}) \neq 0$ *for* $\theta_{ns} \neq \theta_{0,ns}$, *where* $\theta_{ns} = \{\mu, vec(\beta), \lambda_{f,ns}, \lambda_c\}$

3. *Define the following* $(n+nk+k)\times(nk+k+1+k)$ *matrix:* $G_T(\theta) = \frac{dg_{1T}(\theta)}{d\theta'}$. *Assume that* $G_T(\theta) \xrightarrow{p} G(\theta)$ *uniformly in a neighbourhood* N *of* $(\theta_{0,ns}, \lambda_{sp} = \mathbf{0}_{k_2})$, $G(\theta)$ *is continuous in theta.* $G_{ns}(\theta_{ns,0}, \lambda_{sp} = \mathbf{0}_{k_2})$ *is an* $(n+nk+k)\times(nk+k_1+k)$ *submatrix of* $G(\theta_0)$ *and has full column rank.*

4. $W_T(\theta)$ *is a positive definite matrix,* $W_T(\theta) \xrightarrow{p} W(\theta)$ *uniformly in* $\theta \in S$, *where* $W(\theta)$ *is an* $(n+nk+k)\times(n+nk+k)$ *symmetric nonrandom matrix, which is continuous in* θ *and is positive definite for all* $\theta \in S$.

The set of assumptions is fairly standard for the GMM literature and stems from the reliance on the empirical process theory, often used to establish the behaviour of the shrinkage-based GMM estimators (e.g. Caner (2009), Liao (2013)). Most of these assumptions could be further substantially simplified (or trivially established) following the structure of the linear factor model and the moment function for the estimation. However, it is instructive to present a fairly general case. Several comments are in order, however.

Assumption 2.1 presents a widespread sufficient condition for using empirical process arguments, and is very easy to establish for a linear class of models (it also encompasses a relatively large class of processes, including the weak time dependence of the time series and potential heteroscedasticity). For instance, the primary conditions for the two-stage estimation procedure in Shanken (1992) easily satisfy these requirements.

Assumptions 2.2 and 2.3, among other things, provide the identification condition used for the moment function and its parameters. I require the presence of k_2 irrelevant/spurious factors to be the only source for the identification failure, which, once eliminated, should not affect any other parameter estimation. One of the direct consequences is that the first-stage OLS estimates of the betas ($\hat{\beta}$) have a standard asymtotic normal distribution and basically follow the same speed of convergence as in the Fama-McBeth procedure, allowing us to rely on them in formulating the appropriate penalty function.

The following proposition establishes the consistency and asymptotic normality of Pen-GMM:

Proposition 1.5 *Under Assumption 2, if $\beta_{sp} = \mathbf{0}_{n\times k_2}$, $\eta_T = \eta T^{-d/2}$ with a finite constant $\eta > 0$, and $d > 2$, then*

$\hat{\lambda}_{sp} \xrightarrow{p} \mathbf{0}_{k_2}$ *and* $\hat{\theta}_{ns} \xrightarrow{p} \theta_{0,ns}$

Further, if $d > 2$

$$\sqrt{T}(\hat{\lambda}_{pen,sp}) \xrightarrow{d} \mathbf{0}_{k_2}$$
$$\sqrt{T}(\hat{\theta}_{pen,ns} - \theta_{0,ns}) \xrightarrow{d} [G_{ns}(\theta_0)'W(\theta_0)G_{ns}(\theta_0)]^{-1}G_{ns}(\theta_0)W(\theta_0)Z(\theta_0)$$

where $\theta_{ns} = \{\mu, vec(\beta,), \lambda_{f,ns}, \lambda_c\}$, $Z(\theta_0) \equiv N(0, \Gamma(\theta_0))$, *and*
$\Gamma(\theta_0) = lim_{T\to\infty} E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0)\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0)\right]'$

Proof. See Appendix B.5. ■

The intuition behind these results is similar to the two-pass procedure of Fama-MacBeth: the penalty function is formulated in such a way as to capture the effect of factors with extremely weak correlation with asset returns. Not only does the resulting estimator retain consistency, but it also has an asymptotically normal distribution. Bootstrap consistency for constructing confidence bounds could be proved using an argument, similar to the one outlined for the Pen-FM estimator in Propositions 1.3 and 1.4.

1.6 Simulations

Since many empirical applications are characterised by a rather small time sample of available data (e.g. when using yearly observations), it is particularly important to assess the finite sample performance of the estimator I propose. In this section I discuss the small-sample behaviour of the Pen-FM estimator, based on the simulations for the following sample sizes: $T = 30,\ 50,\ 100,\ 250,\ 500,\ 1000$.

For a *correctly specified model* I generate normally distributed returns for 25 portfolios from a one-factor model, CAPM. In order to get factor loadings and other parameters for the data-generating process, I estimate the CAPM on the cross-section of excess returns on 25 Fama-French portfolios sorted on size and book-to-market, using quarterly data from 1947Q2 to 2014Q2 and market excess return, measured by the value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ. The data is taken from Kenneth French website. I then run separate time series regressions of these portfolios excess returns on R^e_{mkt} to get the estimates of market betas, $\hat{\beta}$ (25×1), and the variance-covariance matrix of residuals, $\hat{\Sigma}$ (25×25). I then run a cross-sectional regression of the average excess returns on the factor loadings to get $\hat{\lambda}_0$ and $\hat{\lambda}_1$.

The true factor is simulated from a normal distribution with the empirical mean and variance of the market excess return. A spurious factor is simulated from a normal distribution with the mean and variance of the real per capita nondurable consumption growth, constructed for the same time period using data from NIPA Table 7.1 and the corresponding PCE deflator. It is independent of all the other innovations in the model. Finally, returns are generated from the following equation:

$$R^e_t = \hat{\lambda}_0 + \hat{\beta}'\hat{\lambda}_1 + \beta'\bar{R}^e_{t,mkt} + \epsilon_t$$

where ϵ_t is generated from a multivariate normal distribution $N\left(0, \hat{\Sigma}\right)$.

I then compare the performance of 3 estimators: (a) Fama-MacBeth, using the simulated market return as the only factor (I call this the *oracle* estimator, since it includes only the true risk factor *ex ante*), (b) Fama-MacBeth, using the simulated market return and the irrelevant factor, (c) Pen-FM estimator, using the simulated market return and the irrelevant factor.

For a *misspecified model* the data is generated from a 3 factor model, based on 3 canonical Fama-French factors (with parameters obtained and data generated as in the procedure outlined above). However, in the simulations I consider estimating a 1 factor model (thus, the source of misspecification is omitting the SMB and HML factors). Again, I compare the performance of 3 estimators: (a) Fama-MacBeth, using the simulated market return as the only factor, (b) Fama-MacBeth, using the simulated market return and the irrelevant factor, (c) Pen-FM estimator, using the simulated market return and the irrelevant factor.

For each of the simulations, I also compute conventional measures of fit:

$$R^2_{ols} = 1 - \frac{var\left(\bar{R}^e - \hat{\lambda}_{ols}\hat{\beta}\right)}{var(\bar{R}^e)}, \qquad HJ = \sqrt{\hat{\lambda}'_{ols}\left(\sum_{t=1}^{T} R_t R_t'\right)\hat{\lambda}_{ols}},$$

$$R^2_{gls,1} = 1 - \frac{var\left(\hat{\Omega}^{-1/2}(\bar{R}^e - \hat{\lambda}_{ols}\hat{\beta})\right)}{var(\hat{\Omega}^{-1/2}\bar{R}^e)}, T^2 = \alpha'\left(\frac{(1+\hat{\lambda}'_f\Sigma_f\hat{\lambda}_f)q}{T}\right)^+\alpha,$$

$$R^2_{gls,2} = 1 - \frac{var\left(\hat{\Omega}^{-1/2}(\bar{R}^e - \hat{\lambda}_{gls}\hat{\beta})\right)}{var(\hat{\Omega}^{-1/2}\bar{R}^e)}, q = \alpha'(y\hat{\Omega}y')^+\alpha$$

$$APE = \frac{1}{n}\sum_{i=1}^{n}|\alpha_i|,$$

where R^2_{ols} is the cross-sectional OLS-based R^2, $R^2_{gls,1}$ is the GLS-R^2, based on the OLS-type estimates of the risk premia, $\hat{\Omega}$ is the sample variance-covariance matrix of returns, $R^2_{gls,2}$ is the GLS-R^2, based on the GLS-type estimates of the risk premia $\alpha_i = \bar{R}_i^e - \hat{\lambda}_{ols}\hat{\beta}_i$ is the average time series pricing error for portfolio i, HJ is the Hansen-Jagannathan distance, $^+$ stands for the pseudo-inverse of a matrix, $y = I - \hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}$, T^2 is the cross-sectional test of Shanken (1985), Σ_f is the variance-covariance matrix of the factors, and $\hat{\lambda}_f$ is a $k \times 1$ vector of the factors risk premia (excluding the common intercept).

For the Pen approach, I use the penalty, defined through partial correlations of the factors and returns (since they are invariant to the linear transformation of the data). I set the level tuning parameter, η to $\bar{\sigma}$, the average standard deviation of the residials from the first stage, and the curvature parameter, d, to 4. In Section 1.6.3, I investigate the impact of tuning parameters on the estimator performance, and show that changing values of the tuning parameters has only little effect on the estimator's ability to eliminate or retain strong/weak factors.

1.6.1 Correctly Specified Model

Table 1.1 demonstrates the performance of the three estimation techniques in terms of their point estimates: the Fama-MacBeth two-pass procedure without the useless factor (denoted as the *oracle* estimator), the Fama-MacBeth estimator, which includes both useful and useless factors in the model and the Pen-FM estimator. All three us an identity weight matrix at the second stage. For each of the estimators the table reports the mean point estimate of the risk premia and the intercept, their bias and mean squared error. I also report in the last column the average factor shrinkage rates for the Pen-FM estimator, produced using 10,000 simulations (i.e. how often the corresponding risk premia estimate is set exactly to 0).

The results are striking. The useless factor is correctly identified in the model with the correponding risk premia shrunk to 0 with 100% accuracy even for such a small sample size as 30 observations. At the same time, the useful factor (market excess return) is correctly preserved in the specification, with the shrinkage rate below 1% for all the sample sizes. Starting from $T = 50$, the finite sample bias of the parameter estimates produced by the Pen-FM estimator is much closer to that of the oracle Fama-MacBeth cross-sectional regression, which exludes the useless factor ex ante. For example, when $T = 50$, the average finite sample bias of the useful factor risk premium, produced by the oracle Fama-MacBeth estimator is 0.093 %, 0.114 % for the two-step procedure which includes the useless factor, and 0.091% for the estimates produced by Pen-FM.

The mean squared errors of the estimates demonstrate a similar pattern: for $T \geq 50$ the MSE for Pen-FM is virtually identical to that of the Fama-MacBeth without the useless factor in the model. At the same time, the mean squared error for the standard Fama-MacBeth estimator stays at the same level of about 0.32% regardless of sample size, illustrating the fact that the risk premia estimate of the useless factor is inconsistent, converging to a bounded random variable, centred at 0.

The size of the confidence intervals constructed by bootstrap is slighly conservative (see Table 1.A.1). However, it is not a feature particular to the Pen-FM estimator. Even without the presence of useless factors in the model, bootstrapping risk premia parameters seems to produce similar slighly conservative confidence bounds, as illustrated in Table 1.A.1, Panel A.

Figure 1.A.1-1.A.5 also illustrate the ability of Pen-FM estimator to restore the original quality of fit for the model. Figure 1.A.1 shows the distribution of the cross-sectional R^2 for the various sample size. The measures of fit, produced by the model in the absence of the useless factor and with it, when estimated by Pen-FM, are virtually identical. At the same time, $\bar{R}^2$, produced by the conventional Fama-MacBeth approach seems to be inflated by the presence of a useless factor, consistent with the theoretical findings in Kleibergen and Zhan (2013). The distribution of the in-sample measure of fit seems to be quite wide (e.g. for T=100 it fluctuates a good deal from 0 to 80%), again highlighting the inaccuracy of a single point estimate and a need to construct confidence bounds for the measures of fit (e.g. as suggested in Lewellen, Nagel, and Shanken (2010). Even if we estimate the true model specification, empirically the data contains quite a lot of noise (which was also captured in the simulation design, calibrating data generating parameters to their sample analogues). Thus it is not surprising to find that the probability of getting a rather low value of the R^2 is still high for a moderate sample size. Only when the number of observations is high (e.g. T=1000), does the peak of the probability density function seem to approach 80%; however, even then the domain remains quite wide.

Table 1.1: Estimates of risk premia in a correctly specified model

	True parameter value	Mean Estimate			Bias			MSE			Shrinkage rate
		Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	
					Panel A: T=30						
Intercept	3.277	3.265	3.273	3.25	-0.012	-0.004	-0.027	2.259	2.26	2.203	0
Useful factor	-0.647	-0.673	-0.672	-0.659	-0.026	-0.026	-0.012	2.25	2.247	2.196	0.007
Useless factor	-	-	0.002	0	-	0.002	0	-	0.317	0	1
					Panel B: T=50						
Intercept	3.277	3.17	3.149	3.173	-0.107	-0.128	-0.104	1.877	1.837	1.848	0
Useful factor	-0.647	-0.553	-0.533	-0.556	0.093	0.114	0.091	1.875	1.827	1.848	0.009
Useless factor	-	-	0.01	0	-	0.01	0	-	0.314	0	1
					Panel C: T=100						
Intercept	3.277	3.213	3.195	3.21	-0.064	-0.082	-0.067	1.449	1.447	1.444	0
Useful factor	-0.647	-0.593	-0.575	-0.591	0.054	0.072	0.056	1.421	1.427	1.417	0.003
Useless factor	-	-	0.01	0	-	0.01	0	-	0.318	0	1
					Panel D: T=250						
Intercept	3.277	3.267	3.271	3.266	-0.011	-0.007	-0.011	0.902	0.894	0.901	0
Useful factor	-0.647	-0.642	-0.648	-0.642	0.005	-0.001	0.005	0.887	0.885	0.886	0
Useless factor	-	-	-0.002	0	-	-0.002	0	-	0.325	0	1
					Panel E: T=500						
Intercept	3.277	3.277	3.281	3.276	-0.001	0.004	-0.001	0.628	0.647	0.628	0
Useful factor	-0.647	-0.645	-0.65	-0.645	0.001	-0.003	0.002	0.627	0.646	0.627	0
Useless factor	-	-	0.007	0	-	0.007	0	-	0.31	0	1
					Panel F: T=1000						
Intercept	3.277	3.286	3.278	3.286	0.009	0	0.009	0.435	0.441	0.435	0
Useful factor	-0.647	-0.655	-0.646	-0.654	-0.008	0.001	-0.008	0.421	0.431	0.421	0
Useless factor		-	-0.012	0	-	-0.012	0	-	0.321	0	1

Note. The table summarises the properties of the Fama-MacBeth and Pen-FM estimators with an identity weight matrix in a model for 25 portfolios with a common intercept and one true factor driving the returns. λ_0 is the value of the intercept, λ_1 and λ_2 are the corresponding risk premia of the true risk factor and the useless one. The model is simulated 10 000 times for different values of the sample size (T). The "Oracle" estimator corresponds to the Fama-MacBeth procedure omitting the useless factor, "FM" and "Pen-FM" stand for the Fama-MacBeth and Pen-FM estimators in the model with a useful and a useless factor. The table presents the mean point estimates of the parameters, their bias, and the mean squared error (MSE). The mean shrinkage rate corresponds to the average percentage of times the corresponding coefficient was set to exactly 0 during 10,000 simulations.

Returns are generated from the multivariate normal distribution with the mean and variance-covariance matrix equal to those of the nominal quarterly excess returns on 25 Fama-French portfolios sorted by size and book-to-market ratio during the period 1962Q2 : 2014Q2. The useful factor drives the cross-section of asset returns, and is calibrated to have the same mean and variance as the quarterly excess return on the market. The useless factor is generated from a multivariate normal distribution with the mean and variance equal to their sample analogues of nondurable consumption growth for the same time period. Betas, common intercept and risk premium for the useful factor come from the Fama-MacBeth estimates of a one factor model with market excess return estimated on the cross-section of the 25 Fama-French portfolios.

The *GLS* R^2, based on either OLS or GLS second stage estimates (Figure 1.A.2 and 1.A.3), seem to have a much tighter spread (in particular, if one relies on the OLS second stage). As the sample size increases, the measures of fit seem to better indicate the pricing ability of the true factor. The *GLS* R^2 is less affected by the problem of the useless factor (as demonstrated in Kleibergen and Zhan (2013)), but there is still a difference between the estimates, and if the model is not identified, R^2 seems to be slighly higher, as in the OLS case. This effect, however, is much less pronounced. Once again, the distribution of *GLS* R^2 for Pen-FM is virtually identical to that of the conventional Fama-MacBeth estimator without the useless factor in the model. A similar spurious increase in the quality of fit may be noted, considering the distribution of the average pricing errors (Figure 1.A.5), which is shifted to the left in the presence of a useless factor. The Hansen-Jagannathan distance is also affected by the presence of the useless factor (as demonstrated in Gospodinov, Kan, and Robotti (2014a)); however, not as much (Figure 1.A.4). In contrast to the standard Fama-McBeth estimator, even for a very small sample size the average pricing error and the Hansen-Jagannathan distance produced by Pen-FM are virtually identical to those of the model that does not include the spurious factor ex ante.

Figs. 1.A.11 and 1.A.13 demonstrate the impact of the useless factors on the distribution of the T^2 and q statistics respectively. I compute their values, based on the risk premia estimates produced by Fama-MacBeth approach with or without the useless factor, but not Pen-FM, since that would require an assumption on the dimension of the model, and the shrinkage-based estimation is generally silent about testing the size of the model (as opposed to identifying its parameters). The distribution of q is extremely wide and when the model is contaminated by the useless factors is naturally inflated. The impact on the distribution of T^2 is naturally a combination of the impact coming from the Shanken correction term (which is affected by the identification failure through the risk premia estimates), and q quadratics. As a result, the distribution is much closer to that of the oracle estimator; however, it is still characterised by an appreciably heavy right tail, and is generally slighly inflated.

1.6.2 Misspecified Model

The second simulation design that I consider corresponds to the case of a misspecified model, where the cause of misspecification is the omitted variable bias. The data is generated from a 3-factor model, based on 3 canonical Fama-French factors (with data generating parameters obtained from the in-sample model estimation similar to the previous case). However, in the simulations I consider estimating a one factor model (thus, the source of misspecification is omitting the SMB and HML factors). Again, I compare the performance of 3 estimators: (a) Fama-MacBeth, using the simulated market return as the only factor, (b) Fama-MacBeth, using the simulated market return and the irrelevant factor, (c) Pen-FM estimator, using the simulated market return and the irrelevant factor.

Table 1.2 describes the pointwise distribution of the oracle estimator (Fama-MacBeth with an identity weight matrix, applied using only the market excess return as a risk factor), Fama-MacBeth and Pen-FM estimators, when the model includes both true and useless factors.

The results are similar to the case of the correctly specified model. Pen-FM successfully identifies both strong and useless factors with very high accuracy (the useless one is always eliminated from the model by shrinking its premium to 0 even when $T = 30$). The mean squared error and omitted variable bias for all the parameters are close to those of the oracle estimator. At the same time, column 9 demonstrates that the risk premium for the spurious factor, produced by conventional Fama-MacBeth procedure diverges as the sample size increases (its mean squared error increases from 0.445 for T=50 to 1.979 for T=1000). However, the risk premia estimates remain within a reasonable range of parameters, so even if the Fama-MacBeth estimates diverge, it may be difficult to detect it in practice.

Confidence intervals based on t-statistics for the Fama-MacBeth estimator overreject the null hypohesis of no impact of the useless factors (see Tables 1.A.4 and 1.A.6), and should a researcher rely on them, she would be likely to identify a useless factor as priced in the cross-section of stock returns.

Figures 1.A.6-1.A.10 present the quality of fit measures in the misspecified model contaminated by the presence of a useless factor and the ability of Pen-FM to restore them. Figure 1.A.6 shows the distribution of the cross-sectional R^2 for various sample sizes. The similarity between the measures of fit, produced by the model in the absence of the useless factor and with it, but estimated by Pen-FM, is striking: even for such a small sample size as 50 time series observations, the distributions of the R^2 produced by the Fama-MacBeth estimates in the absence of a useless factor, and Pen-FM in a nonidentified model, are virtually identical. This is expected, since, as indicated in Table 1.2, once the useless factor is eliminated from the model, the parameter estimates produced by Pen-FM are nearly identical to those of the one-factor version of Fama-MacBeth. As the sample size increases, the

Table 1.2: Estimates of risk premia in a missspecified model

	True parameter value (λ)	Mean Estimate			Bias			MSE			Shrinkage rate
		Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	
Panel A: T=30											
Intercept	3.315	3.192	3.041	3.149	-0.123	-0.274	-0.166	1.287	1.514	1.253	0
Useful factor	-1.316	-0.619	-0.629	-0.578	0.698	0.687	0.739	1.392	1.58	1.378	0.022
Useless factor		-	-0.019	0	-	-0.019	0	-	0.34	0	1
Panel B: T=50											
Intercept	3.315	3.184	3.053	3.177	-0.132	-0.262	-0.138	1.105	1.456	1.097	0
Useful factor	-1.316	-0.592	-0.621	-0.587	0.724	0.696	0.729	1.252	1.542	1.25	0.014
Useless factor		-	-0.021	0	-	-0.021	0	-	0.445	0	1
Panel C: T=100											
Intercept	3.315	3.253	3.142	3.247	-0.062	-0.173	-0.068	0.781	1.318	0.78	0
Useful factor	-1.316	-0.639	-0.692	-0.634	0.677	0.624	0.682	0.986	1.407	0.989	0.003
Useless factor		-	0.021	0	-	0.021	0	-	0.605	0	1
Panel D: T=250											
Intercept	3.315	3.261	3.159	3.259	-0.054	-0.156	-0.057	0.488	1.138	0.488	0
Useful factor	-1.316	-0.637	-0.708	-0.635	0.679	0.609	0.681	0.814	1.255	0.816	0
Useless factor		-	-0.004	0	-	-0.004	0	-	0.979	0	1
Panel E: T=500											
Intercept	3.315	3.276	3.246	3.275	-0.04	-0.069	-0.04	0.363	1.117	0.363	0
Useful factor	-1.316	-0.649	-0.794	-0.649	0.667	0.522	0.667	0.745	1.212	0.745	0
Useless factor		-	-0.008	0	-	-0.008	0	-	1.374	0	1
Panel F: T=1000											
Intercept	3.315	3.262	3.157	3.262	-0.053	-0.158	-0.053	0.255	1.053	0.255	0
Useful factor	-1.316	-0.634	-0.703	-0.634	0.682	0.614	0.682	0.72	1.197	0.72	0
Useless factor		-	0.049	0	-	0.049	0	-	1.979	0	1

Note. The table summarises the properties of the Fama-MacBeth and Pen-FM estimators with an identity weight matrix in a model for 25 portfolios with a common intercept and 3 factors driving the returns, but with only the first and a useless one considered in the estimation. λ_0 is the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the first useful factor and the useless one. The model is simulated 10,000 times for different values of the sample size (T). The "Oracle" estimator corresponds to the Fama-MacBeth procedure omitting the useless factor, "FM" and "Pen-FM" stand for the Fama-MacBeth and Pen-FM estimators in the model with a useful and a useless factor. The table summarises the mean point estimates of the parameters, their bias and the mean squared error. The mean shrinkage rate corresponds to the percentage of times the corresponding coefficient was set to exactly 0 during 10 000 simulations.

Returns are generated from the multivariate normal distribution with the mean and variance-covariance matrix equal to those of the quarterly nominal excess returns on 25 Fama-French portfolios sorted on size and book-to-market ratio during the period 1962Q2 : 2014Q2. Returns are simulated from a 3-factor model, the latter calibrated to have the same mean and variance as the three Fama-French factors (market excess return, SMB and HML portfolios). The useless factor is generated from a multivariate normal distribution with the mean and variance equal to their sample analogues of nondurable consumption per capita growth rate during the same time period. Betas, common intercept and risk premium for the useful factor come from the Fama-MacBeth estimates of a 3-factor model on the cross-section of 25 Fama-French portfolios. In the estimation, however, only the market return and the irrelevant factor are used; thus the source of misspecification is the omitted factors.

true sample distribution of R^2 becomes much tighter, and peaks around 10-15%, illustrating the model's failure to capture all the variation in the asset returns, while omitting two out of three risk factors.

The cross-sectional R^2 produced by the conventional Fama-MacBeth method is severely inflated by the presence of a useless factor, and its distribution is so wide that it looks almost uniform on $[0,1]$. This illustration is consistent with the theoretical findings of Kleibergen and Zhan (2013) and Gospodinov, Kan, and Robotti (2014b), who demonstrate that under misspecification, the cross-sectional R^2 seems to be particularly affected by the identification failure.

Figure 1.A.7 describes the distribution of $GLS\ R^2$, when the second stage estimates are produced using the identity weight matrix. Interestingly, when the model is no longer identified, $GLS\ R^2$ tends to be lower than its true in-sample value, produced by Pen-FM or the Fama-MacBeth estimator without the impact of the useless factor. This implies that if a researcher were to rely on this measure of fit, she would be likely to underestimate the pricing ability of the model. Figure 1.A.8 presents similar graphs for the distribution of the $GLS\ R^2$, when the risk premia parameters are estimated by GLS in the second stage. The difference between various methods of estimation is much less pronounced, although Fama-MacBeth tends to somewhat overestimate the quality of fit produced by the model.

The average pricing errors displayed in Figure 1.A.10 also indicate a substantial impact of the useless factor in the model. When such a factor is included, and risk premia parameters are estimated using the conventional Fama-MacBeth approach, the APE seem to be smaller than they actually are, resulting in s spurious improvement in the model's ability to explain the difference in asset returns. Again, this is nearly perfectly restored once the model is estimated by Pen-FM.

The Hansen-Jagannathan distance (Figure 1.A.9) is often used to assess model misspecification, since the greater is the distance between the set of SDFs that price a given set of portfolios and the one suggested by a particular specification, the higher is the degree of mispricing. When a useless factor is included, HJ in the Fama-MacBeth estimation has a much wider support than it normally does; and, on average, it tends to be higher.

Figure 1.A.11 and 1.A.13 demonstrate the impact of the useless factors on the distribution of T^2 and q statistics in a misspecified model. Again, I compute their values on the basis of the risk premia estimates produced by the Fama-MacBeth approach with or without the useless factor, but not Pen-FM, since computing these statistics requires using the matrices

with the dimension, depending on the number of factors in the model (and not just their risk premia values). When the model contains a spurious factor, the distribution of q becomes extremely wide and skewed to the right. The effect of spurious factors on the distribution of T^2 is naturally a combination of the influence coming from the Shanken correction term (which is affected by the identification failure through the risk premia estimates), and q. T^2 is generally biased towards 0, making it harder to detect the model misspecification in the presence of a useless factor.

1.6.3 Robustness Check

In order to assess the numerical stability and finite sample properties of the Pen-FM estimator, I study how the survival rates of useful and useless factors depend on the tuning parameters within the same simulation design of either the correct or the misspecified model descibed in the earlier sections.

Table 1.3 summarises the survival rates for the useful and useless factors as a function of the tuning parameter d, which defines the curvature of the penalty. In Proposition 1.1 I proved the Pen-FM estimator to be consistent and asymptotically normal for all values of $d > 2$. In this simulation I fix the other tuning parameter value, $\eta = \bar{\sigma}$, and vary the value of d from 3 to 10. Each simulation design is once again repeated 10,000 times, and the average shrinkage rates of the factors are reported. Intuitively, the higher the curvature parameter, the harsher is the estimated difference between a strong and a weak factor, and hence, one would also expect a slighly more pronounced difference between their shrinkage rates.

It can be clearly seen that the behaviour of the estimates is nearly identical for different values of the curvature parameter and within 1% difference from each other. The only case that stands out, is when the sample is very small (30-50 observations) and $d = 3$. In this case the useful factor has been mistakenly identified as the spurious one in 1-2.5% of the simulations, but these types of fluctuations are fully expected when dealing with such a small sample with a relatively low signal-to-noise ratio. A similar pattern characterises the shrinkage rates for the useless factors, which are extremely close to 1.

Table 1.4 shows how the shrinkage rates of Pen-FM depend on the value of the other tuning parameter, η, which is responsible for the overall weight on the penalty compared with the standard component of the loss function (see Equation (1.9)) and could be thought of as the level parameter. Once again, I conduct 10,000 simulations of the correctly or incorrectly

Table 1.3: Shrinkage rate dependence on the value of the tuning parameter d

	Shrinkage rates for the									
	useful factor, $\lambda_1 \neq 0$					useless factor				
T	$d=3$	$d=4$	$d=5$	$d=7$	$d=10$	$d=3$	$d=4$	$d=5$	$d=7$	$d=10$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Correctly specified model										
30	0.0137	0.0032	0.0031	0.0014	0.0000	1.0000	0.9947	0.9915	0.9957	0.9981
50	0.0126	0.0011	0.0013	0.0012	0.0011	1.0000	0.9936	0.9926	0.9968	0.9992
100	0.0095	0.0010	0.0001	0.0002	0.0001	1.0000	0.9989	0.9987	1.0000	1.0000
250	0.0011	0.0001	0.0000	0.0001	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000
500	0.0001	0.0001	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1000	0.000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Panel B: Misspecified model										
30	0.0284	0.0063	0.0001	0.0001	0.0000	0.9968	0.9905	0.9821	0.9810	0.9989
50	0.0252	0.0011	0.0000	0.0000	0.0000	0.9976	0.9947	0.9768	0.9922	1.0000
100	0.0063	0.0001	0.0000	0.0000	0.0000	0.9978	0.9968	0.9905	1.0000	1.0000
250	0.0000	0.0000	0.0000	0.0000	0.0000	0.9998	0.9947	1.0000	1.0000	1.0000
500	0.0000	0.0000	0.0000	0.0000	0.0000	0.998	0.9999	1.0000	1.0000	1.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note. The table summarises the shrinkage rates for the useful/useless factor produced by the Pen-FM estimator for various sample sizes (T) and a range of parameters, $d = 2, 3, 5, 7, 10$, when η_0 is set at the average standard deviation of the residuals from the first stage. Simulation designs for the correctly specified and misspecified models correspond to those described in Tables 1.1 and 1.2. Each sample is repeated 10,000 times.

specified model for the various sample size, and compute the shrinkage rates for both useful and useless factors. I fix the curvature tuning parameter, d, at $d = 4$, and vary η.

I consider the following range of parameters:

1. $\eta = \bar{R}^e$, the average excess return on the portfolio;
2. $\eta = ln(\bar{\sigma}^2)$, log of the average volatility of the residuals from the first stage;
3. $\eta = \bar{\sigma}$, the average standard deviation of the first stage residuals;
4. the value of η is chosen by fivefold cross-validation;
5. the value of η is chosen by leave-one-out cross-validation.

I have chosen the values of the tuning parameter η that either capture the scale of the data (for example, whether excess returns are displayed in percentages or not), or are suggested by

Table 1.4: Shrinkage rate dependence on the value of the tuning parameter η_0

	Shrinkage rates for the									
	useful factor, $\lambda_1 \neq 0$					useless factor				
T	$\eta_0 = \bar{R}^e$	$\eta_0 = ln(\sigma^2)$	$\eta_0 = \bar{\sigma}$	$CV(5)$	$CV(n-1)$	$\eta_0 = \bar{R}^e$	$\eta_0 = ln(\sigma^2)$	$\eta_0 = \bar{\sigma}$	$CV(5)$	$CV(n-1)$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Correctly specified model										
30	0.0008	0.0008	0.0031	0.0014	0.0000	0.9888	0.9873	0.9947	0.9957	0.9981
50	0.0016	0.0001	0.0010	0.0012	0.0011	0.9857	0.9944	0.9936	0.9968	0.9992
100	0.0000	0.0000	0.0007	0.0002	0.0001	0.9976	0.9960	0.9989	1.0000	1.0000
250	0.0000	0.0000	0.0001	0.0001	0.0001	0.9992	0.9992	1.0000	1.0000	1.0000
500	0.0000	0.0000	0.0001	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Panel B: Misspecified model										
30	0.0284	0.0063	0.0001	0.0001	0.0000	0.9968	0.9905	0.9640	0.9637	0.9989
50	0.0252	0.0011	0.0000	0.0000	0.0000	0.9976	0.9947	0.9749	0.9912	1.0000
100	0.0063	0.0001	0.0000	0.0000	0.0000	0.9978	0.9968	0.9975	0.9971	1.0000
250	0.0000	0.0000	0.0000	0.0000	0.0000	0.9998	0.9947	1.0000	1.0000	1.0000
500	0.0000	0.0000	0.0000	0.0000	0.0000	0.998	0.9999	1.0000	1.0000	1.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note. The table illustrates the shrinkage rates for the useful/useless factor produced by the Pen-FM estimator for various sample sizes (T) and a range of parameters η_0, while $d = 4$. Simulation designs for the correctly specified and misspecified models correspond to those described in Tables 1.1 and 1.2. Tuning parameter η_0 is set to be equal to 1) average excess return on the portfolio, 2)logarithm of average variance of the residuals from the first stage, 3) average standard deviation of the residuals from the first stage, 4) the average value of the tuning parameter chosen by 5-fold cross-validation, 5) the average value of the tuning parameter chosen by leave-one-out cross-validation. Each sample is repeated 10,000 times.

some of the data-driven techniques[1]. Cross-validation (CV) is intuitively appealing, because it is a data-driven method and it naturally allows one to assess the out-of sample performance of the model, treating every observation as part of the validation set only once. CV-based methods have been extensively used in many different applications, and have proved to be extremely useful[2]. Here I briefly describe the so-called k-fold cross-validation.

The original sample is divided into k equal size subsamples, followed by the following algorithm.

- Pick a subsample and call it a validation set; all the other subsamples form a training set.

- Pick a point on the grid for the tuning parameters. For the chosen values of the tuning

[1]Although the table presents the results for the tuning parameteres selected by cross-validation, I have also considered such alternative procedures as BIC, Generalised BIC and the *pass* selection stability criterion. The outcomes are similar both quantitively and qualitatively, and are available upon request.

[2]For an excellent overview see, e.g. Hastie, Tibshirani, and Friedman (2011)

parameters estimate the model on the training set and assess its performnce on the validation set by the corresponding loss function ($L_T(\hat{\lambda})$).

- Repeat the procedure for all the other subsamples.
- Compute the average of the loss function (CV criterion).
- Repeat the calculations for all the other values of the tuning parameters. Since the location of the minimum CV value is a random variable, it is often suggested that the one to pick the one that gives the largest CV criterion within 1 standard deviation of its absolute minimum on the grid, to ensure the robustness of the result (Friedman, Hastie, and Tibshirani (2010)).

Table 1.4 summarises the shrinkage rates of the useful and useless factors for different values of the level tuning parameter, η. Similar to the findings in Table 1.3, the tuning parameter impact is virtually negligible. The useless factor is successfully identified and eliminated from the model in nearly 100% of the simulations, even for a very small sample size, regardless of whether the model is correctly or incorrectly specified, while the strong factor is successfully retained with an equally high probability. The only setting where it causes some discrepancy (within 2-3% confidence bounds) is the case of a misspecified model and a very small sample size ($T = 30$ or 50); but it is again entirely expected for the samples of such size, and therefore does not raise any concerns.

1.6.4 Comparing Pen-FM with alternatives

In this section I compare the finite sample performance of the sequential elimination procedure proposed in Gospodinov, Kan, and Robotti (2014a) and that of Pen-FM with regard to identifying the strong and useless factors.

I replicate the simulation designs used in Table 4 of Gospodinov, Kan, and Robotti (2014a)[1], to reflect various combinations of the risk drivers in a potential four-factor model: strong factors that are either priced in the cross-section of asset returns or not, and irrelevant factors. For each of the variables I compute the frequency with which it is identified as a strong risk factor in the cross-section of asset returns and consequently retained in the model. Each simulation design is repeated 10,000 times.

[1] I am very grateful to Cesare Robotti for sharing the corresponding routines.

Panel A in Table 1.5 summarises the factor survival rates for a correctly specified model. The top panel focuses on the case of 2 priced strong factors, 1 strong factor that is correlated with returns, but not priced, and 1 purely irrelevant factor, which does not correlate with asset returns[1]. For each of the variables I present its survival rate, based on the misspecification-robust t_m- statistic of Gospodinov, Kan, and Robotti (2014a)[2] for a linear SDF model, the frequency with which the corresponding risk premium estimate was *not* set exactly to 0 by the Pen-FM estimator and one minus the average shrinkage rate from the 10, 000 bootstrap replica. The latter also provides an additional comparison of the performance of the pointwise estimator with its bootstrap analogue. A good procedure should be able to recognise the presence of a strong factor and leave it in the model with probability close to 1. At the same time, faced with the useless factor, one needs to recognise it and eliminate from the model, forcing the survival rate to be close to 0.

Consider the case of a correctly specified model, with 2 useful factors that are priced in the cross-section of asset returns, 1 useful, but unpriced factor (with a risk premium equal to zero), and a useless factor, presented in the top panel of Table 1.5. The useless factor is correctly identified and effectively eliminated from the model by both the misspecification-robust $t-$test and the Pen-FM estimator even for a very small sample size (e.g. for a time series of 100 observations, the useless factor is retained in the model in no more than 1% of the simulations. For the smallest sample size of 50 observations, Pen-FM seems also to outperform the sequential elimination procedure, since it retained the useless factor in less than 1.5% of the models only, while the latter was keeping it as part of the specification in roughly 15% of the simulations.

The t_m-test is designed to eliminate not only the useless factors from the linear model, but also those factors that correlate with asset returns, but are not priced in the cross-section of assets. As a result, in 95-99% of cases the useful factor with $\lambda = 0$ is also eliminated from the model. However, the Pen-FM estimator eliminates only the impact of useless factors, and thus retains the presence of all the strongly identified factors in 92-98% of the simulations, depending on the sample size (the associated risk premia could still be insignificant).

[1]The setting proxies the estimation of a 4-factor model on the set of portfolios similar to 25 size and book-to-market and 17 industry portfolios. For a full description of the simulation design, please refer to Gospodinov, Kan, and Robotti (2014a)

[2]The t_c-statistic for a correctly specified model performs very similar to t_m in terms of the factor survival rates. Since it is not known *ex ante*, whether the model is correctly specified or not, I focus on the outcome of the t_m-test.

Table 1.5: Survival rates of useful and irrelevant factors

Panel A: Correctly specified model

	Useful ($\lambda_1 \neq 0$)			Useful ($\lambda_2 \neq 0$)			Useful ($\lambda_3 = 0$)			Useless		
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0628	1	0.9995	0.1120	0.9166	0.9245	0.0588	0.9233	0.9426	0.1541	0	0.0126
100	0.1760	1	0.9997	0.2431	0.9403	0.9439	0.0207	0.9539	0.9606	0.0072	0	0.0079
150	0.3444	1	0.9998	0.4623	0.9652	0.9598	0.0232	0.9622	0.9695	0.0031	0	0.0053
200	0.5142	1	0.9998	0.6599	0.9787	0.9686	0.0231	0.9788	0.9749	0.0023	0	0.0040
250	0.6614	1	0.9998	0.8035	0.9761	0.9742	0.0231	0.9746	0.9786	0.0017	0	0.0032
600	0.9864	1	0.9998	0.9987	0.9802	0.9784	0.0141	0.9777	0.9813	0.0006	0	0.0027
1000	0.9999	1	0.9999	1.0000	0.9828	0.9815	0.0117	0.9833	0.9829	0.0003	0	0.0023

	Useful ($\lambda_1 \neq 0$)			Useful ($\lambda_2 \neq 0$)			Useless			Useless		
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0573	1	0.9999	0.0984	1	1	0.1537	0	0.0164	0.1485	0	0.0232
100	0.1739	1	0.9999	0.2351	1	1	0.0068	0	0.0065	0.0085	0	0.0119
150	0.2020	1	1	0.2290	1	1	0.0080	0	0.0059	0.0032	0	0.0079
200	0.5265	1	1	0.6582	1	1	0.0017	0	0.0044	0.0025	0	0.0059
250	0.6742	1	1	0.8080	1	1	0.0015	0	0.0035	0.0015	0	0.0040
600	0.9880	1	1	0.9985	1	1	0.0007	0	0.0029	0.0003	0	0.0034
1000	1	1	1	0.9900	1	1	0.0000	0	0.0025	0.0002	0	0.0029

Panel B: Misspecified model

	Useful ($\lambda_1 \neq 0$)			Useful ($\lambda_2 \neq 0$)			Useful ($\lambda_3 = 0$)			Useless		
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0640	1	0.9995	0.1167	0.9453	0.9122	0.0676	0.9132	0.9437	0.1790	0	0.0133
100	0.1696	1	0.9996	0.2353	0.9617	0.9415	0.0224	0.9425	0.9614	0.0142	0	0.0075
150	0.3343	1	0.9997	0.4389	0.9733	0.9566	0.0221	0.9566	0.9699	0.0088	0	0.0052
200	0.5016	1	0.9998	0.6298	0.9787	0.9653	0.0240	0.9652	0.9751	0.0080	0	0.0039
250	0.6526	1	0.9998	0.7750	0.9826	0.9713	0.0238	0.9775	0.9786	0.0079	0	0.0031
600	0.9806	1	0.9998	0.9963	0.9850	0.9758	0.0138	0.9751	0.9812	0.0073	0	0.0026
1000	0.9972	1	0.9998	0.9989	0.9871	0.9792	0.0121	0.9764	0.9830	0.0088	0	0.0022

	Useful ($\lambda_1 \neq 0$)			Useful ($\lambda_2 \neq 0$)			Useless			Useless		
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0406	1	0.9986	0.0815	0.7782	0.7805	0.1669	0	0.0228	0.1660	0	0.0279
100	0.0985	1	0.9992	0.1310	0.9246	0.9181	0.0138	0	0.0150	0.0141	0	0.0184
150	0.1928	1	0.9994	0.2493	0.9781	0.9744	0.0083	0	0.0123	0.0093	0	0.0134
200	0.3058	1	0.9996	0.3840	0.9793	0.9717	0.0074	0	0.0101	0.0081	0	0.0103
250	0.4221	1	0.9997	0.5180	0.9921	0.9927	0.0073	0	0.0082	0.0078	0	0.0087
600	0.9026	1	0.9997	0.9516	1	0.9980	0.0097	0	0.0069	0.0086	0	0.0073
1000	0.9822	1	0.9997	0.9922	1	1	0.0102	0	0.0059	0.0096	0	0.0063

Note. The table summarises the survival rates for the useful/useless factors in the simulations of a 4-factor model (correctly or incorrectly specified) for different sample sizes. For each of the factors, I compute its survival rate from 10,000 simulations, based on the t_m statistic from Gospodinov, Kan, and Robotti (2014a)(Table 4), the pointwise estimates produced by the Pen-FM estimator (e.g. the frequency with which the risk premia estimate was *not* set exactly to 0), and one minus the average shrinkage rate from the Pen-FM estimator in 10,000 bootstrap replicas. For a complete description of the simulation design, please refer to Gospodinov, Kan, and Robotti (2014a).

When the span of the data is not sufficiently large, it is hard to correctly retain a significant factor, even if it is strongly identified in the data. For example, when the sample size is only about 200 observations, the strong factor is mistakenly identified as a useless/insignificant one in 40-50% of the simulations. When $T = 50$, the survival rates for the strong factors are accordingly only 6 and 11%. The inference is restored once the sample size is increased to about $T = 600$, corresponding to roughly 50 years of monthly observations. The Pen-FM estimator seems to be quite promising for the applications relying on quarterly or yearly data, where the sample size is rather small, because it retains strong factors in the model with a very high probability (the first strong factor is retained in 99.9% of the cases for all the sample sizes, while the second one is retained in 92-98% of the simulations). It also worth highlighting that the pointwise and bootstrap shrinkage rates of Pen-FM are very close to each other, with the difference within 2%, supporting the notion that bootstrap replicas approximate the pointwise distribution of the estimates rather well, even for a very small sample size.

The second panel presents findings for a correctly specified model with 2 useful (and priced) and two useless factors. The results are quite similar - both approaches are able to identify the presence of irrelevant factors starting from a very small sample size (again, for $T = 50$, Pen-FM seems to have a little advantage). Pen-FM remains consistent in keeping strongly identified factors in the model regardless of the sample size.

Panel B in Table 1.5 presents the case of a misspecified model, and the results are quite similar to the previous case. The only difference arises for $T = 50$, when the Pen-FM retains the second strong factor in only 77-78% of the simulations compared with the usual 92-95% observed for this sample size in other simulations designs; for $T = 100$ the strong factor is retained already in 91-92% of the simulations.

Overall, Pen-FM seems to be rather accurate at deciphering the strength of a factor, and could be particularly useful for working with quarterly or yearly data, where the sample size is naturally small.

Table 1.6 summarises the factor survival rates produced by the adaptive lasso in the same simulation design of Gospodinov, Kan, and Robotti (2014a). As discussed in Section 1.4, when the model is no longer identified, the adaptive lasso is not expected to correctly identify the factors that are priced in the cross-section of asset returns.

$$\hat{\lambda}_{AdL} = \arg\min_{\lambda \in \Theta} \left[\bar{R}^e - \hat{\beta}\lambda\right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda\right] + \eta_T \sum_{j=1}^{k} \frac{1}{|\hat{\lambda}_{j,ols}|^d} |\lambda_j|,$$

When the model includes useless factors, prior OLS-based estimates of the risk premia that define the individual weights in the penalty no longer have the desired properties, since weak identification contaminates their estimation. As a result, adaptive lasso produces erratic behaviour for the second stage estimates, potentially shrinking true risk drivers and/or retaining the useless ones. Particular shrinkage rates will depend on the strength of the factor, its relation to the other variables, and the prior estimates of the risk premia.

Table 1.6 summarises the average factor survival rates produced by the Pen-FM estimator with $d = 4$ and $\eta = \bar{\sigma}$ (the baseline scenario) with those of the adaptive lasso, when the tuning parameter is chosen via the BIC[1].

For a correctly specified model (Panel A), the adaptive lasso nearly always retains the second useful factor, but not the first, which is often eliminated from the model for a relatively moderate sample size (e.g. when $T = 250$, it is retained in only 62.6% of the simulations). Furthermore, unlike the Pen-FM, the adaptive lasso estmator is not able to recognise the presence of a useless factor, and it is never eliminated.

If the model is misspecified, the impact of the identification failure on the original penalty weights is particularly severe, which results in worse factor survival rates for the adaptive lasso. The first of the useful factors is eliminated from the model with a high probability (e.g. for $T = 250$, it is retained only in 45.66% and 34.31% of the simulations, respectively, depending on whether the simulation design includes 1 or 2 useless factors). The second useless factor is always retained in the model, and the first one increasingly so (e.g. for a sample of 50 observations it is a part of the model in 56.54% of the simulations, while for $T = 1000$ already in 96.18%). This finding is expected, since as the sample size increases, the risk premia for the useless factors in the misspecified models tend to grow larger (along with their t-statistic) and the adaptive lasso penalty becomes automatically smaller, suggesting that it would be useful to preserve such factors in the model. The simulations confirm the different nature of the estimators and a quite drastic difference in the estimation of risk premia parameters in the presence of useless factors.

[1] I am grateful to Dennis D. Boos for sharing the routine for R, which is available at his webpage, http://www4.stat.ncsu.edu/ boos/var.select/lasso.adaptive.html

Table 1.6: Comparison of the Pen-FM estimator with the adaptive lasso, based on the survival rates of useful and useless factors.

Panel A: Correctly specified model

	Useful ($\lambda_1 \neq 0$)		Useful ($\lambda_2 \neq 0$)		Useful ($\lambda_3 = 0$)		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	0.4172	0.9166	1	0.9233	0.7340	0	1
100	1	0.4745	0.9403	1	0.9539	0.8392	0	1
150	1	0.5173	0.9652	1	0.9622	0.9262	0	1
200	1	0.5743	0.9787	1	0.9748	0.9431	0	1
250	1	0.6260	0.9761	1	0.9746	0.9694	0	1
600	1	0.8132	0.9802	1	0.9777	1	0	1
1000	1	0.9099	0.9828	1	0.9833	1	0	1

	Useful ($\lambda_1 \neq 0$)		Useful ($\lambda_2 \neq 0$)		Useless		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	1	1	0.9322	0	1	0	1
100	1	1	1	0.9851	0	1	0	1
150	1	1	1	0.9955	0	1	0	1
200	1	1	1	1	0	1	0	1
250	1	1	1	1	0	1	0	1
600	1	1	1	1	0	1	0	1
1000	1	1	1	1	0	1	0	1

Panel B: Misspecified model

	Useful ($\lambda_1 \neq 0$)		Useful ($\lambda_2 \neq 0$)		Useful ($\lambda_3 = 0$)		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	0.4691	0.9453	1	0.9132	0.6133	0	1
100	1	0.4782	0.9617	1	0.9424	0.7134	0	1
150	1	0.4784	0.9733	1	0.9566	0.7650	0	1
200	1	0.4870	0.9787	1	0.9652	0.7612	0	1
250	1	0.4566	0.9826	1	0.9775	0.8377	0	1
600	1	0.5179	0.9850	1	0.9751	0.9810	0	1
1000	1	0.6433	0.9989	1	0.9764	0.9959	0	1

	Useful ($\lambda_1 \neq 0$)		Useful ($\lambda_2 \neq 0$)		Useless		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	0.5352	0.7782	1	0	0.5654	0	1
100	1	0.4833	0.9246	1	0	0.6132	0	1
150	1	0.5090	0.9781	1	0	0.6911	0	1
200	1	0.4566	0.9793	1	0	0.7177	0	1
250	1	0.3431	0.9921	1	0	0.7432	0	1
600	1	0.3217	1	1	0	0.9210	0	1
1000	1	0.2918	1	1	0	0.9618	0	1

Note. The table summarises the survival rates for the useful/useless factors in the simulations of a 4-factor model (correctly or incorrectly specified) for different sample sizes. For each of the factors, I compute its survival rate from 10,000 simulations, based on the shrinkage rate of Pen-FM estimator ($d = 4$ and $\nu_0 = \bar{\sigma}$) in 10,000 bootstrap replicas. I then compute the corresponding factor survival rates of the adaptive lasso with the tuning parameter chosen by BIC. Panel A presents the survival rates for the correctly specified model when it is generated with 2 useful and 2 useless factors, or a combination of 2 useful (and priced), 1 useful (but not priced) factors, and 1 useless factor. Panel B presents similar results for a misspecified model. For a complete description of the simulation designs, please refer to Gospodinov, Kan, and Robotti (2014a)

1.7 Empirical applications

1.7.1 Data

I apply the Pen-FM estimator to a large set of models that have been proposed in the empirical literature, and study how using different estimation techniques may alter parameter estimates and the assessment of model model pricing ability[1]. I focus on the following list of models/factors for the cross-section of stock returns.

CAPM. The model is estimated using monthly excess returns on a cross-section of 25 Fama-French portfolios, sorted by size and book-to-market ratio. I use 1-month Treasury rate as a proxy for the risk-free rate of return. The market portfolio is the value-weighted return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ. Data is taken from Kenneth French website. To be consistent with other applications, relying on tradable factors, I consider the period of January 1972 - December, 2013[2].

Fama-French 3 factor model. The model is estimated using monthly excess returns on a cross-section of 25 Fama-French portfolios, sorted by size and book-to-market ratio. I use 1-month Treasury rate as a proxy for the risk-free rate of return. Following Fama and French (1992), I use market excess return, SMB and HML as the risk factors. SMB is a zero-investment portfolio formed by a long position on the stocks with small capitalisation (cap), and a short position on big cap stocks. HML is constructed in a similar way, going long on high book-to-market (B/M) stocks and short on low B/M stocks.

Carhart 4 factor model. I consider two cross-sections of asset returns to test the Carhart (1997) model: 25 Fama-French portfolios, sorted by size and book-to-market, and 25 Fama-French portfolios, sorted by value and momentum. In addition to the 3 Fama-French factors, the model includes the momentum factor (UMD), a zero-cost portfolio constructed by going long the previous 12-month return winners and short the previous 12-month loser stocks.

"Quality-minus-junk". A quality-minus-junk factor (QMJ), suggested in Asness, Frazzini, and Pedersen (2014), is constructed by forming a long/short portfolio of stocks sorted

[1] I have applied the new estimator to a wide set of models; however, for reasons of brevity, in this chapter I focus on a particular subset. Additional empirical results are available upon request.

[2] I have also estimated the models, using other time samples, e.g. the largest currently available, 1947-2013, 1961-2013, or the samples used at the time of the papers publication. There was no qualitative difference between the relative performance of Pen-FM and the Fama-MacBeth estimator (i.e. if the factor has been identified as a strong/weak one, it continues to be so when a different time span is used to estimate the model). Additional empirical results are available upon request.

by their quality (which is measured by profitability, growth, safety and payout). I use the set of excess returns on Fama-French 25 portfolios, sorted by size and book-to-market as the test assets, and consider a 4 factor model, which includes market excess return, SMB, HML and QMJ.

q-factor model. I consider the so-called q-factor model, various specifications of which have been suggested in the prior literature linking stock performance to investment-related factors (e.g. Liu, Whited, and Zhang (2009), Hou, Xue, and Zhang (2014), Li and Zhang (2010)). I consider the 4 factor specification adopted in Hou, Xue, and Zhang (2014), and that includes market excess return, the size factor (ME), reflecting the difference between the portfolios of large and small stocks, the investment factor (I/A), reflecting the difference in returns on stocks with high/low investment-to-assets ratio, and the profitability factor, built in a similar way from sorting stocks on their return-on-equity (ROE)[1]. I apply the model to several collections of test assets: excess returns on 25 Fama-French portfolios sorted by size and book-to-market, 25 Fama-French portfolios sorted by value and momentum, 10 portfolios sorted on momentum, and 25 portfolios sorted on price/earnings ratio.

cay-CAPM. This is the version of scaled CAPM suggested by Lettau and Ludvigson (2001b); it uses the long-run consumption-wealth cointegration relationship in addition to the market factor and their interaction term. I replicate their results for exactly the same time sample and a cross-section of the portfolios that were used in the original paper. The data is quarterly, 1963Q3-1998Q3.

cay-CCAPM. Similar to cay-CAPM, the model relies on nondurable consumption growth, *cay*, and their interaction term.

Human Capital CAPM. Jagannathan and Wang (1996) suggested using return on human capital (proxied by after-tax-labour income), as an additional factor for the cross-section of stock returns. I estimate the model on the same dataset, as in Lettau and Ludvigson (2001b).

Durable consumption model. Yogo (2006) suggested a model of the representative agent, deriving utility from the flow of nondurable goods, and the stock of durables. In the linearised version, the model includes three factors: market excess returns and nondurable/durable consumption growth. I estimate the model using several cross-sections: 25 portfolios sorted by size and book-to-market, 24 portfolios sorted by book-to-market within industry, and 24 portfolios sorted by market and HML betas. The data is quarterly, 1951Q3-2001Q4.

[1] I am very grateful to Lu Zhang and Chen Xue for sharing the factors data.

1.7.2 Tradable Factors and the Cross-Section of Stock Returns

Panel A in Table 1.7 below summarises the estimation of the linear factor models that rely on tradable factors. For each of the specifications, I provide the p-value of the Wald test[1] for the corresponding factor betas to be jointly equal to 0. I also apply the sequential elimination procedure of Gospodinov, Kan, and Robotti (2014a), based on the t_m test statistic[2] and indicate whether a particular factor survives it. I then proceed to estimate the models using the standard Fama-MacBeth approach and Pen-FM, using the identity weight matrix. For the estimates produced by the Fama-MacBeth cross-sectional regression, I provide standard errors and p-values, based on t-statistics with and without Shanken correction, and the p-values based on 10,000 replicas of the stationary bootstrap of Politis and Romano (1994), and cross-sectional R^2 of the model fit. For the Pen-FM estimator, I provide the point estimates of risk premia, their average bootstrap shrinkage rates, bootstrap-based p-values and cross-sectional R^2. To be consistent, when discussing the statistical significance of the parameters, I refer to bootstrap-based p-values for both estimators. Greyshading indicates the factors that are identified as weak (or irrelevant) and eliminated from the model by Pen-FM.

There is no difference whether CAPM parameters are estimated by the Fama-MacBeth or the Pen-FM estimator. Both methods deliver identical risk premia (-0.558% per month for market excess return), bootstrap-based p-values and R^2 (13%). A similar result is obtained when I estimate the Fama-French 3 factor model, where both methods deliver identical pricing performance. Market premium is significant at 10%, but negative. This is consistent with other empirical estimates of the market risk premium (e.g. Lettau and Ludvigson (2001b) also report a negative, but insignificant market premium for the cross-section of quarterly returns). HML, however, is significant and seems to be a strong factor. Overall, the model captures a large share of the cross-sectional variation, as indicated by the in-sample value of R^2 at 71%. The common intercept, however, is still quite large, at about 1.3%. There is no significant shrinkage for any of the factors in bootstrap, either, and the parameter estimates are nearly identical.

[1] I use heteroscedasticity and autocorrelation-robust standard errors, based on the lag truncation rule in Andrews (1991).

[2] Since it is not known ex ante, whether the model is correctly specified or not, I use the misspecification-robust test. Further note that the test is designed for a GMM-style estimation, and therefore essentially targets a pairwise correlation between a factor and a panel of assets, not the *partial* one.

Table 1.7: Models for the cross-section of stock returns

				Fama-MacBeth estimator							Pen-FM estimator			
Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
Panel A: tradable factors														
25 portfolios, sorted by size and book-to-market														
CAPM	Intercept	-	-	1.431***	0.4282	0.0008	0.4325	0.0009	0.002	19	1.4307***	0	0.002	19
	MKT	0	yes	-0.658	0.4256	0.1222	0.4764	0.1674	0.184		-0.658	0	0.184	
25 portfolios, sorted by size and book-to-market														
Fama and French	Intercept	-	-	1.252	0.2987	0	0.3054	0	0	70	1.2533	0	0	70
(1992)	MKT	0	yes	-0.703*	0.3035	0.0205	0.3721	0.0587	0.06		-0.704*	0	0.06	
	SMB	0	no	0.145	0.0291	0	0.1424	0.3083	0.376		0.145	0	0.376	
	HML	0	yes	0.43***	0.031	0	0.1376	0.0018	0.008		0.429***	0	0.008	
25 portfolios, sorted by size and book-to-market														
"Quality-minus-junk"	Intercept	-	-	0.7*	0.3257	0.0317	0.3422	0.0409	0.092	84	0.576	0	0.212	83
Asness, Frazzini	MKT	0	yes	-0.327	0.327	0.3177	0.4013	0.4155	0.412		-0.206	0	0.684	
Pedersen (2014)	SMB	0	no	0.174	0.0217	0	0.1459	0.2325	0.288		0.172	0	0.292	
	HML	0	yes	0.398**	0.0265	0	0.1387	0.0041	0.016		0.416***	0	0.008	
	QMJ	0	no	0.44**	0.1102	0.0001	0.1599	0.006	0.016		0.324*	0.084	0.084	
25 portfolios, sorted by size and book-to-market														
Carhart	Intercept	-	-	0.684**	0.3199	0.0325	0.381	0.0726	0.032	76	1.289***	0	0	71
(1997)	MKT	0	no	-0.001	0.3208	0.998	0.411	0.9984	0.546		-0.648	0.001	0.085	
	SMB	0	no	0.111	0.0263	0	0.1044	0.2869	0.276		0.106	0.001	0.301	
	HML	0	yes	0.408***	0.024	0	0.0997	0	0		0.385***	0	0	
	UMD	0	no	2.257	0.5999	0.0002	0.7276	0.0019	0.29		0	0.996	0.996	
25 portfolios sorted by value and momentum														
	Intercept	-	-	0.898	0.396	0.0233	0.4076	0.0275	0.12	90	1.074*	0	0.052	90
	MKT	0	no	-0.146	0.3765	0.6989	0.416	0.7263	0.724		-0.304	0.002	0.434	
	SMB	0	no	0.201	0.0696	0.0039	0.1227	0.1013	0.196		0.224*	0	0.1	
	HML	0	no	0.122	0.253	0.6305	0.2773	0.6608	0.832		0	0.374	0.87	
	UMD	0	yes	0.811***	0.0258	0	0.1397	0	0		0.804***	0	0	
25 portfolios, sorted by size and book-to-market														
q-factor model	Intercept	-	-	1.045***	0.3164	0.001	0.3354	0.0018	0.004	77	1.034***	0	0	70
Hou, Xue and Zhang	MKT	0	yes	-0.553	0.3168	0.0807	0.3937	0.16	0.166		-0.505	0	0.184	
(2014)	M/E	0	yes	0.363**	0.0542	0	0.1513	0.0165	0.05		0.255	0.002	0.158	
	I/A	0	no	0.407***	0.0976	0	0.1329	0.0022	0.004		0.363**	0.004	0.012	
	ROE	0	no	0.494**	0.2029	0.0148	0.2446	0.0432	0.042		0	0.822	0.822	
25 portfolios sorted by value and momentum														
	Intercept	-	-	0.256	0.5046	0.6115	0.5381	0.6339	0.66	88	0.454	0	0.218	88
	MKT	0	yes	0.285	0.4758	0.5489	0.5474	0.6024	0.604		0.105	0.001	0.921	
	M/E	0	no	0.5***	0.0658	0	0.1566	0.0014	0.004		0.482***	0	0.006	
	I/A	0	yes	0.063	0.2438	0.796	0.273	0.8174	0.788		0	0.759	0.979	
	ROE	0	yes	0.665***	0.1467	0	0.1951	0.0006	0.006		0.63***	0	0.004	
10 portfolios sorted on momentum														
	Intercept	-	-	1.164	0.7529	0.1222	0.8086	0.1502	0.432	93	-0.064	0	0.582	90
	MKT	0	yes	-0.631	0.7234	0.3834	0.8037	0.4327	0.73		0.578	0.001	0.951	
	M/E	0	yes	0.73	0.7362	0.3213	0.803	0.3632	0.614		0	0.968	0.968	
	I/A	0	no	0.02	0.5142	0.9685	0.5585	0.971	0.91		0	0.582	0.6	
	ROE	0	no	0.468	0.3192	0.1425	0.3621	0.1961	0.206		0.742**	0.005	0.033	
19 portfolios sorted by P/E ratio														
	Intercept	-	-	2.71	1.447	0.0611	1.7586	0.1233	0.504	81	0.2578	0	0.544	76
	MKT	0	yes	-2.124	1.4002	0.1293	1.714	0.2153	0.7		0.272	0	0.968	

Table 1.7: Models for the cross-section of stock returns

Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	Fama-MacBeth estimator: λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	Pen-FM estimator: λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
	M/E	0	yes	1.132	0.5639	0.0447	0.6995	0.1056	0.54		0	0.957	0.967	
	I/A	0	yes	0.056	0.2395	0.8144	0.3027	0.8527	0.374		0.443*	0.051	0.095	
	ROE	0	no	0.072	0.281	0.798	0.3609	0.842	0.946		0	0.669	0.845	
Panel B: nontradable factors														
25 portfolios sorted by size and book-to-market														
Lettau and Ludvigson (2001)	Intercept	-	-	3.24	0.6601	0	0.7261	0	0	16	3.8128	0.0000	0	0
	Δc	0.0061	no	0.216	0.1653	0.1922	0.1861	0.2468	0.398		0	0.889	0.975	
	Intercept	-	-	3.681**	0.9501	0.0001	1.4126	0.0092	0.012	31	4.106***	0	0.004	1
	cay	0.2006	no	-0.516	2.1813	0.813	3.3786	0.8786	0.518		0	0.958	0.976	
(scaled CAPM)	MKT	0	no	-0.056	0.9751	0.9545	1.6072	0.9724	0.72		-0.251	0	0.902	
	$cay\times$ MKT	0	no	1.136	0.3035	0.0002	0.4602	0.0136	0.398		0	0.839	0.851	
(scaled CCAPM)	Intercept	-	-	4.281***	0.7043	0	1.0093	0	0	70	4.958***	0	0	25
	cay	0.001	no	-0.125	0.2784	0.6526	0.41	0.7599	0.584		-0.385	0.319	0.513	
	Δc	0.7378	no	0.024	0.1095	0.8292	0.1619	0.884	0.654		0	0.689	0.721	
	$cay \times \Delta c$	0.0002	no	0.057	0.0171	0.0009	0.0254	0.0246	0.298		0	0.989	0.989	
(HC-CAPM)	Intercept	-	-	4.467***	0.9389	0	1.684	0.008	0.008	58	4.160***	0.0000	0	1
	MKT	0	no	-1.097	0.9322	0.2392	1.81	0.5445	0.798		-0.296	0	0.918	
	Δy	0.6566	no	1.259	0.3641	0.0005	0.6569	0.0552	0.1		0	0.815	0.815	
(scaled HC-CAPM)	Intercept	-	-	5.184***	0.9293	0	1.5628	0.0009	0.016	77	4.268***	0	0.002	7
	cay	0.0534	no	-0.445	0.2629	0.0908	0.4521	0.3254	0.502		0	0.987	0.987	
	MKT	0	no	-1.987	0.9226	0.0313	1.6995	0.2424	0.564		-0.438	0	0.8	
	Δy	0.3859	no	0.557	0.254	0.0282	0.4331	0.1982	0.23		0	0.792	0.792	
	$cay\times$ MKT	0	no	0.341	0.1841	0.0643	0.3225	0.2908	0.422		0	0.858	0.87	
	$cay \times \Delta y$	0.5135	no	-0.167	0.0678	0.014	0.1153	0.1485	0.422		-0.009	0.809	0.809	
25 portfolios, sorted by size and book-to-market														
Durable consumption CAPM	Intercept	-	-	2.335**	0.9331	0.0123	1.5056	0.1209	0.03	55	3.445***	0	0.002	11
	Δc_{ndur}	0.1116	no	0.641	0.2197	0.0035	0.3565	0.0721	0.126		0	0.974	0.974	
	Δc_{dur}	0.6711	no	0.013	0.1305	0.9215	0.2139	0.952	0.884		0	0.99	0.996	
	MKT	0	no	-0.152	0.9662	0.8754	1.6615	0.9273	0.592		-1.03	0.001	0.359	
24 portfolios, sorted by book-to-market within industry														
	Intercept	-	-	1.767	0.862	0.0404	0.9431	0.061	0.414	11	1.317	0	0.344	3
	Δc_{ndur}	0.1513	no	0.232	0.1061	0.029	0.1221	0.0579	0.526		0	0.993	0.995	
	Δc_{dur}	0.6878	no	-0.002	0.1836	0.9891	0.2043	0.9902	0.738		0	0.976	0.998	
	MKT	0	no	0.44	0.8346	0.5977	1.0789	0.6831	0.46		0.89	0.002	0.488	
24 portfolios, sorted by market and HML betas														
	Intercept	-	-	1.558**	0.6859	0.0231	0.9654	0.1066	0.014	44	2.185***	0	0.004	1
	Δc_{ndur}	0.9222	no	0.522	0.1578	9e-04	0.2253	0.0206	0.272		0	0.999	0.999	
	Δc_{dur}	0.021	no	0.112	0.1277	0.3823	0.1836	0.5434	0.456		0	0.996	0.998	
	MKT	0	no	0.338	0.6665	0.6122	1.1002	0.7587	0.842		-0.169	0	0.942	

Table 1.7: Models for the cross-section of stock returns

				Fama-MacBeth estimator							Pen-FM estimator			
Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)

Note. The table presents the risk premia estimates and fit for different models of the cross-section of stocks. Panel A summarises results for the models that rely on tradable risk factors, while Panel B demonstrated similar results for the models,relying on nontradable factors. First column describes the estimated model, or refers to the paper where the original factor was first proposed. Column 2 presents the list of the risk factors used in the corresponding speicification. Column 3 presents the p-value of the Wald test for the factor being a useless one, based on the first stage estimates of betas and heteroscedasticity and autocorrelation-robust standard errors, based on the lag truncation rule suggested in Andrews (1991). Column 4 indicates whether a particular risk factor has survived the sequential elimination procedure based on the misspecification-robust t_m-statistic of Gospodinov, Kan, and Robotti (2014a). Columns 5-11 present the results of the model estimation based on the Fama-MacBeth procedure with an identity weight matrix ($W = I_n$), and include point estimates of the risk premia, OLS and Shanken standard errors, the corresponding p-values, and the p-value based on 10,000 pairwise block stationary bootstrap of Politis and Romano (1994). Column 11 presents the cross-sectional R^2 of the model estimated by the Fama-MacBeth procedure. Columns 12-15 describe Pen-FM estimation of the model, and summarise the point estimates of the risk premia, their shrinkage rate in the 10,000 bootstrap samples, the corresponding p-value of the parameter, and the cross-sectional R^2. Grey areas highlight the factors that are identified as useless/weak by the Pen-FM estimator (and, hence, experience a substantial shrinkage rate)

Including the quality-minus-junk factor improves the fit of the model, as R^2 increases from 71 to 83-84%. The QMJ factor risk premium is set exactly to 0 in 8.4% of bootstrap replicas; however, its impact remains significant at 10%, providing further evidence that including this factor improves the pricing ability of the model. In the Fama-MacBeth estimation, the common intercept was weakly significant at 10%, however, in the case of Pen-FM, it is no longer significant, decreasing from 0.7 to 0.57% (which is partly due to a slightly larger risk premium for HML).

The Carhart (1997) 4-factor model is estimated on two cross-sections of porfolios, highlighting a rather interesting, but at the same time expected, finding, that the sorting mechanism used in portfolio construction affects the pricing ability of the factors. When I estimate the 4-factor model on the cross-section of 25 portfolios, sorted by size and book-to-market ratio, momentum factor is identified by the Pen-FM estimator as the irrelevant one, since the correponsing risk premia is shrunk exactly to 0 in 99.6% of the bootstrap replicas. As a result of this elimination, cross-sectional R^2 in the model estimated by Pen-FM is the same as for the 3-factor Fama-French model, 71%.

On the other hand, when portfolios are sorted on value and momentum, HML is indicated as the irrelevant one, while momentum clearly drives most of the cross-sectional variation. Both models exhibit the same R^2, 90%. Interestingly, once HML is eliminated by Pen-FM from the model, the risk premium on SMB becomes weakly significant at 10%, recovering the true impact of the size factor. This illustration of different pricing ability of the risk factors, when facing different cross-sections of asset returns, is not new, but it is interesting to note that the impact can be so strong as to affect the model identification.

Hou, Xue, and Zhang (2014) suggest a 4 factor model that, the authors claim, manages to explain most of the puzzles in empirical finance literature, with the main contribution coming from investment and profitability factors. Their specification outperforms Fama-French and Carhart models with regards to many anomalies, including operating accrual, R&D-to-market and momentum. Therefore, it seems to be particularly interesting to assess model performance on various test assets. For 25 Fama-French portfolios, the profitability factor impact is not strongly identified, as it is eliminated from the model in 82.2% of the bootstrap replica. At the same time, investment remains a significant determinant of the cross-sectional variation, commanding a premium of 0.36%. A different outcome is observed when using the cross-section of stocks sorted by value and momentum. In this case the profitability factor is removed from the model as the weak one. Size and ROE factors are identified as strong determinants of the cross-sectional variation of returns, with risk premia estimates of 0.484% and 0.63% accordingly. It is interesting to note that, although the I/A factor is eliminated from the model, the cross-sectional R^2 remains at the same high level of 88%.

A particular strength of the profitability factor becomes apparent when evaluating its performance on the cross-section of stocks sorted on momentum. When the conventional Fama-MacBeth estimator is applied to the data, none of the factors command a significant risk premium, although the model explains 93% of the cross-sectional dispersion in portfolio excess returns. Looking at the estimates produced by Pen-FM, one can easily account for this finding: it seems that size and investment factors are only weakly related to momentum-sorted portfolio returns, while it is the profitability factor that drives nearly all of their variation. The model delivers a positive (but highly insignificant) market risk premium, and a large and positive risk premium for ROE (0.742%). Although both M/E and I/A are eliminated from the model, the cross-sectional R^2 is at an impressive level of 90%. This may be due to an identification failure, caused by the presence of useless (or weak) factors, which was masking the impact of the true risk drivers.

When stocks are sorted in portfolios based on their price/earnings ratio, the Fama-MacBeth estimator results in high cross-sectional R^2 (81%), but insignificant risk premia for all the four factors, and a rather large average mispricing at 2.71%. In contrast, the Pen-FM estimator shrinks the impact of the size and profitability factors (which are elimininated in 96.8% and 84.5% of the bootstrap replicas, respectively). As a result, investment becomes weakly significant, commanding a premium of 0.44%, the market premium is also positive (but insignificant) at 0.27%, while the common intercept, which is often viewed as the sign of model misspecification, is only 0.25% (and insignificant). The model again highlights the ability of the Pen-FM estimator to identify and eliminate weak factors from the cross-section of returns, while maintaining the impact of the strong ones. In particular, investment and market factors alone explain 76% of the cross-sectional variation in portfolios, sorted by the P/E ratio.

1.7.3 Nontradable Factors and the Cross-Section of Stock Returns

Standard consumption-based asset pricing models feature a representative agent who trades in financial securities in order to optimize her consumption flow (e.g. Lucas (1976), Breeden

(1979)). In this framework the only source of risk is related to the fluctuations in consumption, and hence, all the assets are priced in accordance with their ability to hedge against it. In the simplest version of the CCAPM, the risk premium associated with a particular security is proportional to its covariance with the consumption growth:

$$E[R_i^e] \approx \lambda \, cov\left(R_{t,i}^e, \Delta c\right)$$

If the agent has the CRRA utility function, λ is directly related to the relative risk aversion, γ, and hence, one of the natural tests of the model consists in estimating this parameter and comparing it with the plausible values for the risk aversion (i.e. < 10). Mehra and Prescott (1985) and Weil (1989) show that in order to match historical data, one would need to have a coefficient of risk aversion much larger than any plausible empirically supported value, thus leading to the so-called equity premium and risk-free rate puzzles. The model was strongly rejected on US data (Hansen and Singleton (1982), Hansen and Singleton (1983), Mankiw and Shapiro (1986)), but led to a tremendous growth in the consumption-based asset pricing literature, which largely developed in two main directions: modifying the model framework in terms of preferences, production sector and various frictions related to decision-making, or highlighting the impact of the data used to validate the model[1].

Not only the estimates of the risk aversion parameter turn out to be unrealistically large, but they are also characterised by extremely wide confidence bounds (e.g. Yogo (2006) reports $\hat{\gamma} = 142$ with the standard errors of 25 when estimating the CCAPM using the Fama-French 25 portfolios). The impact of low covariance between consumption and asset returns could not merely explain a high estimate of the risk aversion, but also lead to the models being weakly identified, implying a potential loss of consistency, nonstandard asymptotic distribution for the conventional OLS or GMM estimators, and the need to rely on identification-robust inference procedures.

Panel B in Table 1.7 reports estimation of some widely used empirical models, relying on nontradable factors, such as consumption. The scaled version of CAPM, motivated by the long-run relationship between conumption and wealth dynamics in Lettau and Ludvigson (2001a), seems to be rather weakly identified, as both *cay* and its product with the market return are eliminated from the model by the Pen-FM estimator in 97.6% and 85.1% of the

[1] The literature on consumption-based asset pricing is vast; for an overview see Campbell (2003) and Ludvigson (2013)

bootstrap replicas, respectively. The resulting specification includes only the market excess return as the only factor for the cross-section of quarterly stock returns, which leads to the well-known illustration of the inability of the classical CAPM to explain any cross-sectional variation, delivering the R^2 of only 1%. The scaled version of Consumption-CAPM also seems to be contaminated by identification failure. Not only the estimates of the risk preia of all three factors are shrunk to 0 with a very high frequency, but even the Wald test for the vector of betas indicates nondurable consumption growth as a rather weak risk factor.

This finding provides a new aspect to the well-known failure of the CCAPM and similar specifications to both match the equity premium and explain the cross-sectional variation in returns.

One of the natural solutions to the problem could lie in using alternative measures for consumption and investment horizons. Kroencke (2013) explicitly models the filtering process used to construct NIPA time series, and finds that the unfiltered flow consumption produces a much better fit of the basic consumption-based asset pricing model and substantially lowers the required level of risk aversion. Daniel and Marshall (1997) show that while the contemporaneous correlation of consumption growth and returns is quite low for the quarterly data, it is substantially increased at lower frequency. This finding would be consistent with investors' rebalancing their portfolios over longer periods of time, either due to transaction costs (market frictions or the costs of information processing), or due to external constraints (e.g. some of the calendar effects). Lynch (1996) further studies the effect of decision frequency and its synchronisation between agents, demonstrating that it could naturally result in a lower contemporaneous correlation between consumption risk and returns. Jagannathan and Wang (2007) state that investors are more likely to make decisions at the end of the year, and, hence, consumption growth, if evaluated then, would be a more likely determinant of the asset returns. These papers could also be viewed as a means to improve model identification.

Jagannathan and Wang (1996) and Santos and Veronesi (2004) argue that human capital (HC) should be an important risk driver for financial securities. I estimate their HC-CAPM on the dataset used in Lettau and Ludvigson (2001b), and find that this model is also contaminated by the identification problem. While the true risk factor may command a significant premium, the model is still poorly identified, as indicated by Table 1.7, and after-tax labour income, as a proxy for human capital, is eliminated by Pen-FM from the model for stock returns. The scaled version of the HC-CAPM also seems to be weakly identified, since the only robust risk factor seems to be market excess return.

Unlike the baseline models that mainly focus on nondurable consumption goods and services, Yogo (2006) argues that the stock of durables is an important driver of financial returns, and taking it into account substantially improves the ability of the model to match not only the level of aggregate variables (e.g. the equity premium, or the risk-free rate), but also the cross-sectional spread in portfolios, sorted on various characteristics. Table 1.7 illustrates the estimation of durable consumption CAPM, that includes market returns, as well as durable and nondurable consumption growth as factors on several cross-sections of portfolios. Both consumption-related factors seem to be rather weak drivers for the cross-section of stocks, and are eliminated in roughly 99% of the bootstrap replicas. This finding is also robust across the different sets of portfolios. Once the weak factors are eliminated from the model, only the market excess return remains; however, its price of risk is negative and insignificant, while the resulting R^2 is rather low at only 1-11%.

One of the potential explanations behind such a subpar performance of the nontradable risk factors consists in the measurement error problem. Indeed, if the nondurable consumption growth (or any other variable) is observed with a measurement error, it causes an attenuation bias in the estimates of betas, which could in turn lead to a weak factor problem in small sample[1]. I address this issue by constructing mimicking portfolios of the nontradable factors using a simple linear projection on the cross-section of the corresponding stock returns. By construction, the resulting projection preserves the pricing impact of the original variable, however, it does not have the same measurement error component, as before.

Table 1.8 illustrates the use of mimicking portfolios for some of the models with nontradable factors. While there is considerable improvement in the performance of the nondurable consumption (unless the market return is also included into the model), the main finding remains unchanged: the model still suffer from the identification failures. Cross-products of the consumption-to-wealth ratio and consumption, durable consumption growth, labour and its cross-product still do not generate enough asset exposure to the risk factors to identify the associated risk premia, even when used as mimicking portfolios.

[1]Note, that the classical measurement error leads to a a multiplicative attenuation bias, and therefore can be the sole reason for the lack of identification. In finite sample, however, its presence makes the inference unreliable and, if large enough, could substantially exacerbate the underlying problem

Table 1.8: Mimicking portfolios of the nontradeable factor and the cross-section of stock returns

				Fama-MacBeth estimator							Pen-FM estimator			
Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
				25 portfolios sorted by size and book-to-market										
	Intercept	-	-	2.604	1.0306	0.0115	1.2044	0.0306	0.01	27	3.5769	0	0.008	8
	cay	0	no	-2.46	0.5413	0	0.7895	0.0018	0.105		-0.382	0.9725	0.9735	
Scaled CAPM	MKT	0	yes	1.249	1.0998	0.2563	1.4605	0.3926	0.765		0.232	0.0005	0.8125	
	$cay\times$ MKT	0.011	no	-0.114	0.0959	0.2344	0.1211	0.3463	0.93		0	0.9955	0.9955	
Scaled CCAPM	Intercept	-	-	2.253	0.9141	0.0137	1.0663	0.0346	0.019	69	2.4462	0	0.023	68
	cay	0	yes	-0.111	0.0731	0.128	0.0974	0.2538	0.321		-0.099	0.071	0.311	
	Δc	0	yes	0.175	0.0567	0.002	0.0721	0.015	0.024		0.164	0.006	0.015	
	$cay \times \Delta c$	0.056	no	-0.005	0.0049	0.3124	0.0061	0.4124	0.806		0	0.9805	0.9845	
HC-CAPM	Intercept	-	-	0.704	1.3241	0.595	1.6608	0.6716	0.759	93	4.7047	0	0.111	9
	MKT	0	no	2.472	1.3211	0.0613	1.7963	0.1688	0.078		-0.774	0	0.811	
	Δy	0	no	0.641	0.1689	0.0001	0.2237	0.0042	0.001		0	0.812	0.812	
Scaled HC-CAPM	Intercept	-	-	0.338	1.04	0.7454	1.3891	0.8079	0.628	94	4.7453	00	0.1184	8
	cay	0	no	0.483	0.3568	0.1757	0.671	0.4715	0.4196		0	0.976	0.9776	
	MKT	0	yes	2.856	1.0269	0.0054	1.537	0.0632	0.0716		-0.839	0	0.9384	
	Δy	0.98	no	0.68	0.1213	0	0.1771	0.0001	0.0008		0	0.6938	0.6938	
	$cay\times$ MKT	0.001	no	0.138	0.1173	0.2388	0.1632	0.3974	0.6792		0	0.9888	0.9896	
	$cay \times \Delta y$	0.194	no	0.025	0.0209	0.2332	0.0284	0.3811	0.4956		0	0.9948	0.9948	
Durable consumption model	Intercept	-	-	2.333	0.9333	0.0124	1.0883	0.0321	0.027	55	3.6587	0	0	21
	Δc_{ndur}	0	no	0.136	0.0525	0.0096	0.0643	0.0346	0.073		0	0.976	0.976	
	Δc_{dur}	0	no	-0.019	0.0284	0.5011	0.0381	0.6161	0.987		0	0.998	1	
	MKT	0	yes	-0.19	0.9624	0.8433	1.2608	0.88	0.71		-1.252	0	0.22	

Note. The table presents the risk premia estimates and fit for different models of the cross-section of stocks using mimicking portfolios for the nontradable factors. First column describes the estimated model, or refers to the paper where the original factor was first proposed. Column 2 presents the list of the risk factors used in the corresponding specification. Column 3 presents the p-value of the Wald test for the factor being a useless one, based on the first stage estimates of betas and heteroscedasticity and autocorrelation-robust standard errors, based on the lag truncation rule suggested in Andrews (1991). Column 4 indicates whether a particular risk factor has survived the sequential elimination procedure based on the misspecification-robust t_m-statistic of Gospodinov, Kan, and Robotti (2014a). Columns 5-11 present the results of the model estimation based on the Fama-MacBeth procedure with an identity weight matrix ($W = I_n$), and include point estimates of the risk premia, OLS and Shanken standard errors, the corresponding p-values, and the p-value based on 10,000 pairwise block stationary bootstrap of Politis and Romano (1994). Column 11 presents the cross-sectional R^2 of the model estimated by the Fama-MacBeth procedure. Columns 12-15 describe Pen-FM estimation of the model, and summarise the point estimates of the risk premia, their shrinkage rate in the 10,000 bootstrap samples, the corresponding p-value of the parameter, and the cross-sectional R^2. Grey areas highlight the factors that are identified as useless/weak by the Pen-FM estimator (and, hence, experience a substantial shrinkage rate)

1.8 Conclusion

Identification conditions play a major role in model estimation, and one must be very cautious when trying to draw quantitaive results from the data without considering this property first. While in some cases this requirement is fairly easy to test, the use of more complicated techniques sometimes makes it more difficult to analyze. This chapter deals with one particular case of underidentification: the presence of useless factors in the linear asset pricing models. I proposed a new estimator that can be used simulatenously as a model diagnostic and estimation technique for the risk premia parameters. While automatically eliminating the impact of the factors that are either weakly correlated with asset returns (or do not correlate at all), the method restores the identification of the strong factors in the model, their estimation accuracy, and quality of fit.

Applying this new technique to real data, I find support for the pricing ability of several tradable factors (e.g. the three Fama-French factors or the 'quality-minus-junk' factor). I further demonstrate that the profitability factor largely drives the cross-section of momentum-sorted portfolios, contrary to the outcome of the standard Fama-MacBeth estimation.

It seems that much of the cross-sectional research with nontradable factors, however, should also be considered through the prism of model identification, as nearly all the specifications cosidered are contaminated by the problem of rank deficiency. How and whether the situation is improved in nonlinear models are undoubtedly very important questions, and form an interesting agenda for future research.

Appendix

1.A Graphs and Tables

Table 1.A.1: Empirical size of the bootstrap-based confidence bounds in a correctly specified model

	λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Fama-MacBeth estimator in a model with only a useful factor									
30	0.065	0.029	0.003	0.033	0.009	0.000	-	-	-
50	0.075	0.037	0.009	0.047	0.016	0.001	-	-	-
100	0.096	0.055	0.015	0.062	0.026	0.005	-	-	-
250	0.103	0.049	0.009	0.054	0.023	0.003	-	-	-
500	0.106	0.057	0.008	0.059	0.027	0.003	-	-	-
1000	0.101	0.043	0.008	0.048	0.018	0.002	-	-	-
Panel B: Fama-MacBeth estimator in a model with a useful and a useless factor									
25	0.045	0.023	0.003	0.024	0.006	0.001	0.008	0.001	0
50	0.061	0.032	0.002	0.029	0.011	0.002	0.007	0.002	0
100	0.068	0.029	0.004	0.033	0.009	0.000	0.009	0.002	0
250	0.069	0.027	0.003	0.034	0.011	0.001	0.006	0.002	0
500	0.071	0.036	0.008	0.043	0.018	0.001	0.005	0.003	0.001
1000	0.063	0.028	0.007	0.037	0.012	0.006	0.005	0	0
Panel C: Pen-FM estimator in a model with a useful and a useless factor									
25	0.051	0.027	0.006	0.027	0.007	0.001	0.002	0	0
50	0.081	0.038	0.005	0.036	0.016	0.002	0	0	0
100	0.09	0.041	0.005	0.047	0.014	0.001	0	0	0
250	0.093	0.05	0.008	0.048	0.025	0.001	0	0	0
500	0.095	0.054	0.013	0.055	0.026	0.004	0	0	0
1000	0.097	0.042	0.01	0.061	0.021	0.007	0	0	0

Note. The table summarises the empirical size of the bootstrap-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with the identity weight matrix in the second stage and at various significance levels (α=10%, 5%, 1%). The model includes a true risk factor and a useless one. λ_0 stands for the value of the intercept, λ_1 and λ_2 are the corresponding risk premia of the factors. Panel A corresponds to the case of the Fama-MacBeth estimator with an identity weight matrix, when the model includes only the useful factor. Panels B and C present the empirical size of the confidence bounds of the risk premia when the model includes both a useful and a useless factor, and the parameters are estimated by Fama-MacBeth or Pen-FM estimator accordingly. The model is simulated 10,000 times for different values of the sample size (T). The confidence bounds are constructed from 10,000 pairwise bootstrap replicas.

For a detailed description of the simulation design, please refer to Table 1.1.

Table 1.A.2: Empirical size of the confidence bounds, based on the t-statistic in a correctly specified model

	Intercept, λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Fama-MacBeth estimator in a model with only a useful factor, without Shanken correction									
25	0.1265	0.073	0.0345	0.127	0.07	0.0325	-	-	-
50	0.114	0.064	0.025	0.1155	0.057	0.0255	-	-	-
100	0.1115	0.058	0.0275	0.105	0.055	0.0285	-	-	-
250	0.107	0.051	0.019	0.1065	0.0575	0.0175	-	-	-
500	0.096	0.0465	0.0195	0.1025	0.052	0.021	-	-	-
1000	0.09	0.047	0.018	0.095	0.043	0.0175	-	-	-
Panel B: Fama-MacBeth estimator in a model with only a useful factor, with Shanken correction									
30	0.095	0.0435	0.011	0.035	0.0085	0.001	-	-	-
50	0.092	0.04	0.014	0.04	0.0115	0.003	-	-	-
100	0.097	0.0445	0.0185	0.051	0.02	0.003	-	-	-
250	0.1	0.0445	0.0145	0.0585	0.016	0.003	-	-	-
500	0.0925	0.045	0.0175	0.056	0.023	0.0055	-	-	-
1000	0.089	0.046	0.017	0.0495	0.019	0.007	-	-	-
Panel C: Fama-MacBeth estimator in a model with a useless factor, without Shanken correction									
30	0.1325	0.076	0.035	0.1305	0.0725	0.037	0.123	0.0695	0.034
50	0.117	0.0625	0.0255	0.1225	0.062	0.028	0.1115	0.0595	0.029
100	0.1115	0.0565	0.0245	0.1065	0.053	0.025	0.106	0.0505	0.0205
250	0.101	0.05	0.0195	0.099	0.051	0.02	0.11	0.049	0.0195
500	0.1075	0.0495	0.021	0.111	0.0515	0.0225	0.0935	0.048	0.023
1000	0.089	0.0485	0.0145	0.09	0.0485	0.0175	0.113	0.058	0.021
Panel D: Fama-MacBeth estimator in a model with a useless factor, with Shanken correction									
30	0.0875	0.0355	0.011	0.031	0.0055	0.001	0.0285	0.0055	0
50	0.074	0.0365	0.0085	0.033	0.007	0.002	0.033	0.0085	0.0015
100	0.0765	0.033	0.0125	0.037	0.015	0.0035	0.024	0.0065	0.0005
250	0.0755	0.0305	0.0085	0.0435	0.0145	0.0025	0.034	0.005	0
500	0.0835	0.0345	0.01	0.0445	0.0165	0.0025	0.0305	0.0075	0.0005
1000	0.0675	0.0335	0.0095	0.041	0.013	0.005	0.0365	0.0075	0

Note. The table presents the empirical size of the t-statistic-based confidence bounds for the Fama-MacBeth estimator with an identity weight matrix in a model with a common intercept for 25 portfolios and a single risk factor, with or without a useless one. λ_0 is the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. The model is simulated 10,000 times for different values of the sample size (T). Panels A and C present the size of the t-statistic, computed using OLS-based heteroscedasticity-robust standard errors. Panels B and D present results based on Shanken correction.

For a detailed description of the simulation design, please refer to Table 1.1.

Table 1.A.3: Empirical size of the bootstrap-based confidence bounds for true values in a misspecified model

	λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Fama-MacBeth estimator in a model with only a useful factor									
25	0.004	0	0	0.003	0	0	-	-	-
50	0.005	0	0	0.003	0.001	0	-	-	-
100	0.004	0.001	0	0.005	0.002	0	-	-	-
250	0.001	0	0	0.009	0.001	0	-	-	-
500	0	0	0	0.05	0.005	0	-	-	-
1000	0.002	0.001	0	0.183	0.049	0.005	-	-	-
Panel B: Fama-MacBeth estimator in a model with a useful and a useless factor									
25	0.01	0.003	0	0.002	0.001	0	0.002	0	0
50	0.005	0	0	0.002	0	0	0.011	0.002	0
100	0.003	0.001	0	0.002	0.001	0	0.055	0.02	0.001
250	0.004	0.002	0	0.005	0.001	0	0.093	0.052	0.014
500	0.002	0	0	0.01	0.001	0	0.088	0.05	0.01
1000	0.003	0.001	0	0.019	0.005	0	0.122	0.066	0.019
Panel C: Pen-FM estimator in a model with a useful and a useless factor									
25	0.007	0.001	0	0.004	0.001	0	0	0	0
50	0.002	0	0	0.002	0	0	0	0	0
100	0.003	0	0	0.003	0	0	0	0	0
250	0.002	0.001	0	0.009	0	0	0	0	0
500	0.004	0.001	0	0.055	0.012	0	0	0	0
1000	0.002	0	0	0.161	0.046	0.002	0	0	0

Note. The table summarises the empirical size of the bootstrap-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with an identity weight matrix in the second stage and at various significance levels (α=10%, 5%, 1%). The misspecified model includes only 1 out of 3 true risk factors, and is further contaminated by the presence of a useless one. λ_0 stands for the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. Panel A corresponds to the case of the Fama-MacBeth estimator with an identity weight matrix, when the model includes only one useful factor. Panels B and C present empirical size of the confidence bounds of the risk premia when the model includes both a useful and a useless factor, and their parameters are estimated by the Fama-MacBeth or Pen-FM procedures accordingly. The model is simulated 10 000 times for different values of the sample size (T). The confidence bounds are constructed from 10 000 pairwise bootstrap replicas.

For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

Table 1.A.4: Empirical size of the confidence bounds for the true values of the risk premia, based on the t-statistic in a mispecified model

	Intercept, λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Fama-MacBeth estimator in a model with only a useful factor, without Shanken correction									
30	0.0145	0.004	0.001	0.02	0.006	0.002	-	-	-
50	0.01	0.003	0.0015	0.015	0.004	0.001	-	-	-
100	0.0035	0	0	0.015	0.002	0	-	-	-
250	0.0025	0.001	0	0.0555	0.014	0.002	-	-	-
500	0.004	0.0015	0	0.1535	0.051	0.009	-	-	-
1000	0.0035	0	0	0.408	0.206	0.0785	-	-	-
Panel B: Fama-MacBeth estimator in a model with only a useful factor, with Shanken correction									
30	0.011	0.0015	5e-04	0.002	0	0	-	-	-
50	0.0065	0.002	0.001	0.003	0.001	0	-	-	-
100	0.0015	0	0	0.001	0	0	-	-	-
250	0.0025	0.001	0	0.014	0.0015	0	-	-	-
500	0.004	0.0015	0	0.0585	0.0115	0.002	-	-	-
1000	0.003	0	0	0.238	0.086	0.0115	-	-	-
Panel C: Fama-MacBeth estimator in a model with a useless factor, without Shanken correction									
25	0.0355	0.015	0.002	0.0435	0.016	0.003	0.135	0.055	0.016
50	0.0435	0.0175	0.0055	0.0555	0.022	0.007	0.2885	0.139	0.0465
100	0.0745	0.0375	0.015	0.0935	0.0465	0.02	0.5945	0.441	0.2375
250	0.178	0.1115	0.0605	0.2325	0.154	0.084	0.805	0.7595	0.696
500	0.301	0.229	0.162	0.3845	0.2935	0.202	0.872	0.8425	0.806
1000	0.4095	0.347	0.2815	0.5355	0.451	0.3625	0.932	0.915	0.8885
Panel D: Fama-MacBeth estimator in a model with a useless factor, with Shanken correction									
25	0.018	0.004	0.0015	0.003	0.001	0	0.0185	0.003	0.001
50	0.0165	0.005	0.0015	0.007	0.001	0	0.054	0.0065	5e-04
100	0.0185	0.0065	0.0015	0.0155	0.003	0.0005	0.249	0.044	0.002
250	0.0305	0.008	0.0015	0.0375	0.011	0.0015	0.672	0.3305	0.051
500	0.0445	0.018	0.006	0.0725	0.027	0.0095	0.8105	0.5845	0.2095
1000	0.056	0.0235	0.011	0.124	0.0595	0.025	0.8895	0.731	0.3965

Note. The table presents the empirical size of the t-statistic-based confidence bounds for the true risk premia values for the Fama-MacBeth estimator with the identity weight matrix in a model with a common intercept for 25 portfolios and a single risk factor, with or without a useless one. λ_0 is the value of the intercept, λ_1 and λ_2 are the corresponding risk premia of the factors. The model is simulated 10,000 times for different values of the sample size (T). Panels A and C present the size of the t-statistic computed using heteroscedasticity-robust standard errors. Panels B and D present the results based on Shanken correction.

For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

Table 1.A.5: Empirical size of the bootstrap-based confidence bounds for the pseudo-true values in a misspecified model

	λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Fama-MacBeth estimator in a model with only a useful factor									
25	0.004	0	0	0	0	0	-	-	-
50	0.005	0	0	0	0	0	-	-	-
100	0.004	0.001	0	0.001	0	0	-	-	-
250	0	0	0	0	0	0	-	-	-
500	0	0	0	0	0	0	-	-	-
1000	0.001	0	0	0	0	0			
Panel B: Fama-MacBeth estimator in a model with a useful and a useless factor									
25	0.01	0.003	0	0.001	0	0	0.002	0	0
50	0.001	0	0	0	0	0	0.011	0.002	0
100	0.003	0.001	0	0	0	0	0.055	0.02	0.001
250	0.004	0	0	0.001	0	0	0.093	0.052	0.014
500	0.002	0	0	0.001	0	0	0.088	0.05	0.01
1000	0.004	0.001	0	0.003	0	0	0.122	0.066	0.019
Panel C: Pen-FM estimator in a model with a useful and a useless factor									
25	0.007	0.001	0	0.001	0	0	0	0	0
50	0.002	0	0	0	0	0	0	0	0
100	0.003	0	0	0	0	0	0	0	0
250	0.002	0.001	0	0.001	0	0	0	0	0
500	0.003	0.001	0	0	0	0	0	0	0
1000	0.002	0	0	0	0	0	0	0	0

Note. The table summarises the empirical size of the bootstrap-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with an identity weight matrix at the second stage and various significance levels (α=10%, 5%, 1%). The misspecified model includes only 1 out of 3 true risk factors, and is further contaminated by the presence of a useless one. λ_0 stands for the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. The pseudo-true values of the risk premia are defined as the limit of the risk premia estimates in a misspecified model without the influence of the useless factor. Panel A corresponds to the case of the Fama-MacBeth estimator with an identity weight matrix, when the model includes only one useful factor. Panels B and C present the empirical size of the confidence bounds of risk premia when the model includes both a useful and a useless factor, and their parameters are estimated by the Fama-MacBeth or Pen-FM procedures accordingly. The model is simulated 10 000 times for different values of the sample size (T). The confidence bounds are constructed from 10 000 pairwise bootstrap replicas.

For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

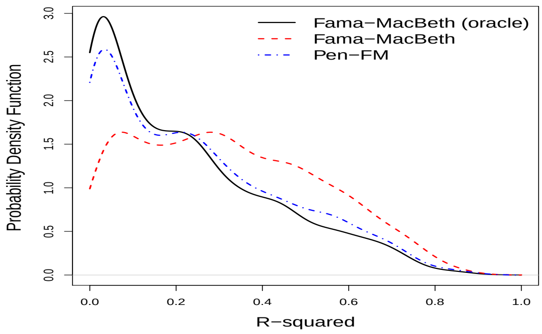
Table 1.A.6: Empirical size of the confidence bounds for the pseudo-true values of risk premia, based on the t-statistic in a mispecified model

	Intercept, λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Fama-MacBeth estimator in a model with only a useful factor, without Shanken correction									
30	0.015	0.004	0.001	0.0085	0.002	0	-	-	-
50	0.0085	0.003	0.0015	0.007	0.002	0.0005	-	-	-
100	0.003	0	0	0.0015	0	0	-	-	-
250	0.0025	0.001	0	0.002	0	0	-	-	-
500	0.0045	0.0015	0	0.0035	5e-04	0	-	-	-
1000	0.003	0	0	0.0005	0	0	-	-	-
Panel B: Fama-MacBeth estimator in a model with only a useful factor, with Shanken correction									
30	0.01	0.0015	0.0005	0	0	0	-	-	-
50	0.0055	0.0015	0.001	0.0015	0	0	-	-	-
100	0.0015	0	0	0	0	0	-	-	-
250	0.002	0.001	0	0	0	0	-	-	-
500	0.004	0.0015	0	0.001	0	0	-	-	-
1000	0.002	0	0	0	0	0	-	-	-
Panel C: Fama-MacBeth estimator in a model with a useless factor, without Shanken correction									
30	0.0345	0.015	0.002	0.0225	0.0055	0.0015	0.135	0.055	0.016
50	0.04	0.017	0.0045	0.0315	0.01	0.004	0.2885	0.139	0.0465
100	0.0715	0.036	0.0145	0.06	0.027	0.01	0.5945	0.441	0.2375
250	0.18	0.11	0.0595	0.168	0.1015	0.0515	0.805	0.7595	0.696
500	0.305	0.2245	0.1605	0.297	0.2275	0.158	0.872	0.8425	0.806
1000	0.405	0.342	0.283	0.4145	0.347	0.28	0.932	0.915	0.8885
Panel D: Fama-MacBeth estimator in a model with a useless factor, with Shanken correction									
30	0.0175	0.0035	0.0005	0.001	0	0	0.0185	0.003	0.001
50	0.015	0.0055	0.0015	0.005	0.001	0	0.054	0.0065	0.0005
100	0.0175	0.0065	0.001	0.0055	0.001	0	0.249	0.044	0.002
250	0.028	0.0065	0.0015	0.011	0.002	0	0.672	0.3305	0.051
500	0.0445	0.018	0.006	0.0305	0.007	0.0015	0.8105	0.5845	0.2095
1000	0.055	0.022	0.009	0.0445	0.0155	0.0015	0.8895	0.731	0.3965

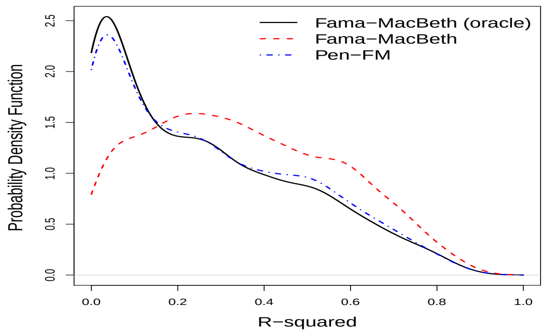
Note. The table summarises the empirical size of the t-statistic-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with an identity weight matrix at the second stage and at various significance levels (α=10%, 5%, 1%). The misspecified model includes only 1 out of 3 true risk factors, and is further contaminated by the presence of a useless factor. λ_0 stands for the value of the common intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. The pseudo-true values of the risk premia are defined as the limit of the risk premia estimates in a misspecified model without the influence of the useless factor. Panels A and C present the size of the t-statistic confidence bounds, computed using OLS-based heteroscedasticity-robust standard errors that do not take into account the error-in-variables problem of the second stage. The model is estimated with/without the useless factor. Panels B and D present similar results for the case of Shanken correction. The model is simulated 10,000 times for different values of the sample size (T).

For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

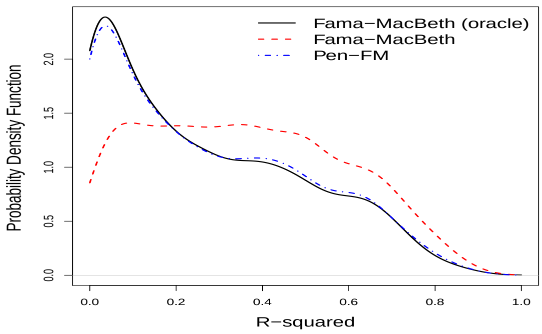
Figure 1.A.1: Distribution of the cross-sectional R^2 in a correctly specified model



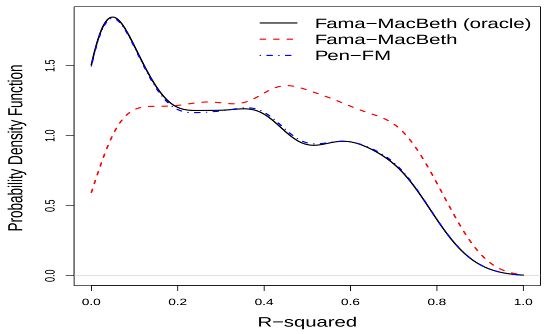
(a) T=30



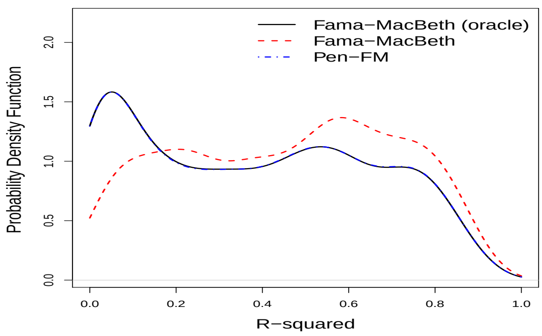
(b) T=50



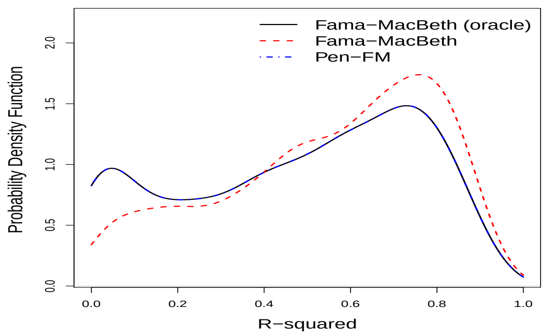
(c) T=100



(d) T=250



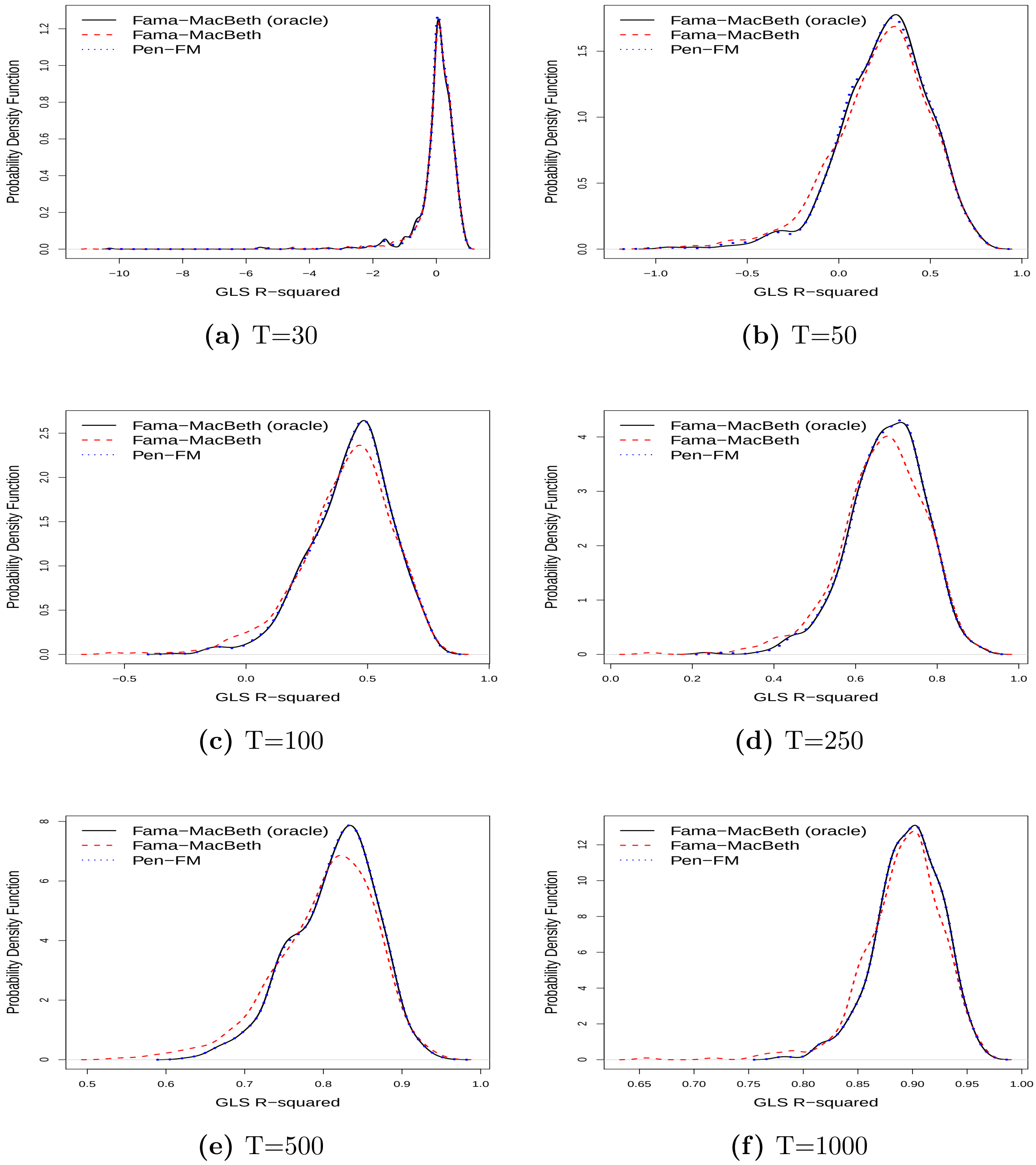
(e) T=500



(f) T=1000

Note. The graphs present the probability density function for the cross-sectional R-squared in a simulation of a correctly specified model, potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). For each of the sample sizes, the solid line represents the p.d.f. of the R-squared in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional R-squared when the model is estimated by the Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for the R^2 when the Pen-FM estimator is employed in the same scenario of the contaminated model. For a detailed description of the simulation design, please refer to Table 1.1.

Figure 1.A.2: Distribution of the cross-sectional $GLS\ R^2$ in a correctly specified model based on the OLS risk premia estimates in the second stage



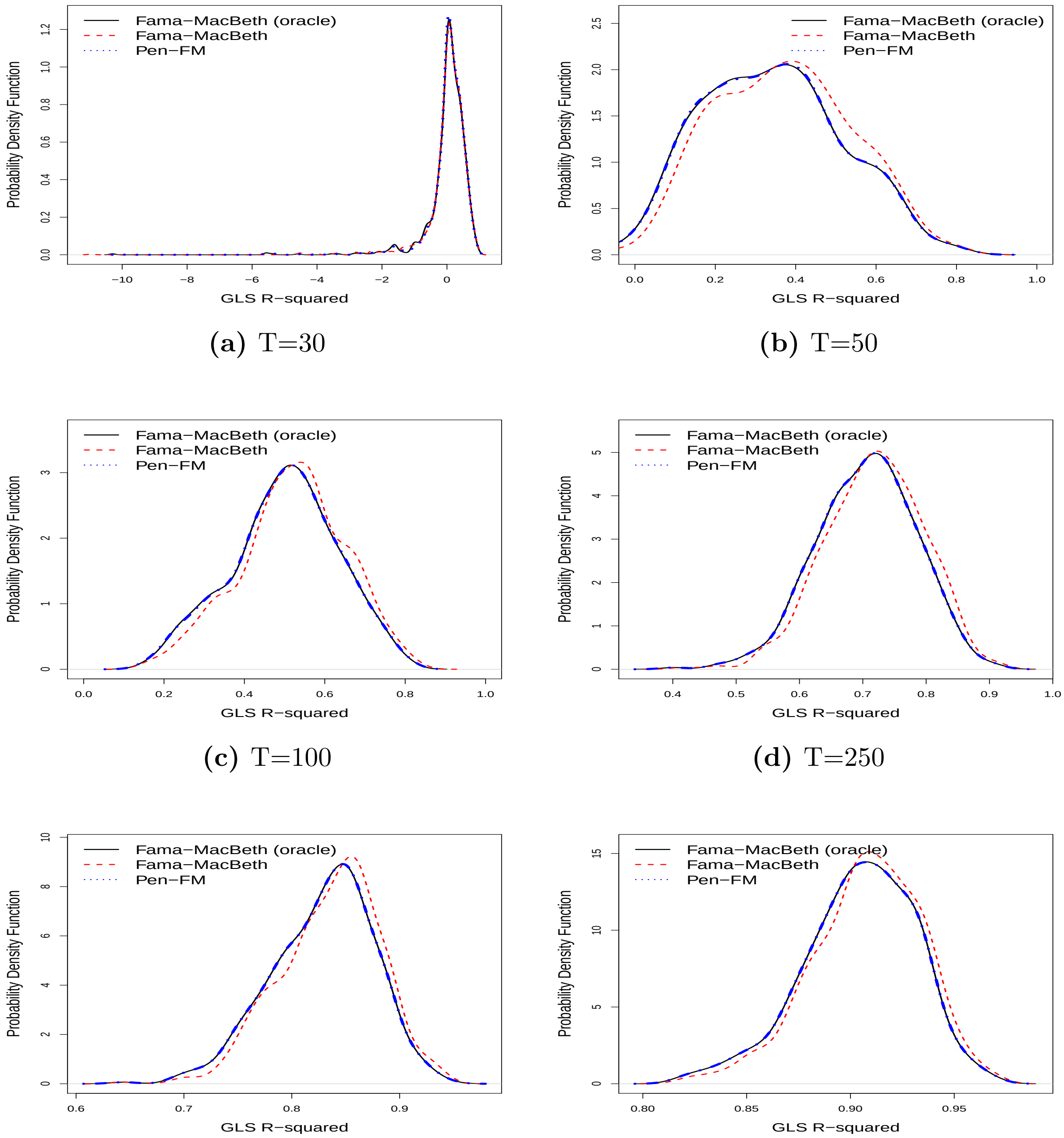
(a) T=30 **(b)** T=50

(c) T=100 **(d)** T=250

(e) T=500 **(f)** T=1000

Note. The graphs demonstrate the probability density function for the cross-sectional $GLS\ R^2$ in a simulation of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and estimated using an identity weight matrix on the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents p.d.f. of the $GLS\ R^2$ in the model without a useless factor, when risk premia are estimated by Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional $GLS\ R^2$ when the model is estimated by Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for the $GLS\ R^2$ when Pen-FM estimator is employed in the same scenario of the contaminated model. For a detailed description of the simulation design, please refer to Table 1.1.

Figure 1.A.3: Distribution of the cross-sectional GLS R^2 in a correctly specified model based on the GLS risk premia estimates in the second stage



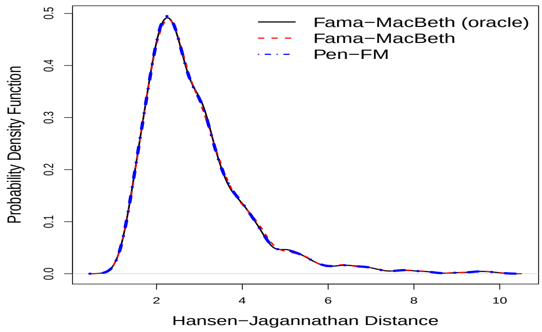
(a) T=30 **(b)** T=50

(c) T=100 **(d)** T=250

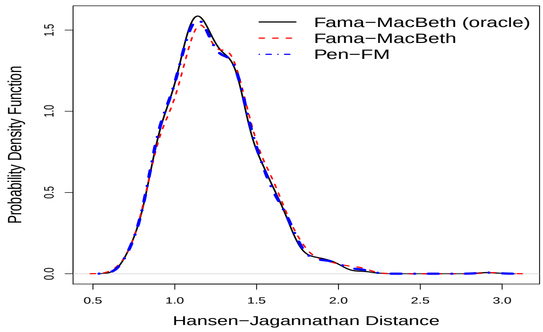
(e) T=500 **(f)** T=1000

Note. The graphs present the probability density function for the cross-sectional $GLS\ R^2$ in a simulation of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and estimated using the FGLS weight matrix on the second stage ($W = \hat{\Omega}^{-1}$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of the $GLS\ R^2$ in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional $GLS\ R^2$ when the model is estimated by Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for the $GLS\ R^2$ when the Pen-FM estimator is employed in the same scenario of the contaminated model. For a detailed description of the simulation design, please refer to Table 1.1.

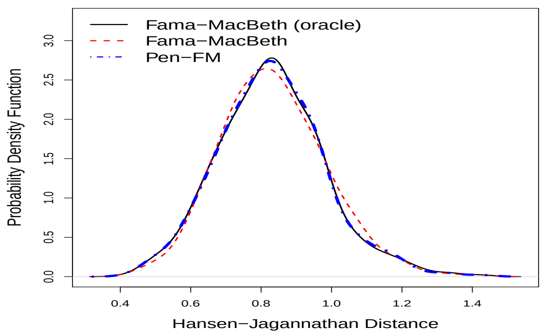
Figure 1.A.4: Distribution of the Hansen-Jagannathan distance in a correctly specified model



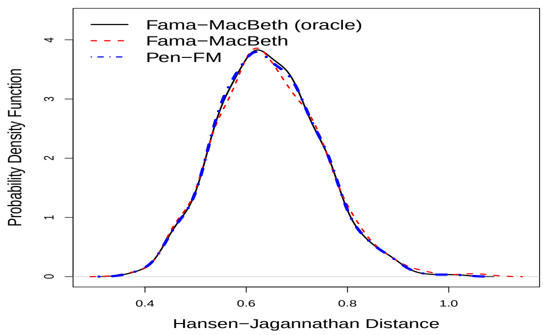
(a) T=30



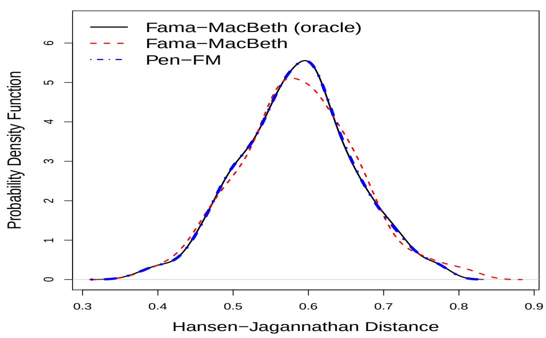
(b) T=50



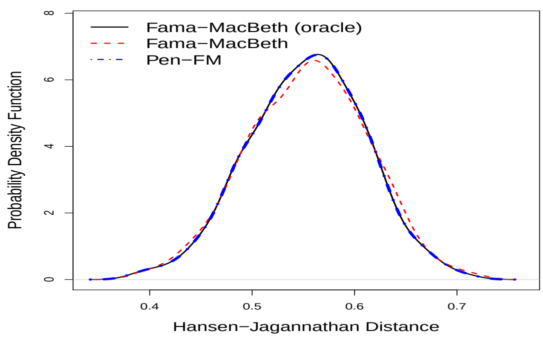
(c) T=100



(d) T=250



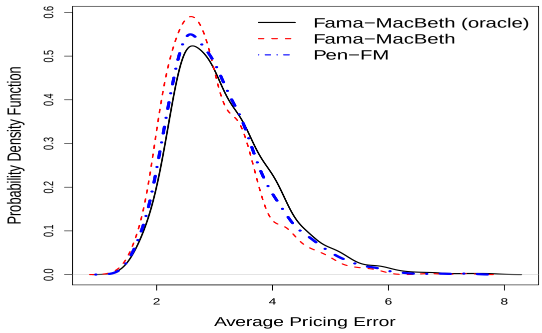
(e) T=500



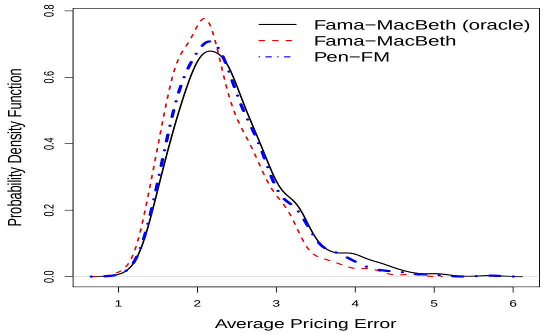
(f) T=1000

Note. The graphs present the probability density function for the Hansen-Jagannathan distance in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix on the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of HJ in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of HJ when the model is estimated by the Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for HJ when the Pen-FM estimator is employed in the same scenario of the contaminated model. For a detailed description of the simulation design, please refer to Table 1.1.

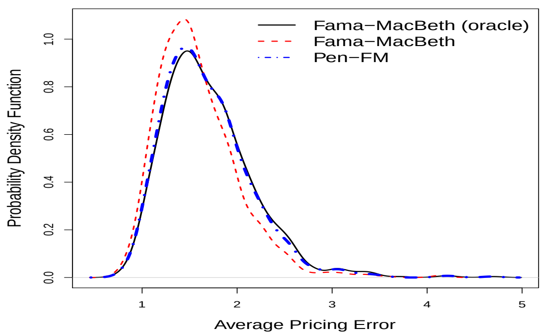
Figure 1.A.5: Distribution of the average pricing error in a correctly specified model



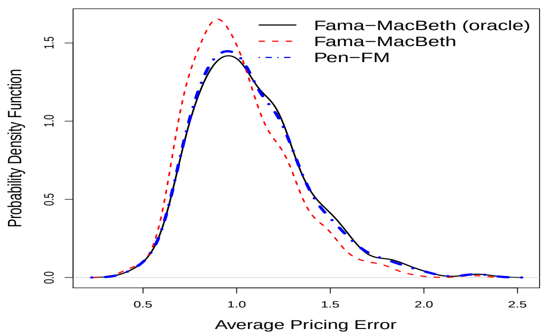
(a) T=30



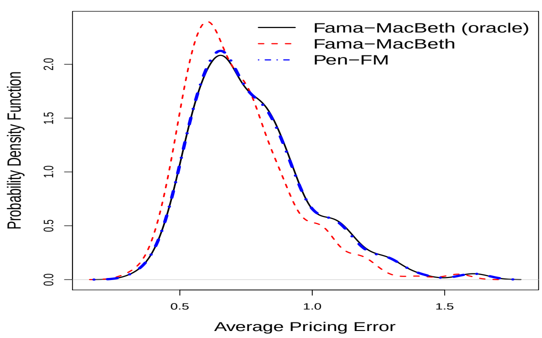
(b) T=50



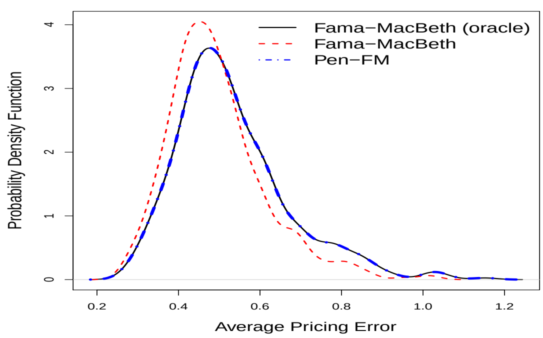
(c) T=100



(d) T=250



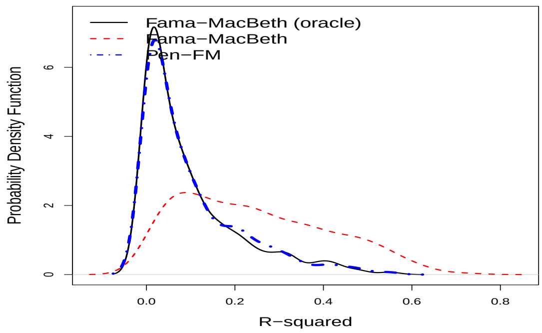
(e) T=500



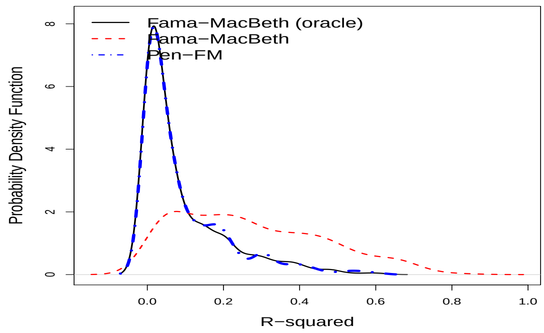
(f) T=1000

Note. The graphs present the probability density function for the average pricing error (APE) in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix on the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of the APE in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of APE when the model is estimated by Fama-MacBeth procedure, and a useless factor is included as well, while the dash-dotted line stands for the APE when the Pen-FM estimator is employed in the same scenario of the contaminated model. For a detailed description of the simulation design, please refer to Table 1.1.

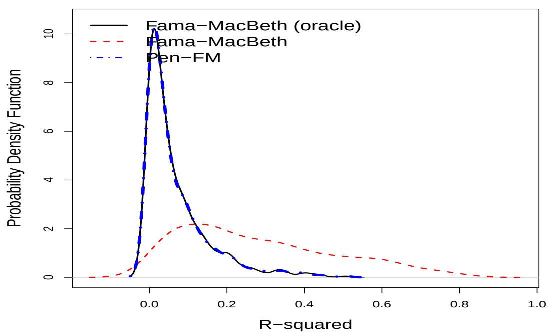
Figure 1.A.6: Distribution of the cross-sectional R^2 in a misspecified model



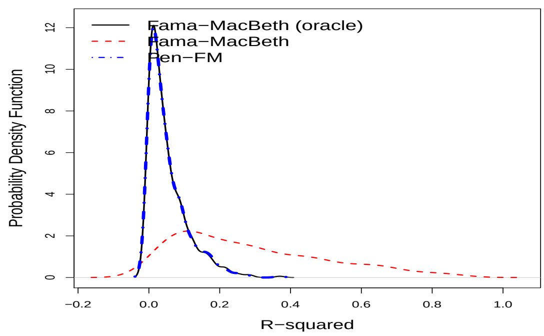
(a) T=30



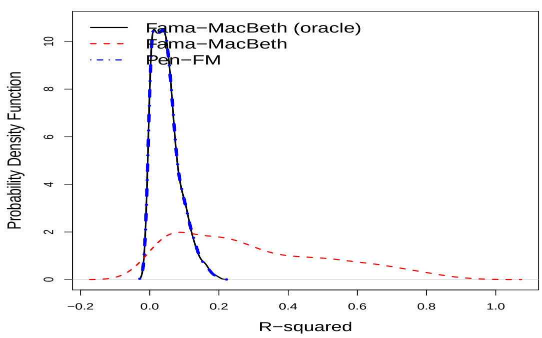
(b) T=50



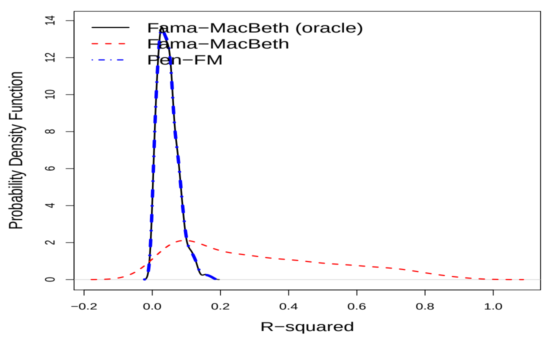
(c) T=100



(d) T=250



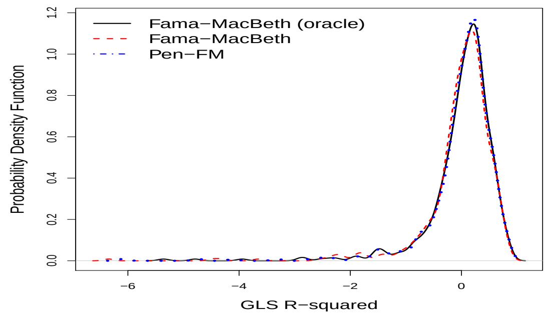
(e) T=500



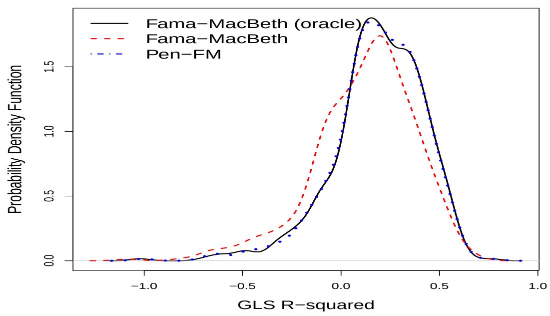
(f) T=1000

Note. The graphs present the probability density function for the cross-sectional R^2 in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents p.d.f. of the R^2 statistic in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional R^2 when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line stands for R^2 when the Pen-FM estimator is employed in the same scenario of the contaminated model. For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

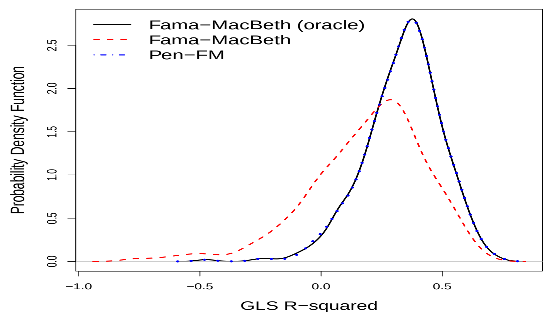
Figure 1.A.7: Distribution of the *GLS* R^2 in a misspecified model based on the OLS estimates of the risk premia in the second stage



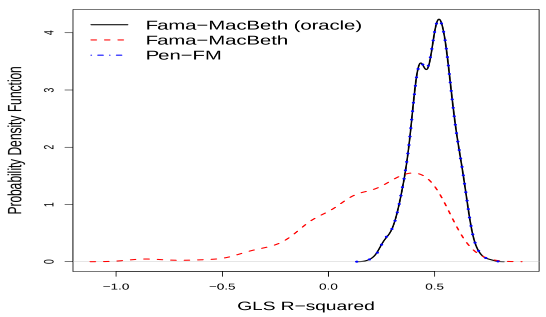
(a) T=30



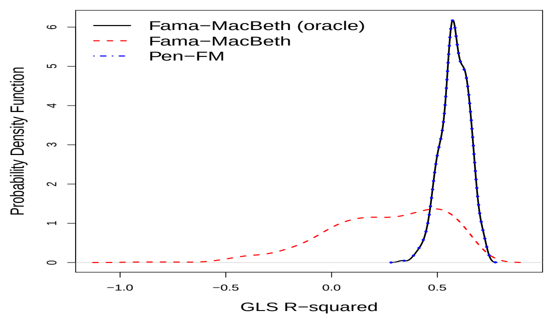
(b) T=50



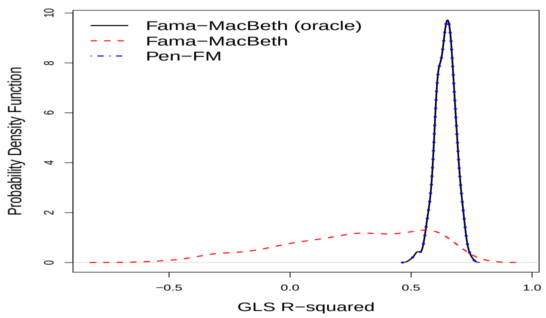
(c) T=100



(d) T=250



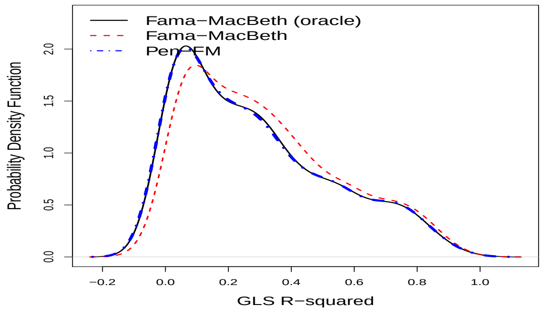
(e) T=500



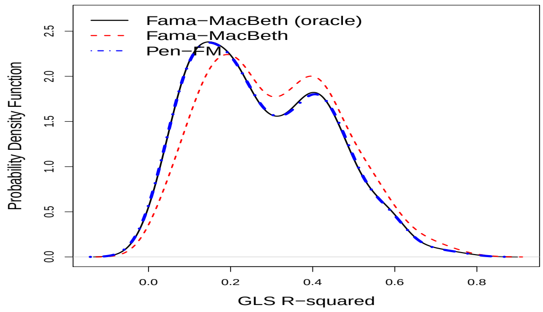
(f) T=1000

Note. The graphs illustrate the probability density function for the cross-sectional *GLS* R^2 in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of the *GLS* R^2 statistic in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional *GLS* R^2 when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line stands for R^2 when the Pen-FM estimator is employed in the same scenario of the contaminated model. For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

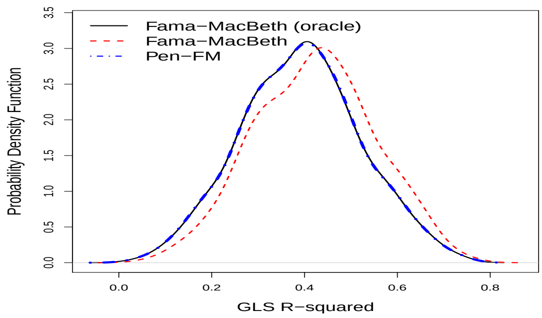
Figure 1.A.8: Distribution of the cross-sectional $GLS\ R^2$ in a misspecified model with risk premia estimates based on the GLS second stage



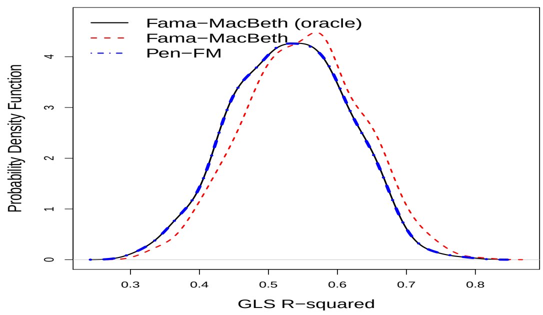
(a) T=30



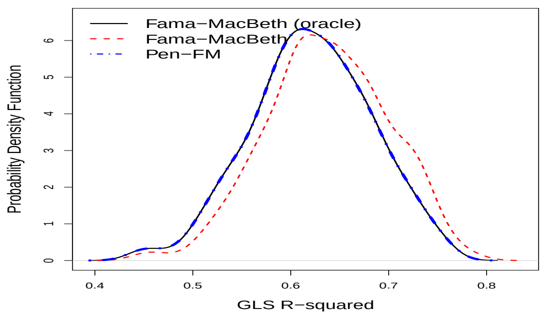
(b) T=50



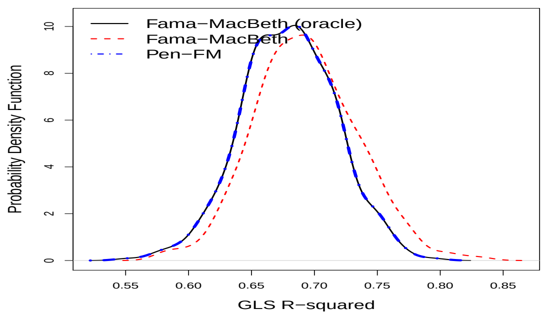
(c) T=100



(d) T=250



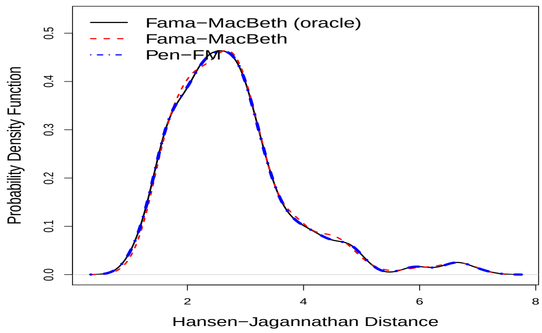
(e) T=500



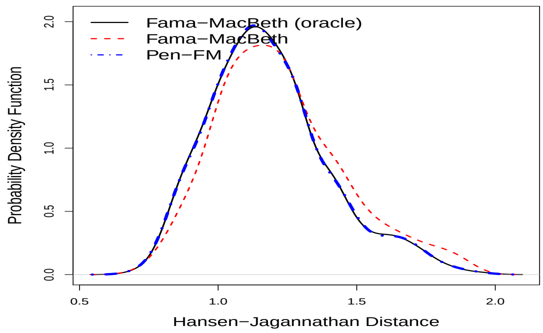
(f) T=1000

Note. The graphs present the probability density function for the cross-sectional $GLS\ R^2$ in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using the FGLS weight matrix ($W = \hat{\Omega}^{-1}$). For each of the sample sizes, the solid line represents the p.d.f. of the $GLS\ R^2$ statistic in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional $GLS\ R^2$ when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line corresponds to the case of the Pen-FM estimator employed in the same scenario of the contaminated model. For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

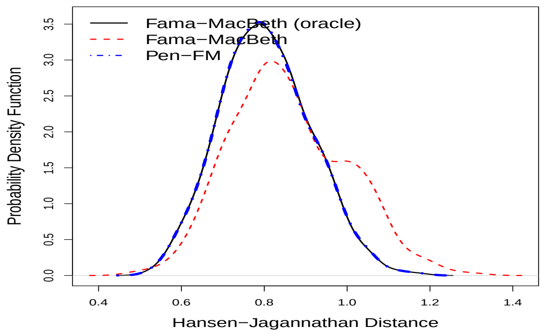
Figure 1.A.9: Distribution of the Hansen-Jagannathan distance in a misspecified model



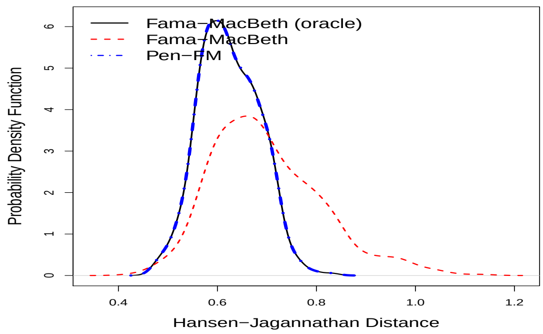
(a) T=30



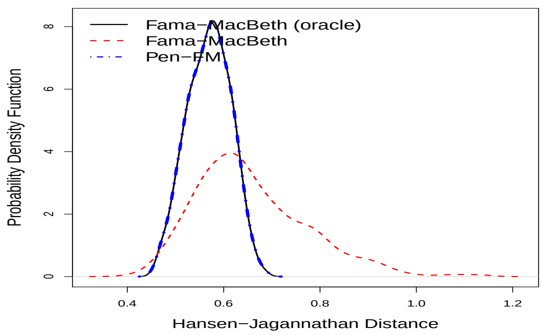
(b) T=50



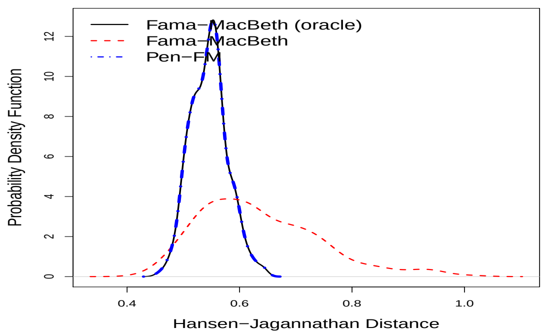
(c) T=100



(d) T=250



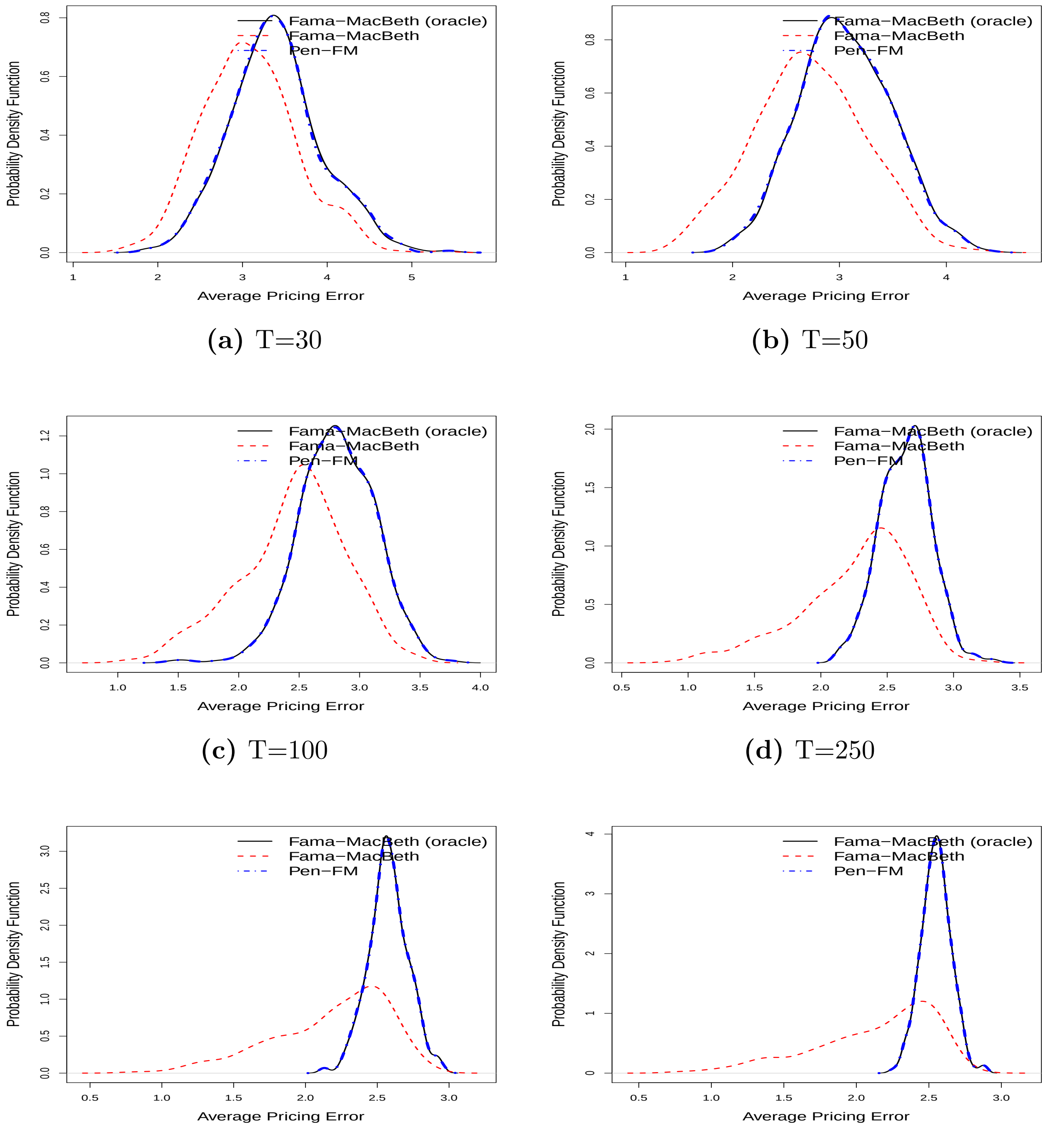
(e) T=500



(f) T=1000

Note. The graphs present the probability density function for the Hansen-Jagannathan distance (HJ) in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of HJ in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of HJ when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line corresponds to the case of the Pen-FM estimator employed in the same scenario of the contaminated model. For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

Figure 1.A.10: Distribution of the average pricing error in a misspecified model



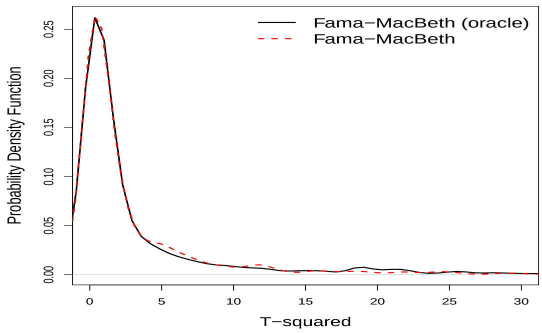
(a) T=30 **(b)** T=50

(c) T=100 **(d)** T=250

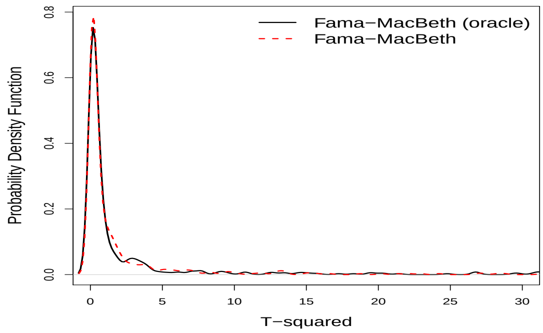
(e) T=500 **(f)** T=1000

Note. The graphs present the probability density function for the average pricing error (APE) in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of APE in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the APE when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line corresponds to the case of the Pen-FM estimator employed in the same scenario of the contaminated model. For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

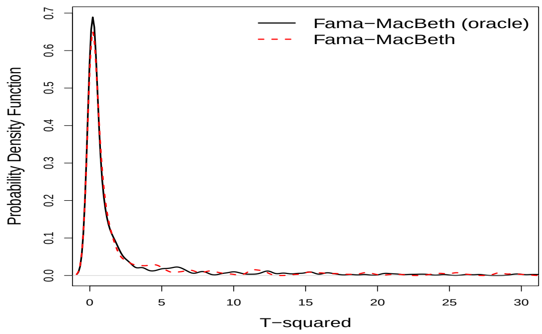
Figure 1.A.11: Distribution of the T^2 statistic in a correctly specified model



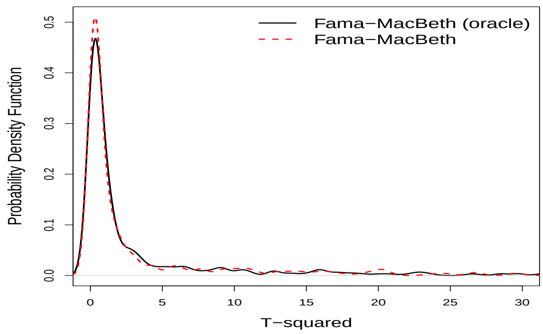
(a) T=30



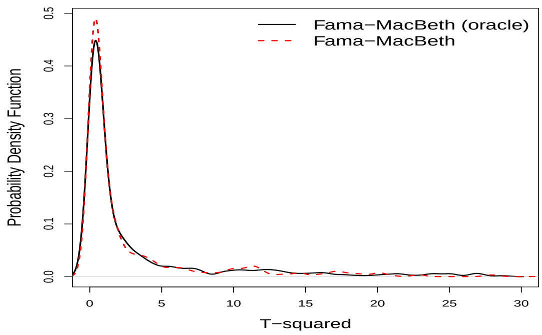
(b) T=50



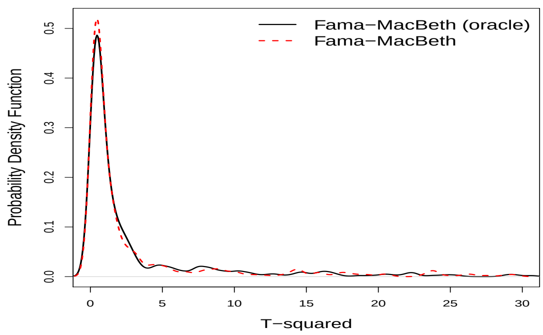
(c) T=100



(d) T=250



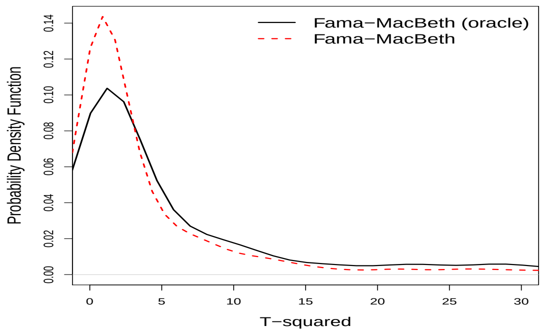
(e) T=500



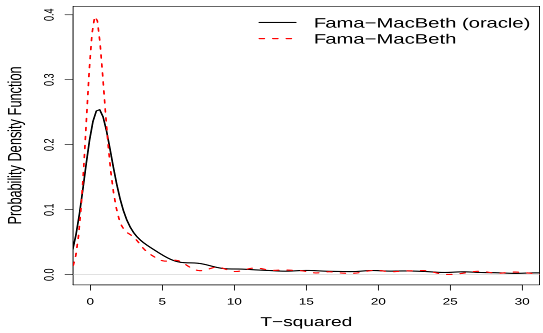
(f) T=1000

Note. The graphs present the probability density function for the T^2 statistic in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix in the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of the T^2 in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of T^2 when the model is estimated by the Fama-MacBeth procedure in the presence of a useless factor. For a detailed description of the simulation design, please refer to Table 1.1.

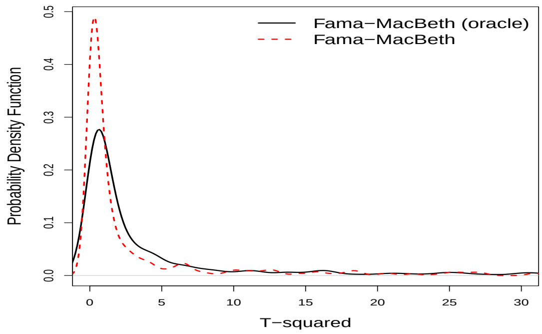
Figure 1.A.12: Distribution of the T^2-statistic in a misspecified model



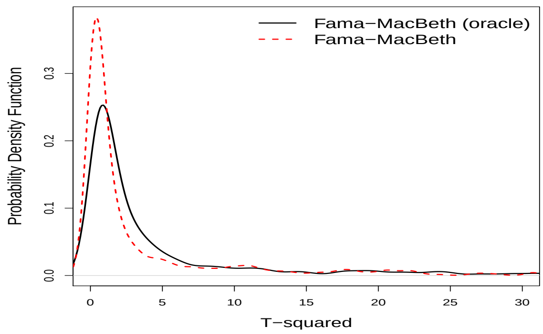
(a) T=30



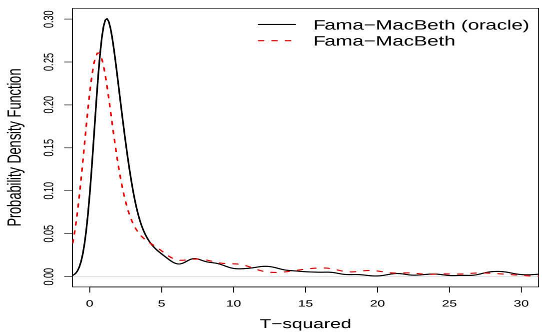
(b) T=50



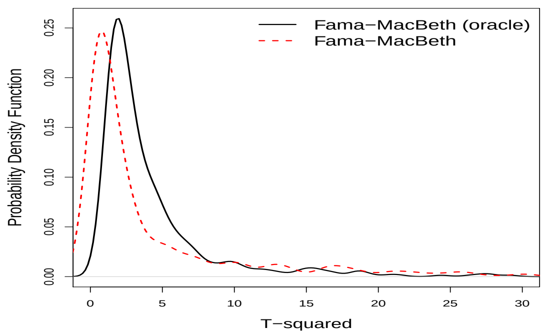
(c) T=100



(d) T=250



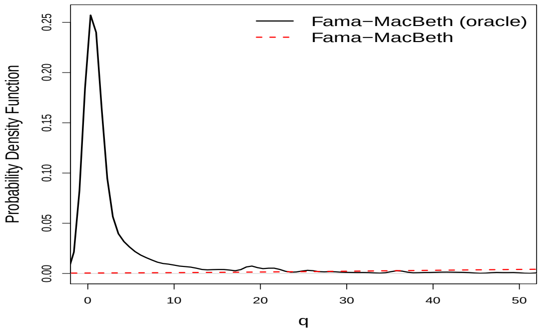
(e) T=500



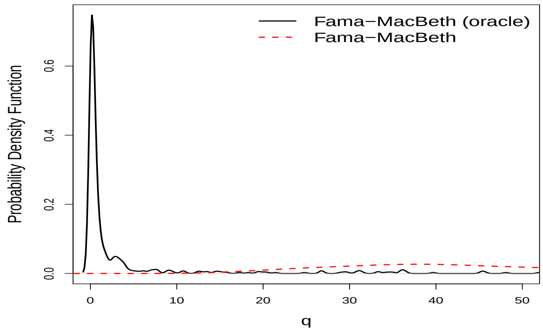
(f) T=1000

Note. The graphs present the probability density function for the T^2-statistic in a simulation of a misspecified model, potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents p.d.f. of T^2 in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of T^2 when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor. For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

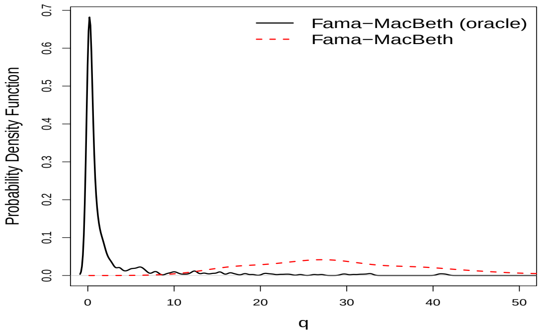
Figure 1.A.13: Distribution of the q-statistic in a correctly specified model



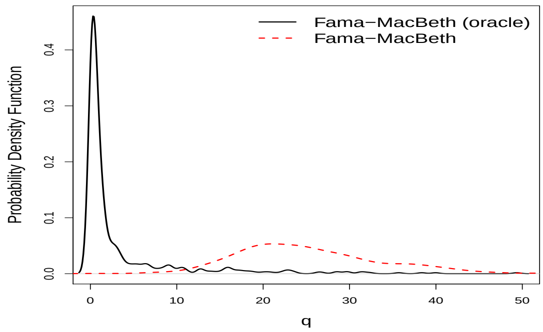
(a) T=30



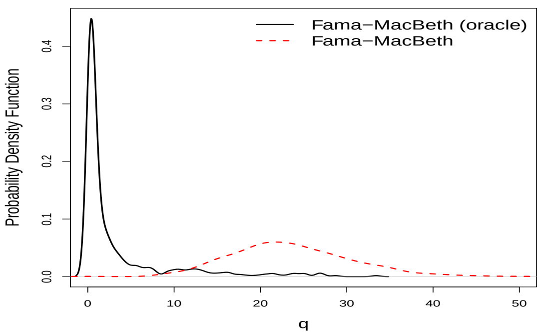
(b) T=50



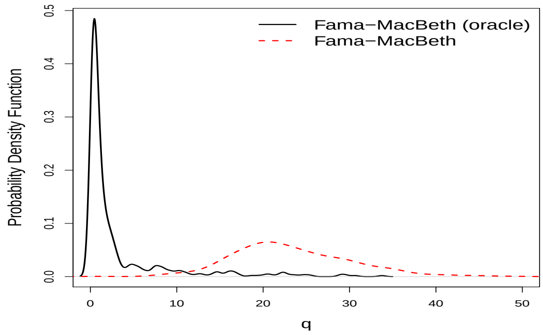
(c) T=100



(d) T=250



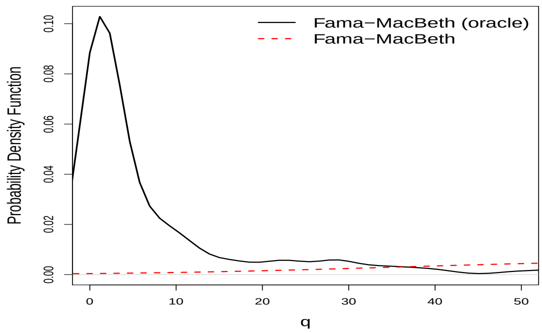
(e) T=500



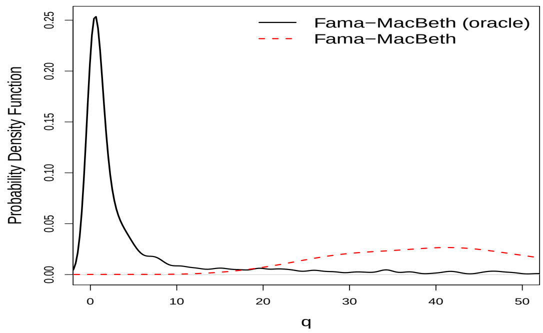
(f) T=1000

Note. The graphs present the probability density function of the q-statistic in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix in the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of q in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), and the dashed line depicts the distribution of q when the model is estimated by the Fama-MacBeth procedure under the presence of a useless factor. For a detailed description of the simulation design, please refer to Table 1.1.

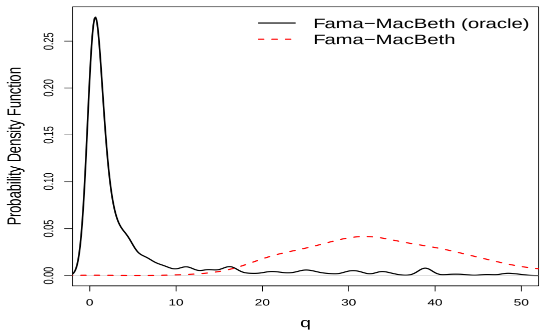
Figure 1.A.14: Distribution of the q-statistic in a misspecified model



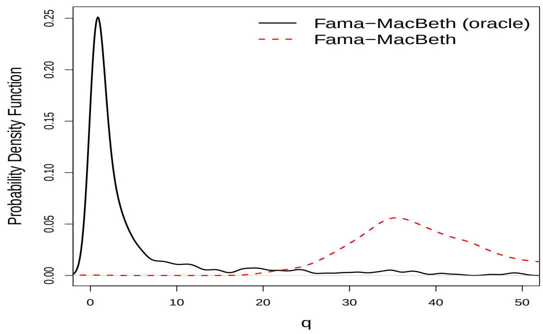
(a) T=30



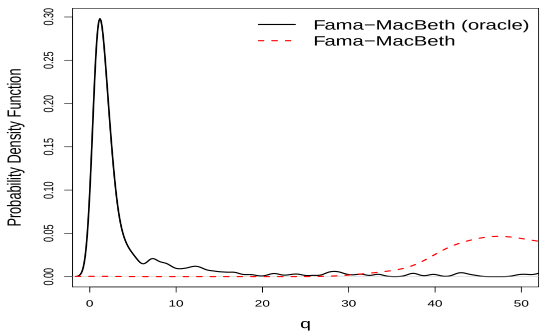
(b) T=50



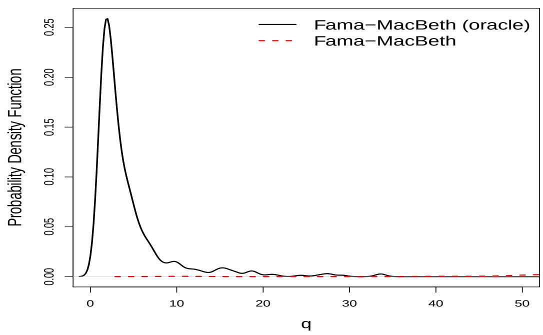
(c) T=100



(d) T=250



(e) T=500



(f) T=1000

Note. The graphs present the probability density function for q-statistic in a simulation of a misspecified model, potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of q in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of q when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor. For a detailed description of the simulation design for the misspecified model, please refer to Table 1.2.

1.B Proofs

1.B.1 Proof of Proposition 1.1

Consider the quadratics in the objective function.

$$\left[\bar{R} - \hat{\beta}\lambda\right]' W_T \left[\bar{R} - \hat{\beta}\lambda\right] \xrightarrow{p} [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}]' W [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}]$$

For the strong factors that have substantial covariance with asset returns (whether their risk is priced or not), $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \overset{a}{\sim} \eta T^{-d/2} O_p(1) \xrightarrow{d} 0$, where $\overset{a}{\sim}$ denotes equivalence of the asymptotic expansion up to $o_p\left(\frac{1}{\sqrt{T}}\right)$. For the useless factors we have $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \overset{a}{\sim} \eta T^{-d/2} c_j T^{d/2} \xrightarrow{d} \widetilde{c}_j > 0$. Therefore, in the limit the objective function becomes the following convex function of λ:

$$[\mathbb{E}[R] - \beta_{ns}\lambda_{ns}]' W [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}] + \sum_{j=1}^{k} \widetilde{c}_j |\lambda_j| \mathbb{1}\{\beta_j = 0\}$$

Since c_j are some positive constants,

$$\mathbf{0} = \arg\min_{\lambda_{sp} \in \Theta_{sp}} [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}]' W [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}] + \sum_{j=1}^{k} \widetilde{c}_j \, |\lambda_j| \, \mathbb{1}\{\beta_j = 0\}$$

The risk premia for the strong factors are still identified, as

$$\begin{aligned} \lambda_{0,ns} = \arg\min_{\lambda_{ns} \in \Theta_{ns}} \; [E[R] - \beta_{ns}\lambda_{ns}]' W [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}] &= \left(\beta_{ns}' W \beta_{ns}\right)^{-1} \beta_{ns}' W \mathbb{E}[R] \\ &= \left(\beta_{ns}' W \beta_{ns}\right)^{-1} \beta_{ns}' W \beta_{ns} \lambda_{0,ns} \end{aligned}$$

By the convexity lemma of Pollard (1991), the estimator is consistent.

To establish asymptotic normality, it is first instructive to show the distribution of the usual Fama-McBeth estimator in the absence of identification failure.

Following Lemma 1.1, the first stage estimates have the following asymptotic representations

$$\hat{\beta}_{ns} = \beta_{ns} + \frac{1}{\sqrt{T}} \Psi_{\beta,ns} + o_p\left(\frac{1}{\sqrt{T}}\right), \qquad \bar{R} = \beta_{ns}\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}} \lambda_{0,sp} + \frac{1}{\sqrt{T}} \psi_R + o_p\left(\frac{1}{\sqrt{T}}\right)$$

where $\Psi_{\beta,ns} = vecinv(\psi_{\beta,ns})$ and $vecinv$ is the inverse of the vectorisation operator.

Consider the WLS estimator of the cross-section regression:

$$\begin{aligned}
\hat{\lambda}_{ns} &= \left(\hat{\beta}'_{ns} W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}'_{ns} W_T \bar{R} \overset{a}{=} \left(\hat{\beta}'_{ns} W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}'_{ns} \left(\beta_{ns}\lambda_{0,ns} + \frac{1}{\sqrt{T}}\psi_R\right) \\
&= \left(\hat{\beta}'_{ns} W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}'_{ns} W_T \left(\hat{\beta}_{ns}\lambda_{0,ns} + (\beta_{ns} - \hat{\beta}_{ns})\lambda_{0,ns} + \frac{1}{\sqrt{T}}\psi_R\right) = \\
&= \lambda_{0,ns} + \left(\hat{\beta}'_{ns} W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}'_{ns} W_T (\hat{\beta}_{ns} - \beta_{ns})\lambda_{0,ns} + \left(\hat{\beta}'_{ns} W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}'_{ns} W_T \frac{1}{\sqrt{T}}\psi_R
\end{aligned}$$

Finally, since as $T \to \infty$

$$\begin{aligned}
\hat{\beta}'_{ns} W_T \hat{\beta}_{ns} &\overset{a}{=} \left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right]' W_T \left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right] \xrightarrow{p} \beta'_{ns} W \beta_{ns} \\
\hat{\beta}_{ns} - \beta_{ns} &\overset{a}{=} -\frac{1}{\sqrt{T}}\Psi_{\beta,ns} = \frac{1}{\sqrt{T}}\Psi_{\beta,ns}
\end{aligned}$$

it follows that

$$\sqrt{T}(\hat{\lambda}_{ns} - \lambda_{0,ns}) \xrightarrow{d} \left[\beta'_{ns} W \beta_{ns}\right]^{-1} \beta'_{ns} W \Psi_{\beta,ns}\lambda_{0,ns} + \left(\beta'_1 W \beta_1\right)^{-1} \beta'_{ns} W \psi_R$$

In order to demonstrate the asymptotic distribution of the shrinkage-based estimator, I reformulate the objective function in terms of the centred parameters $u = \frac{\lambda - \lambda_0}{\sqrt{T}}$:

$$L_T(u) = \left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T \left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right] + \eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}^j\right\|_1^d} \left|\lambda_{0j} + \frac{u}{\sqrt{T}}\right|$$

Solving the original problem in 1.9 w.r.t. λ is the same as optimizing $\bar{L}(u) = T\left(L_T(u) - L_T(0)\right)$ w.r.t. u.

Since

$$\begin{aligned}
&\left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T \left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right] = \\
&= \bar{R}' W_T \bar{R} + \left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \hat{\beta} \left[\lambda_0 + \frac{u}{\sqrt{T}}\right] - 2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \bar{R} \\
&\lambda'_0 \hat{\beta}' W_T \hat{\beta}\lambda_0 + \frac{u'}{\sqrt{T}} \hat{\beta}' W_T \hat{\beta} \frac{u}{\sqrt{T}} + \frac{2}{\sqrt{T}} u' \hat{\beta}' W_T \hat{\beta}\lambda_0 - 2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \bar{R} \\
&= \lambda'_0 \hat{\beta}' W_T \hat{\beta}\lambda_0 + \frac{u'}{\sqrt{T}} \hat{\beta}' W_T \hat{\beta} \frac{u}{\sqrt{T}} - 2\lambda'_0 \hat{\beta}' W_T \bar{R} + \frac{2}{\sqrt{T}} u' \hat{\beta}' W_T (\hat{\beta}\lambda_0 - \bar{R})
\end{aligned}$$

Therefore, in localized parameters u the problem looks as follows:

$$\hat{u} = \underset{u \in K}{\arg\min} \;\; u'\hat{\beta}'W_T\hat{\beta}u + 2\sqrt{T}u'\hat{\beta}'W_T(\hat{\beta}\lambda_0 - \bar{R}) + T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}_j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$
$$= \underset{u \in K}{\arg\min} \;\; u'\hat{\beta}'W_T\hat{\beta}u + 2\sqrt{T}\hat{\beta}'W_T(\hat{\beta} - \beta)\lambda_0 - 2u'\hat{\beta}'W_T\varphi_R +$$
$$+T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}_j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$

where K is a compact set in $\mathbb{R}^k$.

It is easy to show that since as $t \to \infty$

$$\hat{\beta}'W_T\hat{\beta} \stackrel{a}{=} \begin{bmatrix} \beta'_{ns} + \frac{1}{\sqrt{T}}\Psi'_{\beta,ns} \\ \frac{1}{\sqrt{T}}\Psi'_{\beta,sp} \end{bmatrix} W_T \begin{bmatrix} \beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns} & \frac{1}{\sqrt{T}}\Psi_{\beta,sp} \end{bmatrix} \stackrel{a}{=} \begin{bmatrix} \beta'_{ns}W\beta_{ns} & 0 \\ 0 & 0 \end{bmatrix}$$

the following identities hold:

$$u'\hat{\beta}'W_T\hat{\beta}u \stackrel{a}{=} \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \beta'_{ns}W\beta_{ns} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{ns} \\ u_{sp} \end{bmatrix} = u'_{ns}\left[\beta'_{ns}W\beta_{ns}\right]u_{ns} \; ,$$
$$u'\hat{\beta}'W_T(\hat{\beta} - \beta)\lambda_0 \stackrel{a}{=} \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \beta'_{ns} + \frac{1}{\sqrt{T}}\Psi'_{\beta,ns} \\ \frac{1}{\sqrt{T}}\Psi'_{\beta,sp} \end{bmatrix} W \begin{bmatrix} \frac{1}{\sqrt{T}}\Psi_{\beta,ns} & \frac{1}{\sqrt{T}}\Psi_{\beta,sp} \end{bmatrix} \begin{bmatrix} \lambda_{0,ns} \\ 0 \end{bmatrix} \stackrel{a}{=}$$
$$= \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{T}}\beta'_{ns}W\Psi_{\beta,ns} & \frac{1}{\sqrt{T}}\beta'_{ns}W\Psi_{\beta,sp} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{0,ns} \\ 0 \end{bmatrix} = \frac{1}{\sqrt{T}}u'_{ns}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} \; ,$$
$$u'\hat{\beta}'W_T\varphi_R \stackrel{a}{=} \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \beta'_{ns} + \frac{1}{\sqrt{T}}\Psi'_{\beta,ns} \\ \frac{1}{\sqrt{T}}\Psi'_{\beta,sp} \end{bmatrix} W\varphi_R \; .$$

Finally, this implies that the overall objective function asymptotically looks as follows:

$$\bar{L}_T(u) \stackrel{a}{=} u'_{ns}\left[\beta'_{ns}W\beta_{ns}\right]u_{ns} + 2u'_{ns}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} - 2u'_{ns}(\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns})'W\varphi_R$$
$$-\frac{2}{\sqrt{T}}u'_{sp}\Psi'_{\beta,sp}W\varphi_R + T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}^j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] \stackrel{a}{=}$$
$$= u'_{ns}\beta'_{ns}W\beta_{ns}u_{ns} - 2u'_{ns}\beta'_{ns}W(\varphi_R - \Psi_{\beta,ns}\lambda_{0,ns}) + T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}^j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$

Now, for a spurious factor: $T\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d}\left[\left|\lambda_{0j}+\frac{u_j}{\sqrt{T}}\right|-|\lambda_{0j}|\right]=\sqrt{T}\eta T^{-d/2}c_jT^{d/2}\,|u_j|=\sqrt{T}\widetilde{c}_j\,|u_j|$,

while for the strong ones: $T\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d}\left[\left|\lambda_{0j}+\frac{u_j}{\sqrt{T}}\right|-|\lambda_{0j}|\right]=c_j\sqrt{T}T^{-d/2}u_j sgn(\lambda_{0j})\to 0$, since $d>2$.

Therefore, as $T\to\infty$, $\bar{L}_T(u)\xrightarrow{d}\bar{L}_n$ for every u, where

$$\bar{L}(u)=\begin{cases} -u_{ns}^{'}\beta_{ns}^{'}W\beta_{ns}u_{ns}-2u_{ns}^{'}W\beta_{ns}\left(\phi_R-\Psi_{\beta,ns}\lambda_{0,ns}\right) & \text{if } u_{sp}=0\\ \infty & \text{otherwise}\end{cases}$$

Note that $\bar{L}_T(u)$ is a convex function with a unique optimum given by

$$\left(\left[\beta_{ns}'W\beta_{ns}\right]^{-1}\beta_{ns}'W\Psi_{\beta,ns}\lambda_{0,ns}+\left[\beta_{ns}'W\beta_{ns}\right]^{-1}\beta_{ns}'W\psi_R,\,0\right)'.$$

Therefore, due to the epiconvergence results of Pollard (1994) and Knight and Fu (2000), we have that

$$\hat{u}_{ns}\xrightarrow{d}\left[\beta_{ns}'W\beta_{ns}\right]^{-1}\beta_{ns}'W\Psi_{\beta,ns}\lambda_{0,ns}+\left[\beta_{ns}'W\beta_{ns}\right]^{-1}\beta_{ns}'W\psi_R\,,$$
$$\hat{u}_{sp}\xrightarrow{d}0\,.$$

Hence, the distribution of the risk premia estimates for the useful factors coincides with the one without the identification problem. Therefore, Pen-FM exhibits the so-called *oracle* property.

1.B.2 Proof of Proposition 1.2

I am going to prove consistency first. Consider the objective function. As $T\to\infty$

$$\left[\bar{R}-\hat{\beta}\lambda\right]'W_T\left[\bar{R}-\hat{\beta}\lambda\right]\xrightarrow{p}[\mathbb{E}\left[R\right]-\beta_{ns}\lambda_{ns}]'W[\mathbb{E}\left[R\right]-\beta_{ns}\lambda_{ns}]$$

Also note that for the strong factors $\eta_T\frac{1}{\|\hat{\beta}_j\|_1^d}\sim\eta T^{-d/2}O_p(1)\to 0$, while for the weak ones $\eta_T\frac{1}{\|\hat{\beta}_j\|_1^d}\sim\eta T^{-d/2}c_jT^{d/2}\to\widetilde{c}_j>0$.

Therefore, the limit objective function becomes

$$[\mathbb{E}\left[R\right]-\beta_{ns}\lambda_{ns}]'W[\mathbb{E}\left[R\right]-\beta_{ns}\lambda_{ns}]+\sum_{j=1}^{k}\widetilde{c}_j\,|\lambda_j|\,\mathbb{1}\left\{\beta_j=O_p\left(\frac{1}{\sqrt{T}}\right)\right\}$$

Since $\widetilde{c}_j$ are positive constants,

$$0=\underset{\lambda_{sp}\in\Theta_{sp}}{\arg\min}[\mathbb{E}\left[R\right]-\beta_{ns}\lambda_{ns}]'W[\mathbb{E}\left[R\right]-\beta_{ns}\lambda_{ns}]+\sum_{j=1}^{k}\widetilde{c}_j\,|\lambda_j|\,\mathbb{1}\left\{\beta_j=O_p\left(\frac{1}{\sqrt{T}}\right)\right\}$$

However, the risk premia for the strong factors are still strongly identified, since

$$\arg\min_{\lambda_j \in \Theta_{ns}} [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}]' W [\mathbb{E}[R] - \beta_{ns}\lambda_{ns}] = \lambda_{0,ns} + \frac{1}{\sqrt{T}} \left(\beta_{ns}' W \beta_{ns}\right)^{-1} \beta_{ns}' W B_{sp} \lambda_{0,sp} \to \lambda_{0,ns}$$

Therefore, once again, due to the convexity lemma of Pollard (1991), the estimator is consistent.

Again, I first demonstrate the asymptotic distribution in the usual Fama-McBeth estimator in the absence of weak factors. Recall that

$$\hat{\beta}_{ns} = \beta_{ns} + \frac{1}{\sqrt{T}} \Psi_{\beta,ns} + o_p\left(\frac{1}{\sqrt{T}}\right), \qquad \bar{R} = \beta_{ns}\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} + \frac{1}{\sqrt{T}}\psi_R + o_p\left(\frac{1}{\sqrt{T}}\right)$$

where $\Psi_{\beta,ns} = vecinv(\psi_{\beta,ns})$.

Therefore, the second stage estimates have the following asymptotic expansion

$$\begin{aligned}
\hat{\lambda}_{ns} = \left(\hat{\beta}_{ns}' W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}_{ns}' W_T \bar{R} &\overset{a}{=} \left(\hat{\beta}_{ns}' W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}_{ns}' \left(\beta_{ns}\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} + \frac{1}{\sqrt{T}}\psi_R\right) = \\
&= \left(\hat{\beta}_{ns}' W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}_{ns}' W_T \left(\hat{\beta}_{ns}\lambda_{0,ns} + (\beta_{ns} - \hat{\beta}_{ns})\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} + \frac{1}{\sqrt{T}}\psi_R\right) = \\
&= \lambda_{0,ns} + \left(\hat{\beta}_{ns}' W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}_{ns}' W_T (\hat{\beta}_{ns} - \beta_{ns})\lambda_{0,ns} + \left(\hat{\beta}_{ns}' W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}_{ns}' W_T \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} \\
&\quad + \left(\hat{\beta}_{ns}' W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}_{ns}' W_T \frac{1}{\sqrt{T}}\psi_R
\end{aligned}$$

Finally, since

$$\hat{\beta}_{ns}' W_T \hat{\beta}_{ns} \overset{a}{=} \left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right]' W_T \left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right] \xrightarrow{p} \beta_{ns}' W \beta_{ns}$$
$$\hat{\beta}_{ns} - \beta_{ns} \overset{a}{=} -\frac{1}{\sqrt{T}}\Psi_{\beta,ns} = \frac{1}{\sqrt{T}}\Psi_{\beta,ns}$$

we get

$$\sqrt{T}(\hat{\lambda}_{ns} - \lambda_{0,ns}) \xrightarrow{d}$$
$$\left[\beta_{ns}' W \beta_{ns}\right]^{-1} \beta_{ns}' W \Psi_{\beta,ns}\lambda_{0,ns} + \left(\beta_1' W \beta_1\right)^{-1} \beta_{ns}' W \psi_R + \left(\hat{\beta}_{ns}' W_T \hat{\beta}_{ns}\right)^{-1} \hat{\beta}_{ns}' W_T B_{sp}\lambda_{0,sp}$$

The asymptotic distribution of risk premia estimates has three components:

- $\left[\beta_{ns}' W \beta_{ns}\right]^{-1} \beta_{ns}' W \Psi_{\beta,ns}\lambda_{0,ns}$, which arises due to the error-in-variables problem, since we observe not the true values of betas, but only their estimates, i.e. the origin for Shanken (1992) correction;
- $\left(\beta_1' W \beta_1\right)^{-1} \beta_{ns}' W \psi_R$, which corresponds to the usual sampling error, associated with the

WLS estimator;

- $\left(\hat{\beta}_{ns}'W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}_{ns}'W_TB_{sp}\lambda_{0,sp}$, which is the $\frac{1}{\sqrt{T}}$ omitted variable bias, due to eliminating potentially priced weak factors from the model.

Similar to the previous case, in order show the asymptotic distribution of the Pen-FM estimator, I rewrite the objective function in terms of the localised parameters, $u = \frac{\lambda-\lambda_0}{\sqrt{T}}$, as follows:

$$\hat{u} = \underset{u\in K}{\arg\min}\ u'\hat{\beta}'W_T\hat{\beta}u + 2\sqrt{T}u'\hat{\beta}'W_T(\hat{\beta}\lambda_0 - \bar{R}) + T\eta_T\sum_{j=1}^{k}\frac{1}{\left\|\hat{\beta}_j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] ,$$

since

$$\left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T\left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right] =$$

$$= \bar{R}'W_T\bar{R} + \left[\lambda_0 + \frac{u}{\sqrt{T}}\right]'\hat{\beta}'W_T\hat{\beta}\left[\lambda_0 + \frac{u}{\sqrt{T}}\right] - 2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]'\hat{\beta}'W_T\bar{R} ,$$

$$\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]'\hat{\beta}'W_T\hat{\beta}\left[\lambda_0 + \frac{u}{\sqrt{T}}\right] = \lambda_0'\hat{\beta}'W_T\hat{\beta}\lambda_0 + \frac{u'}{\sqrt{T}}\hat{\beta}'W_T\hat{\beta}\frac{u}{\sqrt{T}} + \frac{2}{\sqrt{T}}u'\hat{\beta}'W_T\hat{\beta}\lambda_0$$

$$-2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]'\hat{\beta}'W_T\bar{R} = -2\lambda_0'\hat{\beta}'W_T\bar{R} - \frac{2}{\sqrt{T}}u'\hat{\beta}'W_T\bar{R} .$$

Recall that

$$\hat{\beta}'W_T\hat{\beta} \overset{a}{=} \begin{bmatrix}\beta_{ns}' + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}' \\ \frac{1}{\sqrt{T}}(B_{sp}' + \Psi_{\beta,sp}')\end{bmatrix} W_T\begin{bmatrix}\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns} & \frac{B_{sp}}{\sqrt{T}} + \frac{1}{\sqrt{T}}\Psi_{\beta,sp}\end{bmatrix} \overset{a}{=}$$

$$= \begin{bmatrix}\beta_{ns}'W\beta_{ns} + \frac{2}{\sqrt{T}}\Psi_{\beta,ns}'W\beta_{ns} & \frac{1}{\sqrt{T}}\beta_{ns}'W(B_{sp} + \Psi_{\beta,sp}) \\ \frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,sp})W\beta_{ns} & 0\end{bmatrix} .$$

Hence,

$$u'\hat{\beta}'W_T\hat{\beta}u \overset{a}{=} \begin{bmatrix}u_{ns}' & u_{sp}'\end{bmatrix}\begin{bmatrix}\beta_{ns}'W\beta_{ns} + \frac{2}{\sqrt{T}}\Psi_{\beta 1}'W\beta_{ns} & \frac{1}{\sqrt{T}}\beta_{ns}'W(B_{sp} + \Psi_{\beta,sp}) \\ \frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,ns})'W\beta_{ns} & 0\end{bmatrix}\begin{bmatrix}u_{ns} \\ u_{sp}\end{bmatrix} =$$

$$= u_{ns}'\left[\beta_{ns}'W\beta_{ns} + \frac{2}{\sqrt{T}}\Psi_{\beta,ns}'W\beta_{ns}\right]u_{ns} + u_{sp}'\left[\frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,sp})'W\beta_{ns}\right]u_{ns} + u_{ns}'\left[\frac{1}{\sqrt{T}}\beta_{ns}'W(B_{sp} + \Psi_{\beta,sp})\right]u_{sp}$$

$$u'\hat{\beta}'W_T\hat{\beta}\lambda_0 - u'\hat{\beta}'W_T\bar{R} = u'\hat{\beta}'W_T\left[\hat{\beta}_{ns}\lambda_{0,ns} - \beta_{ns}\lambda_{0,ns} - \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} - \frac{1}{\sqrt{T}}\varphi_R\right] \stackrel{a}{=}$$
$$= \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \beta'_{ns} + \frac{1}{\sqrt{T}}\Psi'_{\beta,ns} \\ \frac{1}{\sqrt{T}}(B'_{sp} + \Psi'_{\beta,sp}) \end{bmatrix} W_T \left[\frac{1}{\sqrt{T}}\Psi_{\beta,ns}\lambda_{0,ns} - \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} - \frac{1}{\sqrt{T}}\varphi_R\right] \stackrel{a}{=}$$
$$= \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{T}}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} - \frac{1}{\sqrt{T}}\beta'_{ns}WB_{sp}\lambda_{0,sp} - \frac{1}{\sqrt{T}}\beta'_{ns}W\varphi_R \\ 0 \end{bmatrix}$$

Finally, this implies that the overall objective function asymptotically looks as follows:

$$\bar{L}_T(u) \stackrel{a}{=} u'_{ns}\left[\beta'_{ns}W\beta_{ns} + \frac{2}{\sqrt{T}}\Psi'_{\beta,ns}W\beta_{ns}\right]u_{ns} + u'_{sp}\left[\frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,sp})'W\beta_{ns}\right]u_{ns}$$
$$+u'_{ns}\left[\frac{1}{\sqrt{T}}\beta'_{ns}W(B_{sp} + \Psi_{\beta,sp})\right]u_{sp} + 2u'_{ns}\left[\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} - \beta'_{ns}WB_{sp}\lambda_{0,sp} - \beta'_{ns}W\varphi_R\right]$$
$$+T\eta_T\sum_{j=1}^{k}\frac{1}{\left\|\hat{\beta}^j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] \stackrel{a}{=}$$
$$= u'_{ns}\beta'_{ns}W\beta_{ns}u_{ns} + 2u'_{ns}\left[\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} - \beta'_{ns}WB_{sp}\lambda_{0,sp} - \beta'_{ns}W\varphi_R\right]$$
$$+T\eta_T\sum_{j=1}^{k}\frac{1}{\left\|\hat{\beta}^j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$

Now, for a spurious factor

$$T\eta_T\frac{1}{\left\|\hat{\beta}_j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = \sqrt{T}\eta T^{-d/2}c_2T^{d/2}|u_j| = \sqrt{T}\widetilde{c}\,|u_j|\,,$$

while for the strong ones

$$T\eta_T\frac{1}{\left\|\hat{\beta}_j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = c_2\sqrt{T}T^{-d/2}u_j sgn(\lambda_{0j}) \to 0,$$

since $d > 2$.

Hence, as $T \to \infty$, $\bar{L}_T(u) \xrightarrow{d} \bar{L}_n$ for every u, where

$$\bar{L}(u) = \begin{cases} -u'_{ns}\beta'_{ns}W\beta_{ns}u_{ns} - 2u'_{ns}W\beta_{ns}(\phi_R + B_{sp}\lambda_{0,ns} - \Psi_{\beta,ns}\lambda_{0,ns}) & \text{if } u_{sp} = 0 \\ \infty & \text{otherwise} \end{cases}$$

Note that $\bar{L}_T(u)$ is a convex function with the unique optimum given by

$$\left(\left[\beta_{ns}' W \beta_{ns}\right]^{-1} \beta_{ns}' W \left[\Psi_{\beta,ns}\lambda_{0,ns} + B_{sp}\lambda_{0,sp}\right] + \left[\beta_{ns}' W \beta_{ns}\right]^{-1} \beta_{ns}' W \psi_R,\, 0 \right)' .$$

Therefore, due to the epiconvergence results of Pollard (1994) and Knight and Fu (2000),

$$\hat{u}_{ns} \xrightarrow{d} \left[\beta_{ns}' W \beta_{ns}\right]^{-1} \beta_{ns}' W B_{sp}\lambda_{0,sp} + \left[\beta_{ns}' W \beta_{ns}\right]^{-1} \beta_{ns}' W (\psi_R + \Psi_{\beta,ns}\lambda_{0,ns}) ,$$
$$\hat{u}_{sp} \xrightarrow{d} 0 .$$

1.B.3 Proof of Proposition 1.3

Consider the bootstrap counterpart of the second stage regression.

$$\hat{\lambda}^* = \underset{\lambda\in\Theta}{\arg\min}\, (\bar{R}^* - \hat{\beta}^*\lambda)' W_T^* (\bar{R}^* - \hat{\beta}^*\lambda) + \mu_T \sum_{j=1}^{k} \frac{1}{||\hat{\beta}_j^*||^d} [|\lambda_j|]$$

Similar to Proposition 1.1, in terms of localised parameters, $\lambda = \hat{\lambda}_{pen} + \frac{u}{\sqrt{T}}$, the centred problem becomes

$$\hat{u}^* = \underset{u\in K}{\arg\min}\, (\hat{\lambda}_{pen} + \tfrac{u}{\sqrt{T}})' \hat{\beta}^{*\prime} W_T^* \hat{\beta}^* (\hat{\lambda}_{pen} + \tfrac{u}{\sqrt{T}}) - 2(\hat{\lambda}_{pen} + \tfrac{u}{\sqrt{T}})' \hat{\beta}^{*\prime} W_T^* \bar{R}^* +$$
$$+\mu_T \textstyle\sum_{j=1}^{k} \frac{1}{||\hat{\beta}_j^*||^d} [|\hat{\lambda}_{j,pen} + \frac{u}{\sqrt{T}}| - |\hat{\lambda}_{j,pen}|]$$

where K is a compact set on $\mathbb{R}^{k+1}$. Note that the problem is equivalent to the following one

$$\hat{u}^* = \underset{u\in K}{\arg\min}\;\; u'\hat{\beta}^{*\prime} W_T^* \hat{\beta}^* u + 2\sqrt{T} u' \hat{\beta}^{*\prime} W_T^* \hat{\beta}^* \hat{\lambda}_{pen} - 2\sqrt{T} u' \hat{\beta}_{*'} W_T^* \bar{R}^* +$$
$$+\mu_T \textstyle\sum_{j=1}^{k} \frac{1}{||\hat{\beta}_j^*||^d} [|\hat{\lambda}_{j,pen} + \frac{u}{\sqrt{T}}| - |\hat{\lambda}_{j,pen}|] .$$

If $\beta_{sp} = 0$

$$\begin{bmatrix} u_{ns}'' & u_{sp}' \end{bmatrix} \begin{bmatrix} \hat{\beta}_{ns}' + \frac{\Psi_{\beta,ns}}{\sqrt{T}}' \\ \hat{\beta}_{sp}' + \frac{\Psi_{\beta,sp}}{\sqrt{T}}' \end{bmatrix} W_T^* \begin{bmatrix} \hat{\beta}_{ns} + \frac{\Psi_{\beta,ns}}{\sqrt{T}} & \hat{\beta}_{sp} + \frac{\Psi_{\beta,ns}}{\sqrt{T}} \end{bmatrix} \begin{bmatrix} u_{ns} \\ u_{sp} \end{bmatrix} \overset{a}{=} u_{ns}' \beta_{ns}' W \beta_{ns} u_{ns}$$

$$2\sqrt{T} u' \hat{\beta}^{*\prime} W_T^* \hat{\beta}^* \hat{\lambda}_{pen} - 2\sqrt{T} u' \hat{\beta}^{*\prime} W_T^* \bar{R}^* = 2u' \begin{bmatrix} \hat{\beta}_{ns}' + \frac{\Psi_{\beta,ns}}{\sqrt{T}}' \\ \hat{\beta}_{sp}' + \frac{\Psi_{\beta,sp}}{\sqrt{T}}' \end{bmatrix} W_T^* \sqrt{T} \left[\hat{\beta}^* \hat{\lambda}_{pen} - \hat{\beta}^* \bar{R}^*\right]$$

Further,

$$\hat{\beta}^*\hat{\lambda}_{pen} - \hat{\beta}^*\bar{R}^* = -\frac{1}{\sqrt{T}}\psi_R + \frac{1}{\sqrt{T}}\hat{\beta}\hat{\lambda}_{pen} + \left[\bar{R} - \hat{\beta}\hat{\lambda}_{pen}\right]$$

$$\left[\beta_0 + \tfrac{1}{\sqrt{T}}\Psi_\beta\right]' W \left[\beta_0\lambda_0 + \tfrac{1}{\sqrt{T}}\psi_R\right] - \left[\beta_0 + \tfrac{1}{\sqrt{T}}\Psi_\beta\right]' W \left[\beta_0 + \tfrac{1}{\sqrt{T}}\Psi_\beta\right]\left[\lambda_0 + \tfrac{1}{\sqrt{T}}\psi_{pen}\right] = o_p\left(\tfrac{1}{\sqrt{T}}\right)$$

since $\psi_{pen} = [(\beta_{ns}'\beta_{ns}]^{-1}W\beta_{ns}'[-\psi_R + \beta_{ns}\Psi_R]$

This in turn implies that the bootstrap counterpart of the second stage satisfies

$$\hat{u}^* = \arg\min_{u\in K} u'\beta_{ns}'W\beta_{ns}u + 2u_{ns}'\beta_{ns}'W(-\psi_R + \Psi_{\beta,ns}\lambda_{0,ns}) + \eta_T\textstyle\sum_{j=1}^k \frac{1}{||\hat{\beta}_j^*||^d}[|\hat{\lambda}_{j,pen} + \frac{u}{\sqrt{T}}| - |\hat{\lambda}_{j,pen}|]$$

The weak convergence of $\sqrt{T}(\hat{\lambda}_{pen}^* - \hat{\lambda}_{pen})$ to $\sqrt{T}(\hat{\lambda}_{pen} - \lambda_0)$ now follows from the argmax theorem of Knight and Fu (2000).

1.B.4 Proof of Proposition 1.4

The condition in Proposition 1.4 requires the strict monotonicity of the cdf to the right of a particular α-quantile. This implies that if $B_T \to B$ weakly, then $B_T^{-1}(\alpha) \to B^{-1}(\alpha)$ as $T \to \infty$. Hence, $P(\lambda_0 \in I_{T,\alpha}) \to \alpha$ as $T \to \infty$.

If there is at least one non-spurious component (e.g. a common intercept for the second stage or any useful factor), the limiting distribution of the estimate will be a continuous random variable, thus implying the monotonicity of its cdf, and again, driving the desired outcome.

1.B.5 Proof of Proposition 1.5

The argument for the consistency and asymptotic normality of the Pen-GMM estimator is derived on the basis of the empirical process theory. The structure of the argument is similar to the existing literature on the shrinkage estimators for the GMM class of models, e.g. Caner (2009), Liao (2013), and Caner and Fan (2014). I first demonstrate the consistency of the estimator.

The sample moment function can be decomposed in the following way:

$$\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) = \frac{1}{T}\sum_{t=1}^{T}(g_t(\theta) - \mathbb{E}g_t(\theta)) + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}g_t(\theta)$$

Under Assumption 2, by the properties of the empirical processes (Andrews (1994))

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(g_t(\theta) - \mathbb{E}g_t(\theta)) = O_p(1)$$

Further, by Assumption 2.2

$$\mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\right) \xrightarrow{p} g_1(\theta)$$

Also note that for the strong factors $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \sim \eta T^{-d/2} O_p(1) \to 0$, while for the spurious ones $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \sim \eta T^{-d/2} c_j T^{d/2} \to \widetilde{c}_j > 0$

Therefore, the whole objective function converges uniformly in $\theta \in S$ to the following expression

$$g_1(\theta)' W(\theta) g_1(\theta) + \sum_{j=1}^{k} \widetilde{c}_j \left|\lambda_j\right| \mathbb{1}\{\beta_j = 0\}$$

Finally, since $g_1(\theta_{0,ns}, \lambda_{sp}) = g_1(\theta_{0,ns}, \mathbf{0}_{k2})$, and $\{\mu_f, vec(\beta_f), \lambda_{0,ns}, \lambda_{0,c}\}$ are identified under Assumption 2.4, $\{\theta_{0,ns}, \mathbf{0}_{k_2}\}$ is the unique minimum of the limit objective function.

Therefore

$$\hat{\theta}_{pen} \xrightarrow{p} \arg\min_{\theta \in S} \; g_1(\theta)' W(\theta) g_1(\theta) + \sum_{j=1}^{k} \widetilde{c}_j \left|\lambda_j\right| \mathbb{1}\{\beta_j = 0\}$$

and

$$\hat{\theta}_{pen,ns} = \{\hat{\mu}_f, vec(\hat{\beta}), \hat{\lambda}_{ns}, \hat{\lambda}_c\} \to \theta_{0,ns} = \{\mu_f, vec(\beta_f), \lambda_{0,ns}, \lambda_{0,c}\}$$
$$\hat{\lambda}_{sp} \xrightarrow{p} \mathbf{0}_{k2}$$

Similar to the case of the Fama-MacBeth estimator, in order to derive the asymptotic distribution of the Pen-GMM, I rewrite the original optimization problem in the centred parameters $u = \sqrt{T}(\hat{\theta}_{pen} - \theta_0)$:

$$\hat{u} = \arg\min_{u \in K} L_T(u)$$

where

$$L_T(u) = \left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T\left(\theta_0 + \frac{u}{\sqrt{T}}\right)\left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right)\right]$$
$$-\left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0\right)\right]' W_T\left(\theta_0\right)\left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0\right)\right] + \eta_T \sum_{j=1}^{k} \frac{1}{||\hat{\beta}||_1^d}\left(\left|\lambda_{j,0} + \frac{u_{\lambda,j}}{\sqrt{T}}\right| - \left|\lambda_{j,0}\right|\right)$$

and K is a compact subset in $\mathbb{R}^{nk+2k+1}$.

Using the empirical process results (Andrews (1994)), from Assumption 2.1 it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t\left(\theta_0+\frac{u}{\sqrt{T}}\right)-Eg_t\left(\theta_0+\frac{u}{\sqrt{T}}\right)\Rightarrow Z(\theta_0)\equiv N(0,\Gamma)$$

Now, since $Eg_t(\theta_0)=0$ and by Assumption 2.3,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}g_t\left(\theta_0+\frac{u}{\sqrt{T}}\right)\to G(\theta_0)u$$

uniformly in u.

Therefore,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}g_t\left(\theta_0+\frac{u}{\sqrt{T}}\right)\Rightarrow Z(\theta_0)+G(\theta_0)u$$

Note that under the presence of useless factors, $G(\theta_0)u=G_{ns}(\theta_0)u_{ns}$ for $u\in K$, where $u_{ns}=\{u_\mu,u_\beta;,u_{\lambda,ns},u_{\lambda,c}\}$, that is all the localized parameters, except for the those corresponding to the risk premia of the spurious factors.

Therefore, by Assumption 2.4 the first part of the objective function becomes

$$\begin{aligned}V_T(u)=TL_T(u)&\Rightarrow[Z(\theta_0)+G_{ns}(\theta_0)u_{ns}]'\,W(\theta_0)\,[Z(\theta_0)+G_{ns}(\theta_0)u_{ns}]-Z(\theta_0)'W(\theta_0)Z(\theta_0)\\&=u_{ns}'G_{ns}(\theta_0)'W(\theta_0)G_{ns}(\theta_0)u_{ns}+2u_{ns}'G_{ns}(\theta_0)'W(\theta_0)Z(\theta_0)\end{aligned}$$

Now, for the spurious factors: $T\eta_T\frac{1}{\|\hat{\beta}_j\|_1^d}\left[\left|\lambda_{0j}+\frac{u_{\lambda,j}}{\sqrt{T}}\right|-|\lambda_{0j}|\right]=\sqrt{T}\eta T^{-d/2}c_2T^{d/2}\,|u_{\lambda,j}|=\sqrt{T}\tilde{c}\,|u_{\lambda,j}|$,

while for the usual ones: $T\eta_T\frac{1}{\|\hat{\beta}_j\|_1^d}\left[\left|\lambda_{0j}+\frac{u_{\lambda,j}}{\sqrt{T}}\right|-|\lambda_{0j}|\right]=c_2\sqrt{T}T^{-d/2}u_{\lambda,j}sgn(\lambda_{0j})\to 0$, since $d>2$

Therefore, $\bar{V}_T(u)\xrightarrow{d}\bar{L}_n$ for every u, where

$$\bar{L}(u)=\begin{cases}u_{ns}'G_{ns}(\theta_0)'W(\theta_0)G(\theta_0)u+2u_{ns}'G(\theta_0)'W(\theta_0)Z(\theta_0) & \text{if } u_{\lambda,sp}=\mathbf{0}_{k_2}\\ \infty & \text{otherwise}\end{cases}$$

Due to the epiconvergence theorem of Knight and Fu (2000),

$$\sqrt{T}(\hat{\lambda}_{pen,sp})\xrightarrow{d}\mathbf{0}_{k_2}$$
$$\sqrt{T}(\hat{\theta}_{pen,ns}-\theta_{0,ns})\xrightarrow{d}[G_{ns}(\theta_0)'W(\theta_0)G_{ns}(\theta_0)]^{-1}G_{ns}(\theta_0)W(\theta_0)Z(\theta_0)$$

where $\theta_{0,ns}=\{\mu_f,vec(\beta_f),\lambda_{0,ns},\lambda_{0,c}\}$

Chapter 2

Term Structure of Interest Rates and Unspanned Factors

2.1 Introduction

Understanding the movements of the yield curve along with forecasting their direction and magnitude are crucial for any economy. While monetary policy is tightly related to the short term interest rates, most business and individuals heavily rely on the medium and long term yields in their financing and investment decisions. At the same time, information contained in the yield spread predicts not only future interest rates (Fama and Bliss (1987), Cochrane and Piazzesi (2005), Campbell and Shiller (1991)), but also the level of economic activity (Hamilton and Kim (2002), Ang, Piazzesi, and Wei (2006)). It is therefore natural that pricing the cross-section of bond returns, in the attempt to uncover the relation between the yield curve and present/future states of the economy, has been one of the active areas of research in both theoretical and empirical finance.

Affine term structure models (e.g. Vasicek (1977), Cox, Ingersoll, and Ross (1985), Duffie and Kan (1996b)) have proven to be a particularly popular modelling choice, due to their ability to capture the no-arbitrage links between bonds of different maturities, as well as many stylised time series features of bond returns, while remaining easily tractable. In particular, Litterman and Scheinkman (1991) demonstrated that 3 common factors ('level', 'slope' and 'curvature') can successfully capture most of the variation in bond returns. Since then, these factors have been used for a variety of applications related to the term structure of interest rates (see the overview in Piazzesi (2010)). At the same time, the recent years have seen a considerable and growing body of literature on the role of *hidden*, or *unspanned* factors, that are not identified from the cross-section of bonds due to the lack or covariance with bond excess returns, but are nevertheless useful for forecasting the yield curve. Usually, these include various macroeconomic indicators related to inflation and economic activity (Ang and Piazzesi (2002), Chernov and Mueller (2012), Joslin, Priebsch, and Singleton (2012), Favero, Niu, and Sala (2012), Ludvigson and Ng (2009)). The impact of these factors can often be accommodated within the standard estimation approach, but their treatment usually involves a pretesting procedure to distinguish between the two types of variables and consecutive risk premia recovery, e.g. as in the regression-based approach of Adrian, Crump, and Moench (2013). This could have a non-trivial impact on the risk premia estimation, fitting the yield curve and producing its forecast.

In this chapter I propose an alternative regresion-based approach to estimating the affine term structure model, the Adaptive Ridge Estimation (ARES). Compared to the alternative settings, ARES does not require an *ex ante* distinction between the spanned and unspanned factors. It combines the ability of adaptive group lasso to correctly identify the groups of non-zero coefficients (in this case, bond returns exposure to particular spanned risk factors) with the ridge-type regression to automatically identify the correct nature of the factor and treat its risk-neutral pricing impact accordingly. In fact, I show that the factor selection procedure embedded in this estimation does not affect recovering the associated risk premia, and that the estimates follow the same asymptotic distribution as if the true nature of the factors had been known *ex ante*. The key technical feature of the approach lies in the consistent model selection, produced by adaptive group lasso: as the sample sizes increases, the method correctly distinguishes between the spanned and hidden factors. Moreover, I demonstrate that the probability of incorrectly classifying a particular factor tends to 0 at an exponential rate. At the same time, not only all the non-zero coefficients are correctly retained from the cross-section of bond excess returns, but also have the correct sign with probability tending to 1 at the same exponential rate. Combined with an adaptive ridge regression, this approach allows for an efficient recovery of the associated risk-neutral pricing equations, while the embedded factor classification does not affect the asymptotic distribution of the parameters driving the yield curve. Simulations also confirm the good finite sample properties of the new estimator.

The chapter is organised as follows. First, I introduce the exponentially affine factor models, and their implications for bond excess returns and yield curve. I present the regression-based approach of Adrian, Crump, and Moench (2013) for the case when some of the factors could be unspanned. I then introduce Adaptive Ridge Estimation of the term structure and discuss its asymptotic properties and conclude by demonstrating the finite-sample properties of the estimator using simulations.

2.2 Exponentially Affine Factor Model

The exponentially affine term structure model[1] (Adrian, Crump, and Moench (2013), Ang, Piazzesi, and Wei (2006)) assumes that the price of a bond with maturity n is driven by the innovations in the set of state variables F (the *factors*) indexed by the time of their observation, $t = 1 \dots T$.

$$\begin{aligned} P_t^n &= \mathbb{E}_t \left[M_{t+1} P_{t+1}^{n-1} \right], \\ M_{t+1} &= \exp\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\Sigma^{-\frac{1}{2}} v_{t+1} \right), \\ \lambda_t &= \Sigma^{-\frac{1}{2}} \left(\lambda_0 + \lambda_1 F_t \right), \\ F_{t+1} &= \mu + \Phi F_t + v_{t+1} \end{aligned} \tag{2.1}$$

where P_t^n is the price of a bond with maturity n at time t, M_{t+1} is a stochastic discount factor, $\mathbb{E}_t$ stands for the conditional expectation at time t, λ_0 and λ_1 are prices of risk, r_t is the short-term interest rate, $v_{t+1}|F_{s\,s=0}^{\,t} \backsim N(0, \Sigma)$.

Since the log excess return of a bond maturing in n periods can be decomposed as:

$$rx_{t+1}^{n-1} := lnP_{t+1}^{(n-1)} - lnP_t^{(n)} - r_t.$$

the model in Equation (2.1) also implies that

$$1 = \mathbb{E}_t \left[\exp\left(rx_{t+1}^{(n-1)} - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\Sigma^{-\frac{1}{2}} v_{t+1} \right) \right]. \tag{2.2}$$

[1] Th exposition of the affine term structure model and its regression-based estimation closely follow those of Adrian, Crump, and Moench (2013).

Using the properties of the lognormal distribution, Equation (2.2) becomes:

$$\mathbb{E}_t\left[rx_{t+1}^{(n-1)}\right] = cov_t\left(rx_{t+1}^{(n-1)}, \lambda_t'\Sigma^{-1/2}v_{t+1}\right) - \frac{1}{2}var_t\left(rx_{t+1}^{(n-1)}\right).$$

Denoting $\beta_t^{(n-1)} = cov_t\left(rx_{t+1}^{(n-1)}, v_{t+1}'\right)\Sigma^{-1}$, we get

$$\mathbb{E}_t\left[rx_{t+1}^{(n-1)}\right] = \beta_t^{(n-1)\prime}\left(\lambda_0 + \lambda_1 X_t\right) - \frac{1}{2}var_t\left(rx_{t+1}^{(n-1)}\right) \tag{2.3}$$

Unexpected excess returns can be decomposed into the term that correlates with the factor innovations, v_{t+1}, and the one orthogonal to it:

$$rx_{t+1}^{(n-1)} - \mathbb{E}_t\left[rx_{t+1}^{(n-1)}\right] = \beta_t^{(n-1)}v_{t+1} + e_{t+1}^{n-1} \tag{2.4}$$

Assuming that e_{t+1}^{n-1} is distributed i.i.d with zero mean and variance σ^2, it follows from Equations (2.3) and (2.4) that

$$rx_{t+1}^{(n-1)} = \beta_t^{(n-1)\prime}\left(\lambda_0 + \lambda_1 F_t\right) - \frac{1}{2}\left(\beta_t^{(n-1)\prime}\Sigma\beta_t^{(n-1)} + \sigma^2\right) + \beta_t^{(n-1)\prime}v_{t+1} + e_{t+1}^{n-1} \tag{2.5}$$

Stacking the excess returns on the bonds with different maturities together, Equation (2.5) can be written as

$$rx = \beta'\left(\lambda_0 i_T' + \lambda_1 F_-\right) - \frac{1}{2}\left(B^* vec(\Sigma) + \sigma^2 i_N\right)i_T' + \beta' V + E \tag{2.6}$$

where rx is $N \times T$ matrix of excess returns, $\beta = [\beta^1\ \beta^2\ ...\beta^N]$ is a $K \times N$ matrix of factor loadings, i_T and i_N are a $T \times 1$ and $N \times 1$ vectors of ones, $F_- = [F_0\ F_1\ ...\ F_{T-1}]$ is a $K \times T$ matrix of lagged pricing factors, $B^* = [vec(\beta^{(1)}\beta^{(1)\prime})\ ...\ vec(\beta^{(N)}\beta^{(N)\prime})]'$ is an $N \times K^2$ matrix, V is a $K \times T$ matrix, and E is an $N \times T$ matrix.

One can also show that the bond prices in this model are exponentially affine functions of the factors:

$$lnP_t^{(n)} = A_n + B_n' X_t + u_t^{(n)}, \tag{2.7}$$

where $u_t^{(n)}$ is the unobservable error component, implying the following dynamics for the excess returns:

$$rx_{t+1}^{n-1} = A_{n-1} + B_{n-1}'F_{t+1} + u_{t+1}^{(n-1)} - A_n - B_n'F_t - u_t^n + A_1 + B_1'X_t + u_t^{(1)} \tag{2.8}$$

Equations (2.5) and (2.8) together imply that

$$\begin{aligned} &A_{n-1} + B_{n-1}'F_{t+1} + u_{t+1}^{(n-1)} - A_n - B_n'F_t - u_t^n + A_1 + B_1'F_t + u_t^{(1)} \\ &= \beta_t^{(n-1)\prime}(\lambda_0 + \lambda_1 F_t) - \frac{1}{2}(\beta_t^{(n-1)\prime}\Sigma\beta_t^{(n-1)} + \sigma^2) + \beta_t^{(n-1)\prime}v_{t+1} + e_{t+1}^{n-1} \end{aligned}$$

Hence, the following system of recursive linear restrictions for the parameters for the bond prices should hold:

$$\begin{gathered} A_n = A_{n-1} + B_{n-1}'(\mu - \lambda_0) + \frac{1}{2}(B_{n-1}'\Sigma B_{n-1} + \sigma^2) + A_1 \\ B_n' = B_{n-1}'(\Phi - \lambda_1) - +B_1' \\ A_0 = 0,\ B_0' = 0, \\ \beta^{(n)\prime} = B_n', \\ u_{t+1}^{(n-1)} - u_t^n + u_t^n = e_{t+1}^{n-1} \end{gathered} \tag{2.9}$$

which concludes the derivation of the model and can be used to recover the whole yield curve from the risk premia parameters of the cross-section of bonds.

2.3 Regression-Based Estimation and Unspanned factors

Adrian, Crump, and Moench (2013) suggest the following way to estimate the model.

Step 1. Decompose factors into their predictable components and innovations using vector autoregression:

$$F_{t+1} = \mu + \Phi F_t + v_{t+1}$$

Stack the residuals from these regressions, $\hat{v}_{t+1}$ into the matrix $\hat{V}$ and estimate Σ as $\hat{\Sigma} = \frac{\hat{V}\hat{V}'}{T}$.

Step 2. Consider an unrestricted version of Equation (2.6):

$$rx = ai_T' + \beta'\hat{V} + cF_- + E$$

Using $[i_T', \hat{V}', F_-']'$ as regressors, get OLS estimates of $\hat{a}, \hat{\beta}, \hat{c}$ and use the corresponding residuals to form the estimate of the variance: $\hat{\sigma}^2 = trace(\frac{\hat{E}\hat{E}'}{NT})$. Construct the matrix $\hat{B}^*$ from $\hat{\beta}$.

Step 3. According to Equation (2.6)

$$a = \beta'\lambda_0 - \frac{1}{2}\left(B^* vec(\Sigma) + \sigma^2 i_N\right)$$
$$c = \beta'\lambda_1$$

Note, that this system of equations is overidentified. Hence, one approach would be to recover risk premia parameters by OLS[1]:

$$\hat{\lambda}_1 = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\hat{c} \tag{2.10}$$
$$\hat{\lambda}_0 = (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}(\hat{a} + \frac{1}{2}(\hat{B}^* vec(\hat{\Sigma}) + \hat{\sigma}^2 i_N))$$

Adrian, Crump, and Moench (2013) provide the asymptotic distribution and properties of the estimator defined above. These parameters can also be used to forecast the yield curve using the recursive relationships in Equation (2.9).

Unspanned factors can be easily incorporated into the setting. Assume that the state variables can be *ex ante* classified into those which are spanned and those which are not, the

[1]Vectorised estimators of the risk premia parameters can be written as follows

$$vec(\hat{\lambda}_1) = \underset{x_1 \in C \subset \mathbb{R}^{nk\times 1}}{\arg\min} \left[vec(\hat{c}) - (I_k \otimes \hat{\beta}')x_1\right]' \left[vec(\hat{c}) - (I_k \otimes \hat{\beta}')x_1\right]$$
$$vec(\hat{\lambda}_0) = \underset{x_0 \in \tilde{C} \subset \mathbb{R}^{k\times 1}}{\arg\min} \left[vec\left(\hat{a} + \tfrac{1}{2}(\hat{B}^* vec(\hat{\Sigma}) + \hat{\sigma}^2 i_N)\right) - (I_k \otimes \hat{\beta}')x_0\right]' \left[vec\left(\hat{a} + \tfrac{1}{2}(\hat{B}^* vec(\hat{\Sigma}) + \hat{\sigma}^2 i_N)\right) - (I_k \otimes \hat{\beta}')x_0\right],$$

which, after manipulation, leads to the following estimators:

$$vec(\hat{\lambda}_1) = \left[I_k \otimes (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\right] vec(\hat{c})$$
$$\hat{\lambda}_0 = \left[I_k \otimes (\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}\right] \left[\hat{a} + \frac{1}{2}\left(\hat{B}^* vec(\hat{\Sigma}) + \hat{\sigma}^2 i_N\right)\right]$$

Finally, it is easy to see that these expressions are simply the vectorised versions of those in Equation (2.10).

corresponding sets will have s and u subscripts. Without loss of generality, I will assume that the list of factors consists first of the spanned state variables, followed by the unspanned:

$$\begin{pmatrix} F_t^s \\ F_t^u \end{pmatrix} = \mu + \Phi \begin{pmatrix} F_{t-1}^s \\ F_{t-1}^u \end{pmatrix} + \begin{pmatrix} v_t^s \\ v_t^u \end{pmatrix}$$

where F_t^s is of the dimension $k_1 \times 1$ and X_t^u is a $(k - k_1) \times 1$ vector.

Since excess returns enjoy non-zero exposure only to the spanned factors (by definition of the latter), the short rate does not load on the unspanned state variables, and the model can be rewritten as follows:

$$\begin{aligned} rx_{t+1}^{(n-1)} &= \beta_t^{(n-1)\prime} (\lambda_0 + \lambda_1 F_t) - \frac{1}{2} \left(\beta_t^{(n-1)\prime} \Sigma \beta_t^{(n-1)} + \sigma^2 \right) + \beta_t^{(n-1)\prime} v_{t+1} + e_{t+1}^{n-1} \\ &= \beta_t^{(n-1)\prime} (\lambda_0 + \lambda_1 F_t) - \frac{1}{2} \left(\beta_t^{(n-1)\prime} \Sigma \beta_t^{(n-1)} + \sigma^2 \right) + \beta_t^{(n-1)\prime} (F_{t+1} - \mu - \Phi F_t) + e_{t+1}^{n-1} \\ &= -\beta_{s,t}^{(n-1)\prime} (\mu_s^* + \Phi_{ss}^* F_t^s) - \frac{1}{2} \left(\beta_{t,s}^{(n-1)\prime} \Sigma_{ss} \beta_{t,s}^{(n-1)} + \sigma^2 \right) + \beta_{t,s}^{(n-1)\prime} F_{t+1}^s + e_{t+1}^{(n-1)}, \end{aligned} \tag{2.11}$$

where μ_s^* is the upper $k_1 \times 1$ subvector of $(\mu - \lambda_0)$, corresponding to the spanned factors, Φ_{ss}^* is the $k_1 \times k_1$ upper-left block of the risk-neutral transition matrix $(\Phi - \lambda_1)$.

The estimation proceeds in a way similar to the original setting, with the only difference that estimating the unrestricted version of Equation (2.11) includes only the spanned factors, leading to the OLS-type estimate of the upper-left block of the risk-neutral matrix.

Step 1. Decompose factors into their predictable components and innovations using vector autoregression:

$$F_{t+1} = \mu + \Phi F_t + v_{t+1}$$

Stack the residuals from these regressions, $\hat{v}_{t+1}$ into the matrix $\hat{V}$ and estimate Σ as $\hat{\Sigma} = \frac{\hat{V}\hat{V}'}{T}$.

Step 2. Consider an unrestricted version of Equation (2.11), and the regression of excess returns (rx) on the contemporaneous and lagges spanned factor values (F^s and F_{-1}^s correspondingly):

$$rx = a_s i_T' + c_s X_{(-1)}^s + \beta_s' X^s + E,$$

where a_s is a $n \times 1$ vector of intercepts, while β_s and c_s are $n \times k_1$ matrices of the slope coefficients. Form the matrix B^{s*} from $\hat{\beta}^s$ and use the corresponding residuals to get an estimate of the variance: $\hat{\sigma}^2 = trace(\frac{\hat{E}\hat{E}'}{NT})$.

Step 3. According to Equation (2.11)

$$a_s = -\beta_s' \mu_s^* - \frac{1}{2}\left(B^{s*} vec(\Sigma_{ss}) + \sigma^2 i_N\right) \tag{2.12}$$
$$c_s = -\beta_s' \Phi_{ss}$$

Once again, the system of equations is overidentified, and one can apply OLS-type estimation to recover risk-neutral parameters for the spanned state variables:

$$\hat{\mu}_s^* = -(\hat{\beta}_s \hat{\beta}_s')^{-1} \hat{\beta}_s \left(\hat{a}_s + \frac{1}{2}\left(\hat{B}^{s*} vec(\hat{\Sigma}_{ss}) + \hat{\sigma}^2 i_N\right)\right) \tag{2.13}$$
$$\hat{\Phi}_{ss}^* = -(\hat{\beta}_s \hat{\beta}_s')^{-1} \hat{\beta}_s \hat{c}_s$$

Since the risk-free rate does not load on the unspanned factors, $\Phi_{su}^* = \mathbf{0}_{k_1 \times (k-k_1)}$. I also adopt the convention that those parameters which are not identified (and do not matter for the dynamics of the cross-section of bond, like λ_0^u and λ_1^{uu}) are set exactly to zero. Therefore, the following relations complete the estimator design:

$$\begin{aligned}
\Phi_{su}^* &= \mathbf{0}_{k_1 \times (k-k_1)} \\
\hat{\lambda}_1^{su} &= \Phi_{su} \\
\hat{\lambda}_1^{ss} &= \Phi_{ss} - \hat{\Phi}_{ss}^* \\
\hat{\lambda}_0^{u} &= \mathbf{0}_{(k-k_1) \times 1} \\
\hat{\lambda}_1^{uu} &= \mathbf{0}_{(k-k_1) \times (k-k_1)}
\end{aligned} \tag{2.14}$$

Adrian, Crump, and Moench (2013) provide the asymptotic distribution and properties of the resulting risk premia estimates.

2.4 Adaptive Ridge Estimation (ARES)

Adaptive Ridge Estimation of the exponentially affine term structure model combines the advantages of model selection introduced by adaptive lasso, with the ease of estimation of the weighted ridge regression. The weights in the risk premia estimation are designed to automatically distinguish between spanned and unspanned factors, which is not required to be known *ex ante*. At the same time, the method preserves the ease of estimation and interpretation of the conventional regression-based approach. In fact, as Proposition 2.3 reveals, ARES risk premia estimates have exactly the same distribution as the one in Adrian, Crump, and Moench (2013), but do not require the prior knowledge of the spanning restrictions.

The estimation approach can be described as follows.

Step 1. Decompose factors into their predictable components and innovations using a vector autoregression:

$$F_{t+1} = \mu + \Phi F_t + v_{t+1}$$

Stack the residuals from these regressions, $\hat{v}_{t+1}$ into the matrix $\hat{V}$ and estimate Σ as $\hat{\Sigma} = \frac{\hat{V}\hat{V}'}{T}$.

Step 2. Consider an unrestricted version of Equation (2.11), and the regression of excess returns (rx) on the contemporaneous and lagged factor values (F and F_{-} correspondingly):

$$rx = ai_T' + cF_{-} + \beta' F + E = \theta \tilde{F} + E,$$

where a_s is a $n \times 1$ vector of intercepts, while β_s and c_s are $n \times k_1$ matrices of the slope coefficients, $\tilde{F} = (i_T', F_{-}, F)$ and $\theta = (a, c, \beta)$. First, estimate model parameters with OLS:

$$\hat{\theta}^{ols} = (\hat{a}^{ols}, \hat{\beta}^{ols}, \hat{c}^{ols}) = rx\tilde{F}' \left(\tilde{F}\tilde{F}'\right)^{-1}$$

Then run adaptive group lasso, using OLS coefficients to define the weights for the corresponding parameters, i.e. in the vectorised form the estimation looks as follows:

$$vec(\hat{\theta}_{agl}) = \underset{\theta \in Q \subset \mathbb{R}^{2nk+n}}{\arg\min} \left[vec(rx) - (\tilde{F}' \otimes I_n)vec(\theta)\right]' \left[vec(rx) - (\tilde{F}' \otimes I_n)vec(\theta)\right] \tag{2.15}$$
$$+\gamma_T \sum_{j=1}^{k} w_j ||(\beta_j, c_j)||_2,$$

where $(\beta_j, c_j) = (\beta_{j,1}, ..., \beta_{j,n}, c_{j,1}, ..., c_{j,n})'$ is the vector of parameters that correspond to bond returns sensitivities to the current and lagged values of the factor j; $||x||_{l_2} := \sqrt{\sum_{i=1}^{n} x_i^2}$ stands for the L_2 norm of the vector x, $w_j = \frac{1}{||(\hat{\beta}_j^{ols}, \hat{c}_j^{ols})||_{l_2}}$ is the adaptive group lasso weight, which is inversely proportional to the prior OLS estimate of the parameters.

The intuition for this penalty is simple: if the true parameter value is 0, then its OLS estimate will be relatively small and, hence, the penalty of adaptive lasso becomes large enough to additionally shrink it towards 0 even in finite sample. The use of the L_2-norm in the penalty allows for the automatic selection of the groups of variables.

Adaptive lasso allows us to distinguish between spanned and unspanned factors, since the true bond loadings on the unspanned factors should be zero for all the maturities. Some of the columns inside the matrix B^*_{agl} and c will be exactly zeros, reflecting the unspanned nature of the corresponding factors, while others will reflect the non-zero correlation between bond returns and the factors. This structure will be exploited in Step 3 in order to distinguish the risk premia estimation for the spanned and unspanned variables.

Step 3. For each factor j define the following indicator:

$$p_j = \mathbb{1}\left\{\sum_{i=1}^{n} \hat{\beta}^2_{ij} = 0\right\}, \quad j = 1..k$$

The "span" indicator is equal to 1 for those factors, that are estimated to have no covariance with any of the bond holding returns. At the same time, if there is a non-zero exposure with at least one of the portfolios, the indicator is equal to zero. This set of variables defines the weights used in the following ridge regression for the risk premia estimates:

$$\begin{aligned}
\hat{\Phi}^{*r} &= \underset{x_1 \in Q_1 \times Q_2 \subset \mathbb{R}^k \times \mathbb{R}^k}{\arg\min} \left[vec(\hat{c}_{agl}) - (I_k \otimes \hat{\beta}'_{agl})vec(x_1)\right]' \left[vec(\hat{c}_{agl}) - (I_k \otimes \hat{\beta}'_{agl})vec(x_1)\right] + \textstyle\sum_{i,j=1}^{k} \max(p_i, p_j) x^2_{1,ij} \\
\hat{\mu}^{*r} &= \underset{x_0 \in Q_3 \subset \mathbb{R}^k}{\arg\min} \left[vec\left(\hat{a}_{int}\right) - (I_k \otimes \hat{\beta}'_{agl})x_0\right]' \left[vec\left(\hat{a}_{int}\right) - (I_k \otimes \hat{\beta}'_{agl})x_0\right] + \textstyle\sum_{j=1}^{k} p_j x^2_{0,j},
\end{aligned} \tag{2.16}$$

$$\begin{aligned}
\hat{\lambda}^r_{1,ij} &= (1 - \max(p_i, p_j))\left(\hat{\Phi}_{i,j} - \hat{\Phi}^{*r}_{i,j}\right) \quad i, j = 1..k, \\
\hat{\lambda}^r_{0,j} &= (1 - p_i)(\hat{\mu}_j - \hat{\mu}^{*r}_j) \quad j = 1..k,
\end{aligned}$$

where $\hat{a}_{int} = \hat{a}_{agl} + \frac{1}{2}\left(\hat{B}^*_{agl} vec(\hat{\Sigma}_{agl}) + \hat{\sigma}^2_{agl} i_N\right)$ are all the standard intercept components stacked together.

2.5 Intuitive Example

Before I proceed with establishing the asymptotic properties of the adaptive ridge estimator, it may be helpful to present the intuition behind the final stage of the procedure, the weighted ridge regression.

For simplicity, consider the case of a standard linear regression, where some of the variables on the right-hand side are the vectors of zeros, for example, as in the following data generating process, where the last one in the matrix of k regressors is 'ill-behaved':

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_1^1 & X_1^2 & \cdots & X_1^{k-1} & 0 \\ X_2^1 & X_2^2 & \cdots & X_2^{k-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_n^1 & X_n^2 & \cdots & X_n^{k-1} & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Note, that in such a model only the first $(k-1)$ coefficients are identified, and simply running an OLS regression of Y on the set of X variables will result in the lack of identification, due to rank deficiency of the regressors matrix.

Consider, however, adding into the estimation the following ridge penalty:

$$\hat{\beta}_r = \underset{\beta \in B \in \mathbb{R}^k}{\arg\min} (Y - X\beta)'(Y - X\beta) + \sum_{j=1}^{k} p_j \beta_j^2,$$

where $p_j = \mathbb{1}\left\{\sum_{i=1}^{n} X_i^2 = 0\right\}$.

The ridge regression has a very convenient property of being rewritten in terms of the simple OLS regression in the span of augmented variables. In particular, it can be easily shown that $\hat{\beta}_r$ are the coefficients in the regression of Y^* on X^*, where the latter two are defined as follows:

$$Y^* = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{(n+k)\times 1} \qquad X^* = \begin{pmatrix} X_1^1 & X_1^2 & \cdots & X_1^{k-1} & 0 \\ X_2^1 & X_2^2 & \cdots & X_2^{k-1} & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ X_n^1 & X_n^2 & \cdots & X_n^{k-1} & 0 \\ p_1 & 0 & \cdots & 0 & 0 \\ 0 & p_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & p_{k-1} & 0 \\ 0 & 0 & \cdots & 0 & p_k \end{pmatrix}_{(n+k)\times k}$$

In our particular case $p_j = 0$ for $j = 1..k-1$ and $p_k = 1$. Hence, the first $k-1$ regressors are followed by the set of zeros, while the last regressor basically become a dummy variable for the $n+k$ observation. It is a well-known result from the application of Frisch-Waugh-Lowell theorem, that in such a regression of Y^* on X^*, $\hat{\beta}_k^r$ will simply be equal to the value of

Y_{n_k}, which is 0. At the same time, the other parameter estimates would be exactly the same as if we were to run a simple OLS regression of Y on the set of first $k-1$ regressors. In other words, such weighted ridge regression automatically selects a subset of non-zero regressors and estimates the model by OLS, using them as the only available non-trivial regressors. At the same time, the elements of $\hat{\beta}$, corresponding to the zero vectors, are set exactly to 0.

Intuitively, this is exactly what happens in the third stage of the risk premia estimation, if there are unspanned factors in the model, driving the corresponding vectors of β_j to 0. One would also have to be careful when recovering $\hat{\Phi}^{*r}$, keeping different from zero only those elements $\hat{\Phi}^{*r}_{i,j}$ where both i and j correspond to the spanned factors.

2.6 Asymptotic Properties

A model selection procedure does not affect the distribution of the resulting parameter estimates, if it is *consistent* and *fast enough* in the sense that the probability of making an error and incorrectly classifying the given factors between the spanned and hidden, should be *small* relative to the rate of convergence of the risk premia estimates to the true values. In particular, I am going to show that the probability of adaptive group lasso to correctly classify all the factors, tends to 1 at an exponential rate, as the sample size increases. This, in turn, allows me to derive the asymptotic distribution of the associated risk premia estimates that is not affected by the model selection procedure.

Consider again a simple linear regression model with $Y = X\beta + \epsilon$, where the goal is to correctly identify which elements on the vector β are qual to 0, and which are not. The following definition will prove useful when discussing the likelihood of adaptive lasso to correctly distinguish between the two groups of variables.

Proposition 2.1 *Consider an adaptive group lasso estimation as in Equation (2.15). If $\gamma_T = \gamma_0 T^d$ s.t. $d \in (0, 1/2)$ then vec $\left(\hat{\theta}_s^{agl}\right) \xrightarrow{p}$ vec(θ_s) and vec$(\hat{\theta}_u^{agl}) \xrightarrow{p} 0$. Further,*

$$\sqrt{T} \begin{pmatrix} vec\left(\hat{\theta}_s^{agl} - \theta_s\right) \\ vec\left(\hat{\theta}_u^{agl}\right) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_s \\ \mathbf{0} \end{pmatrix}$$

where $\hat{\theta}_s^{agl} = \left(\hat{a}^{agl}, \hat{\beta}_s^{agl}, \hat{c}_s^{agl}\right)$, $Z_s \sim N(0, \Sigma_{\theta_s})$, and Σ_{θ_s} is the asymptotic variance-covariance matrix of the parameter estimates, if only the spanned factors were included into the model *(e.g. as in Step 2 of Adrian, Crump, and Moench (2013)).*

Proof. One can observe that while the vectorised regression has NT observations, the effective rate of convergence is still $\sqrt{T}$, since N is fixed, and hence under the assumptions of the gaussian factor factor all the standard properties of the adaptive group lasso hold. In particular, Proposition 2.1 is a direct application of Theorem 3.2 of Nardi and Rinaldo (2008). ■

In order to define whether a factor selection procedure correctly classifies a set of given state variables, I introduce the notion of *sign consistency*.

Definition 1 *Factor selection, imposed by $\hat{\theta}^{adl}$, is sign consistent, $\hat{\theta}^{agl} =_s \theta$, if and only if*

$$vec(\hat{\theta}^{adl}_u) = \boldsymbol{0}$$
$$sign(\hat{\theta}^{adl}_{i,j}) = sign(\theta_{i,j}) \;\; \forall \, i, j \; s.t. \; \theta_i \neq 0,$$

where $sign(z) = 0$ if $z = 0$, 1 if $z > 0$ and -1 if $z < 0$.

Sign consistent factor selection imposes not only correct distinction between spanned and unspanned factors, setting all the corresponding parameters in the latter case to zero. It further requires that the sign of the factor exposure (e.g. $c_{i,j}$ and $\beta_{i,j}$) is equal to the true one in case there is a non-zero correlation between the excess return on bond with maturity i and a proposed factor j. It is important to correctly determine the type of the factor, since an error in either direction affects the estimation of risk premia. The following result introduces a lower bound on the probability of achieving sign-consistent factor selection.

Proposition 2.2 *If $\gamma_t = \gamma_0 T^d$, where $d \in (0, 1/2)$, then for T large enough*

$$\mathbb{P}\left\{\hat{\theta}^{agl} =_s \theta\right\} \geq 1 - o\left(e^{-T^d}\right)$$

Proof. See Appendix A.1 ■

Proposition 2.2 not only states that as the sample size increases, the probability of correctly identifying the nature of the factor tends to 1, but that it tends towards it at an exponential rate.

Sign consistency of the adaptive group lasso is a natural extension of Zhao and Yu (2006) results on the model selection of the standard lasso estimator. The fundamental difference between the required conditions consists in the fact, that the adaptive group lasso treats model parameters in subsets, and hence, one needs to take into account their joint distribution and magnitude. Further, penalty weights depend on the prior OLS parameter estimates, and hence, are essentially random numbers. This calls for a separate evaluation of their tail behaviour and the corresponding upper/lower bounds.

The main implication of Proposition 2.2 is that the probability of making an error when classifying the factor into the spanned or hidden group, decays too fast to affect the asymptotic distribution of the risk premia. Therefore, embedded factor classification does not affect recovery of the risk premia and other parameters driving the cross-section of yields in the model.

Proposition 2.3 *Consider risk premia estimates in Equation (2.16). If $\gamma_t = \gamma_0 T^d$, then*

$$\begin{cases} \hat{\Phi}^{*r}_{ss} \xrightarrow{p} \Phi^{*r}_{ss}, \quad \hat{\Phi}^{*r}_{su} \xrightarrow{p} \mathbf{0}_{\mathbf{k_1}\times(\mathbf{k}-\mathbf{k_1})}, \quad \hat{\Phi}^{*r}_{us} \xrightarrow{p} \mathbf{0}_{(\mathbf{k}-\mathbf{k_1})\times\mathbf{k_1}}, \quad \hat{\Phi}^{*r}_{uu} \xrightarrow{p} \mathbf{0}_{(\mathbf{k}-\mathbf{k_1})\times(\mathbf{k}-\mathbf{k_1})} \\ \hat{\mu}^{*r}_{s} \xrightarrow{p} \mu^{*}_{s}, \quad \hat{\mu}^{*r}_{s} \xrightarrow{p} \mathbf{0}_{(\mathbf{k}-\mathbf{k_1})\times\mathbf{1}} \\ \hat{\lambda}^{r}_{1,ss} \xrightarrow{p} \lambda_{ss}, \quad \hat{\lambda}^{r}_{0,s} \xrightarrow{p} \lambda^{r}_{0,s} \end{cases}$$

Further,

$$\sqrt{T}\begin{pmatrix} vec(\hat{\lambda}^{r}_{1,ss} - \lambda_{1,ss}) \\ \hat{\lambda}^{r}_{0,s} - \lambda_{0,s} \\ vec(\hat{\lambda}^{r}_{1,su} - \Phi_{su}) \\ \hat{\lambda}^{r}_{1,us} \\ \hat{\lambda}^{r}_{1,uu} \\ \hat{\lambda}^{r}_{0,u} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_{\lambda_{1,s}} \\ Z_{\lambda_{0,s}} \\ Z_{\lambda_{1,su}} \\ \mathbf{0}_{(\mathbf{k}-\mathbf{k_1})\times\mathbf{k_1}} \\ \mathbf{0}_{(\mathbf{k}-\mathbf{k_1})\times(\mathbf{k}-\mathbf{k_1})} \\ \mathbf{0}_{(\mathbf{k}-\mathbf{k_1})\times\mathbf{1}} \end{pmatrix}$$

where $Z_{\lambda_{1,s}} \sim N(0, \Sigma_{\lambda_{1,s}})$, $Z_{\lambda_{0,s}} \sim N(0, \Sigma_{\lambda_{0,s}})$ and $Z_{\lambda_{1,su}} \sim N(0, \Sigma_{\lambda_{1,su}})$ are the ***oracle*** *asymptotic distributions of risk premia estimates, corresponding to the* ***ex ante*** *distinction between spanned and unspanned factors.*

Proof. See Appendix A.2 ■

Proposition 2.3 shows that risk premia estimates, produced by ARES approach, have exactly the same distribution as those derived in Adrian, Crump, and Moench (2013), but without the *ex ante* knowledge of which factors should be treated as spanned and which not. In other words, the factor selection procedure, embedded in the second stage of the estimation, is accurate enough so that it does not influence risk premia recovery.

2.7 Data

I consider excess returns on Treasury bonds with the following maturities: 3 months, 6 months, 1 year, 18 months, and 2-10 years, giving a panel of 13 bond yields over the period January, 1989 - December, 2013. The data on zero coupon yields is from Gurkaynak and Wright (2007), based on fitting Nelson-Siegel-Svensson curves. I use parameter estimates, provided by the authors, to recover the yields of the corresponding maturities. I use 1 month Treasury rate obtained from Kenneth French website as a proxy for the risk-free interest rate. These yields are then transformed into monthly bond holding excess returns, e.g.:

$$rx_{t+1}^{(n-1)} = log(1 + R_{t+1}^{(n-1)}) - log(1 + R_t^1) = (n-1)y_{t+1}^{n-1} - ny_t^n - r_t^1,$$

where $n = (3, 6, 12, 18, 24, 36, 48, 60, 72, 84, 96, 98, 120)$ stands for bond maturity, expressed in months.

I also consider two macroeconomic indicators as potentially unspanned factors in the model:

- CFNAI, Federal Reserve Bank of Chicago National Activity Index, available from Chicago Fed website[1]. The index is a weighted average of 85 different indicators of economic activity and it is released towards the end of each calendar month. Historically, the index fluctuations captured real-time expansion and contraction periods in the economy, as well as those related to higher or lower inflationary pressures, e.g. Stock and Watson (1999), Fisher (2000), Brave (2009).

- PCE Core, monthly inflation in core personal consumption expenditures. Compared to the standard CPI changes, this indicator reflects long-term trends in inflation, because the corresponding basket of goods excludes several components, more subject to high volatility transitory price shocks, such as food and energy expenses.

2.8 Simulations

In order to assess the finite sample performance of the estimator, I build the following simulation design intended to capture the standard time series and cross-sectional properties of the panel of bond yields and their return, as well as the affine term structure consequences.

[1] https://www.chicagofed.org/research/data/cfnai/current-data

I consider a panel of 13 excess returns on Treasury bonds with maturities from 3 months to 10 years during the period of January, 1989 - December 2013. Using the regression-based approach of Adrian, Crump, and Moench (2013), I estimate a five-factor model that includes 3 principal components as spanned factors, CFNA index and PCE Core inflation as the unspanned factors. I use the corresponding risk premia to recover a system of recursive relationships, describing the link between yields and factors. I then simulate a panel of yields and excess returns from the multivariate normal distribution, using the sample model-implied parameter values as a data generating process. In particular, in this design risk premia (λ_1) take the following values:

$$\lambda_1 = \begin{pmatrix} -0.0171 & 0.0844 & -0.1415 & 1.4011 & 0.1229 \\ -0.0071 & -0.0195 & 0.0744 & 0.1196 & 0.0754 \\ -0.0018 & -0.0001 & -0.0773 & -0.0265 & -0.0279 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

I use the sample sizes of $T = 60, 120, 300, 600$ that stand for the 5, 10, 25 and 50 years of monthly observations accordingly. For each simulated path of the cross-section of yields and returns, I estimate model parameters (factor loadings, risk premia, etc) using 2 methods: regression-based estimation, specifying ex ante that the macroeconomic factors should be treated as unspanned, and adaptive ridge estimation that does not require ex ante distinction between the types of the variables. For each sample size I repeat the procedure 2500 times and report the summary statistics. Tuning parameters for the adaptive group lasso are obtained via 10-fold cross-validation.

Table 2.B.1 presents the finite sample bias for the risk premia estimates produced by two alternative approaches. Even for a relatively small sample size, ARES correctly identified the two simulated 'macroeconomic' factors as unspanned with a very high degree of accuracy. The finite sample sample for both estimators is close to each other, and decreases with the sample size, as expected. Table 2.B.2 summarises the mean squared error for risk premia coefficients, obtained by two estimators. As expected, both exhibit very close MSE even for a relatively small sample of $T = 60$.

Fig. 2.B.2 demonstrates the quality of model fit for a baseline specification for various maturities. For most of the bonds, model-implied and historical yields almost perfectly coincide, corresponding to the R-squared of over 98%. This is a well-known result in the

literature, since the first three principal components already summarise most of the variation in the yields. As a result, an affine model with these variables or similar ones (in terms of their span) typically generates an almost perfect fit. The only case, when the setting did not manage to fit the data exactly, were the yields corresponding to relatively short maturities (up to 2 years), which are notoriously hard to model. The use of ARES in parameter estimation leads to a nearly identical quality of fit, and hence, the graphs are omitted.

Overall, it is easy to see that the finite sample performance of ARES is extremely close to that of the estimator relying on the *ex ante* knowledge of which factor should be treated as spanned, and which - not. Both the small sample bias and the mean squared error are close that of the ideal procedure, which indicates that although the model selection stage is embedded in the parameter estimation, it does not affect the distribution of the parameters and the accuracy of risk premia recovery. This makes the new setting particularly appropriate for the cases when in addition to the standard level, slope and curvature factors one would like to incorporate the impact of macroeconomic (or other) variables, without taking a stand on how the additional factors should be treated within the affine framework.

2.9 Conclusion

This chapter demonstrated the use of lasso-based techniques for the term structure modelling and forecast. The main advantage of the new estimator is that it does not require the knowledge of *ex ante* distinction between spanned and unspanned factors, and hence, could be widely applied for the settings that include various macroeconomic indicators. Although all the results have been derived for the regression-based estimation of the gaussian affine factor model, they could be generalised to the alternative approaches, i.e. Joslin, Singleton, and Zhu (2011) or Hamilton and Wu (2012).

There are several potential extensions of the chapter. One of the ways to further improve estimation efficiency could be to allow for the recovery of within-group sparsity, in case some (but not all) bond returns have zero exposure to a certain factor. While maintaining the same structure of the affine model, such estimation and forecast approach could lead to higher efficiency and robustness due to a more parsimonious nature of the factor loadings. It would be therefore interesting to see whether asymptotic efficiency gains could lead to better out-of-sample forecasts.

Another important question is whether lasso-related techniques could help to select the macroeconomic factors that help to improve the yield curve forecast. Since shrinkage-based techniques are particularly effective in large dimensional settings, it seems plausible that one could adapt existing procedures to be able select the list of useful unspanned factors from a large set of various macroeconomic indicators. This could not only improve our ability to forecast the term structure, but also provide additional insights on the leading indicators of the economic activity and the role of expectations.

Appendix

2.A Proofs

2.A.1 Model selection by adaptive group lasso

Consider the linear model $Y_t = X_t\beta_0 + \epsilon_t$ and the standard parameter estimation by the adaptive group lasso:

$$\hat{\beta}_{agl} = \arg\min_{b\in B} (Y - Xb)'(Y - Xb) + \gamma_T \sum_{j=1}^{k} w_j^a ||\beta_j||_2,$$

where Y_T is a $T \times 1$ vector of the response, X_T is a $T \times nk$ matrix of regressors, B is a convex subset of $\mathbb{R}^{nk}$, $w_j^a = ||\hat{\beta}_j||_{l_2}^{-1}$, for $\in \mathbb{R}$, $d > 0$, $\hat{\beta}$ is a $nk \times 1$ vector of the initial OLS estimates of β_0. W.l.o.g., assume $\beta = (\beta_1, \cdots, \beta_{k_1}, \beta_{k_1+1}, \cdots, \beta_k)$ where $\beta_j \neq \mathbf{0_n}$ for $j = 1, \cdots, k_1$ and $\beta_j = \mathbf{0_n}$ for $j = k_1 + 1, \cdots, k$. Denote by X_s^T and X_s^T the first nk_1 and last $n(k - k_1)$ columns of the matrix X_T. Further, let $C^T = \frac{1}{T}X_T' X_T$ be block-partitioned as follows:

$$C^T = \begin{pmatrix} C_{ss}^T & C_{su}^T \\ C_{us}^T & C_{uu}^T \end{pmatrix}$$

where $C_{ss}^T = \frac{1}{T}X_s'X_s$, $C_{su}^T = \frac{1}{T}X_s'X_u$, $C_{us} = \frac{1}{T}X_u'X_s$, and $C_{uu}^T = \frac{1}{T}X_u'X_u$ respectively.

The following lemmas will be useful when deriving conditions for the estimates sign consistency.

Lemma 2.1 *If* $\sqrt{T}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N(0, \Omega_{\hat{\beta}})$, *then*

a) $\mathbb{P}\left\{||\hat{\beta}_j||_{l_2} \geq c_j\right\} \geq 1 - o(e^{-T})$ *for* $j = 1...k_1$,

b) $\mathbb{P}\left\{||\hat{\beta}_j||_{l_2} \geq c_0 T^{\frac{\delta-1}{2}}\right\} \leq o(e^{-T^\delta})$ *for* $j = k_1 + 1, ..., k$,

where $c_0, c_1, ..., c_{k_1} \in \mathbb{R}$ *are finite constants s.t.* $0 < c_j < \min_{i=1..n}\left\{|\beta_{i,j}|\Big|\beta_{i,j} \neq 0\right\}$, *for* $j = 1..k_1$, $c_0 > 0$ *and* $\delta \in (0, 1)$.

Proof. It is known that for $z \sim N(0,1)$ and $\forall t > 0$ $\mathbb{P}\left\{z \geq t\right\} \equiv 1 - \Phi(t) < \frac{1}{2\pi}\frac{1}{t}e^{-1/2t^2}$, i.e. the tail probability of a normally distributed random variable has as exponential upper bound.

a) $\mathbb{P}\left\{||\hat{\beta}_j||_{l_2} \geq c_j\right\} \geq \mathbb{P}\left\{\sqrt{\sum_{\substack{i=1\\ \beta_i \neq 0}}^{n} \hat{\beta}_{i,j}^2} \geq c_j\right\} \geq \mathbb{P}\left\{\max_{i=1..n, \beta_{i,j}\neq 0} |\hat{\beta}_{i,j}| \geq c_j\right\} = 1 - \sum_{\substack{i=1\\ \beta_{i,j}\neq 0}}^{n} \mathbb{P}\left\{|\hat{\beta}_{i,j}| \leq c_j\right\}$

Consider the case of $\beta_{i,j} > 0$ for some $i = 1..n$ (the case of $\beta_{i,j} < 0$ is proved in a similar way). Let

$$\sqrt{T}(\hat{\beta}_{i,j} - \beta_{i,j}) \xrightarrow{d} z_{i,j} \sim N(0, V_{i,j}),$$

where $V_{i,j} \leq M < \infty$, $M \in \mathbb{R}_+$. Hence, for large enough T:

$$\mathbb{P}\left\{|\hat{\beta}_{i,j}| \leq c_1\right\} = (1+o(1))\mathbb{P}\left\{\left|\beta_{i,j} + \tfrac{z_{i,j}}{\sqrt{T}}\right| \leq c_j\right\} = (1+o(1))\mathbb{P}\left\{2\beta_{i,j} - c_j \leq \beta_{i,j} + \tfrac{z_{i,j}}{\sqrt{T}} \leq 2\beta_{i,j} + c_j\right\} =$$
$$(1+o(1))\mathbb{P}\left\{\sqrt{T}(\beta_{i,j} - c_j) \leq z_{i,j} \leq \sqrt{T}(\beta_{i,j} + c_1)\right\} \leq (1+o(1))\mathbb{P}\left\{\frac{z_{i,j}}{\sqrt{V_{i,j}}} \geq \frac{\sqrt{T}(\beta_{i,j} - c_j)}{M}\right\}$$
$$\leq (1+o(1))\left(1 - \Phi\left(\frac{\sqrt{T}(\beta_{i,j} - c_j)}{M}\right)\right) = o(e^{-T})$$

The same result holds for all the other elements of $\hat{\beta}_j$ s.t. $\beta_{i,j} \neq 0$. Hence, the desired inequality follows.

b) First, note that

$$\mathbb{P}\left\{||\hat{\beta}_j||_{l_2} \geq c_0 T^{\frac{\delta-1}{2}}\right\} \leq \mathbb{P}\left\{\sqrt{n}\max_{i=1..n} |\hat{\beta}_{i,j}| \geq c_0 T^{\frac{\delta-1}{2}}\right\} \leq \sum_{i=1}^{n} \mathbb{P}\left\{|\hat{\beta}_{i,j}| \geq \frac{c_0}{\sqrt{n}} T^{\frac{\delta-1}{2}}\right\}$$

Consider any $\hat{\beta}_{i,j} = 0$, $j = k_1 + 1...k$, $i = 1..n$ s.t. $\sqrt{T}(\hat{\beta}_{i,j} - \beta_{i,j}) \xrightarrow{d} z_{i,j} \sim N(0, V_{i,j})$, where $V_{i,j} \leq M < \infty$, $M \in \mathbb{R}_+$. Again, due to the properties of the normal distribution,

$$\mathbb{P}\left\{|\hat{\beta}_{i,j}| \geq \tfrac{c_0}{\sqrt{n}} T^{\frac{\delta-1}{2}}\right\} = (1+o(1))\mathbb{P}\left\{\left|\tfrac{z_{i,j}}{\sqrt{T}}\right| \geq \tfrac{c_0}{\sqrt{n}} T^{\frac{\delta-1}{2}}\right\} \leq (1+o(1))\left(1 - \Phi\left(\tfrac{c_0/\sqrt{n}T^{\frac{\delta}{2}}}{M}\right)\right) = o\left(e^{-T^{\delta}}\right)$$

■

Lemma 2.2 *$\hat{\beta}_{agl}$ is the solution to the adaptive lasso problem if and only if*

$$\begin{aligned} \left.\frac{d(Y-X\beta)'(Y-X\beta)}{d\beta_{i,j}}\right|_{\beta=\hat{\beta}_{agl}} &= -\gamma_T w_j^a \frac{\hat{\beta}_{i,j}}{||\hat{\beta}_j||_{l_2}} && \textit{for } j \textit{ s.t. } \hat{\beta}_{agl,i,j} \neq 0 \\ \left.\frac{d(Y-X\beta)'(Y_T-X_T\beta)}{db_j}\right|_{\beta=\hat{\beta}_{agl}} &\leq \gamma_T w_j^a && \textit{for } j \textit{ s.t. } \hat{\beta}_{agl,i,j} = 0 \end{aligned} \tag{2.17}$$

Proof. This is a direct application of Karush-Kuhn-Tucker conditions. ■

Lemma 2.3 *The lower bound on the probability of sign consistent adaptive group lasso estimates is given by the following:*

$$\mathbb{P}\left\{\hat{\beta}_{al} =_s \beta\right\} \geq \mathbb{P}\left\{S_s \cap S_u\right\}$$

where

$$S_s = \left\{ veciny \left[\left| \left(C_{ss}^T\right)^{-1} Z_s^T \right| \right]_{i,j} < \sqrt{T}\left(|\beta_{i,j}| - \tfrac{\gamma_T}{T} vecinv \left[\left| \left(C_{ss}^T\right)^{-1} W_s \right| \right]_{i,j} \right) \right\},\ for\ j = 1...k_1,\ i = 1..n,\ s.t.\ \beta_{i,j} \neq 0,$$

$$S_u = \left\{ \left| C_{us}^T \left(C_{ss}^T\right)^{-1} Z_s^T - Z_u^T \right| < \frac{\gamma_T}{2\sqrt{T}} \left[W_u - \left| C_{us}^T \left(C_{ss}^T\right)^{-1} W_s \right| \right] \right\},$$

where $Z_s^T = \frac{1}{\sqrt{T}} X_1' \epsilon$ *and* $Z_u^T = \frac{1}{\sqrt{T}} X_2' \epsilon$. $W_s = vec(w_1, ...w_{k_1})$, $vecinv : \mathbb{R}^{nk_1 \times 1} \to \mathbb{R}^{n \times k_1}$ *is the inverse vectorisation operator, and* $W_u = vec(w_{k_1+1}, .., w_k)$ *are* $nk_1 \times 1$ *and* $n(k - k_1) \times 1$ *vectors of weights* $w_j \in \mathbb{R}^n$ *defined below.*

Proof. Condition S_1 states that the groups of the coefficients with at least some non-zero parameters in them are retained in the model. Further, the signs on the non-zero elements in such groups are correctly recovered. Conditional on S_s, S_u further implies that the groups of coefficients, where all the true values of betas are equal to zero, are correctly identified and set exactly to 0 in sample.

To see this, it is illustrative to rewrite the problem in the centred variables.

Let $\hat{u} = \hat{\beta}_{al} - \beta$. Since $Y_t = X_t\beta + \epsilon_t$, the original problem is equivalent to the following:

$$\hat{u} = \arg\min_{u \in U} \left(\epsilon - Xu\right)' \left(\epsilon - Xu\right) + \gamma_T \sum_{j=1}^{k} w_j^a ||\beta_j + u_j||_{l_2} =$$

$$\arg\min_{u \in U} -2\frac{X'\epsilon}{\sqrt{T}} \left(\sqrt{T}u\right) + \left(\sqrt{T}u\right)' \frac{X'X}{T} \left(\sqrt{T}u\right) + \gamma_T \sum_{j=1}^{k} w_j^a ||\beta_j + u_j||_{l_2} =$$

$$\arg\min_{u \in U} -2Z_T \left(\sqrt{T}u\right) + \left(\sqrt{T}u\right)' C_T \left(\sqrt{T}u\right) + \gamma_T \sum_{j=1}^{k} w_j^a ||\beta_j + u_j||_{l_2}$$

Note that if the optimal $\hat{u} = (u_s', u_u')'$, where u_s is a $k_1 \times 1$ vector, satisfies the following conditions, it will be automatically sign consistent:

$$C_{ss}\left(\sqrt{T}\hat{u}_1\right) - Z_s^T = -\frac{\gamma_T}{\sqrt{T}}W_s$$
$$\forall j = 1\dots k_1 \text{ and } i = 1\dots n \text{ s.t. } \beta_{i,j} \neq 0,\ |\hat{u}_{i,j}| < |\beta_{i,j}|$$
$$\left|C_{us}\left(\sqrt{T}\hat{u}_1\right) - Z_u^T\right| \leq \frac{\gamma_T}{\sqrt{T}}W_u$$

where $W_s = vec(w_1, \dots w_{k_1}) \in \mathbb{R}^{nk_1}$, s.t. $\forall j \in [1, k_1]$ $w_j = \frac{w_j^a \hat{\beta}_j^{agl}}{||\hat{\beta}_j^{adl}||_{l_2}}$ is a $n \times 1$ vector of weights, associated with group j, and $W_u = vec(w_{k_1+1}, \dots w_k) \in \mathbb{R}^{n(k-k_1)}$, s.t. $\forall j \in [k_1+1, k]$ $w_j = w_j^a$ are adaptive group lasso weights.

These conditions should hold element by element and imply that groupwise $\hat{\beta}_j \neq \mathbf{0_n}$ for those cases, where there is at least one non-zero component. Condition $|u_{i,j}| < |\beta_{i,j}|$ further guarantees that the signs of these non-zero components are correctly recovered. At the same time, when all the true value of betas in the group are zeros, the penalty of the adaptive group lasso becomes non-differentiable, and the corresponding parameter estimates are jointly set exactly to zero. It is easy to see that the existence of such $\hat{u}$ follows from the structure of the FOC for the solution (see Lemma 2.2), the constraints imposed in S_1, S_2 and the minimizer uniqueness. ■

Proof of Proposition 2.2. By Bonferroni inequality, it follows from Lemma 3 that

$$\mathbb{P}\left\{\hat{\beta}_{al} =_s \beta\right\} \geq 1 - \mathbb{P}\left\{S_s^c\right\} - \mathbb{P}\left\{S_u^c\right\},$$

where

$$S_s = \left\{vecinv\left[|\left(C_{ss}^T\right)^{-1} Z_s^T|\right]_{i,j} \geq \sqrt{T}\left(|\beta_{i,j}| - \tfrac{\gamma_T}{T} vecinv\left[\left|\left(C_{ss}^T\right)^{-1} W_s\right|\right]_{i,j}\right)\right\}, \text{ for } j = 1...k_1,\ i = 1..n, \text{ s.t. } \beta_{i,j} \neq 0,$$

$$S_2^c = \left\{\left|C_{us}^T\left(C_{ss}^T\right)^{-1} Z_s^T - Z_u^T\right| \geq \frac{\gamma_T}{2\sqrt{T}}\left[W_u - \left|C_{us}^T\left(C_{ss}^T\right)^{-1} W_s\right|\right]\right\}$$

describe the set of events complementary to S_s and S_u. It is left to demonstrate that $\mathbb{P}\left\{S_s^c\right\} \leq o(e^{-T^d})$ and $\mathbb{P}\left\{S_u^c\right\} \leq o(e^{-T^d})$.

Note that under the conditions of the gaussian factor model, $vec(Q_s^T) = (C_{ss}^T)^{-1}Z_s^T \xrightarrow{d} Q \sim N(0, V_q)$, $vec(Q_s^T) = vec(q_1^T, ..., q_{k_1}^T)$ and $vec(\tilde{Q}^T) = vec(\tilde{q}_{k_1+1}^T, ..., \tilde{q}_k^T) = (C_{ss}^T)^{-1}Z_s^T - Z_u^T \xrightarrow{d} \tilde{Q} \sim N(0, \tilde{V}_q)$, where V_q and $\tilde{V}_q$ are variance-covariance matrices such that $\exists\ M_2 \in \mathbb{R}_+,\ M_2 < \infty$, s.t. $V_{i,j} \leq M_2$, $j = 1 \times k_1$, $i = 1 \times n$ and $\tilde{V}_q i, j \leq M_2$, $j = k_1 + 1 \times k$, $i = 1 \times n$.

By Proposition 2.1, $\sqrt{T}\left(\hat{\beta}_s^{agl} - \beta_s\right) \xrightarrow{d} \tilde{Z}$, where $\tilde{Z} \sim N(0, V_s)$ with bounded V_s. Further, Since for $j = 1 \times k_1$ $w_j = w_j^a \hat{\beta}_j^{agl}/||\hat{\beta}_j^{adl}||_{l_2}$, where $w_j^a = ||\hat{\beta}_j^{ols}||_{l_2}^{-1}$. Hence, following the arguments in

Lemma 2.1, the following is true for T large enough,

$$\mathbb{P}\left\{||\hat{\beta}_j^{adl}||_{l_2} \geq c_j\right\} \geq 1 - o(e^{-T})$$
$$\mathbb{P}\left\{|\hat{\beta}_{i,j}^{agl}| \geq c_j\right\} \geq 1 - o(e^{-T}) \quad \forall i = 1..n \ \text{ s.t. } \ \beta_{i,j} \neq 0,$$
$$\mathbb{P}\left\{w_j^a \leq 1/c_j\right\} \geq 1 - o(e^{-T})$$
$$\mathbb{P}\left\{|w_j| \leq \frac{|\beta_j| + c_0 T^{\frac{\delta-1}{2}}}{c_j^2}\right\} \geq 1 - o(e^{-T^\delta})$$

where $0 < c_j \leq \min\limits_{i=1..n} \left\{|\beta_{i,j}| \middle| \beta_{i,j} \neq 0\right\}$, $c_0 > 0$ and $\delta \in (0,1)$.

Hence, for $b = (b_1, ..., b_{k_1}) = vecinv(|C_{ss}^T W_s|)$, $b \in \mathbb{R}^{n \times k_1}$, $\exists\ 0 < M_2 < \infty$, $\underline{c} \in (0, M_2)$ and $\bar{c} \in (0, M_2)$ s.t. $\mathbb{P}\left\{b_{i,j} \leq \underline{c} + \bar{c}T^{\frac{\delta-1}{2}}\right\} \geq 1 - o(e^{-T^\delta})$.

Therefore, since $\gamma_T = o\left(\sqrt{T}\right)$

$$\mathbb{P}\left\{S_s^c\right\} \leq \sum_{j=1}^{k_1} \sum_{\substack{i=1 \\ \beta_{i,j} \neq 0}}^{n} \mathbb{P}\left\{|q_{i,j}^T| \geq \sqrt{T}\left(|\beta_{i,j}| - \frac{\gamma_T}{T} b_{i,j}\right)\right\}$$

$$\leq \sum_{j=1}^{k_1} \sum_{\substack{i=1 \\ \beta_{i,j} \neq 0}}^{n} \left[\mathbb{P}\left\{\left|q_{i,j}^T\right| \geq \sqrt{T}\left(|\beta_{i,j}| - \tfrac{\gamma_T}{T} b_{i,j}\right) \middle| b_{i,j} \leq \underline{c} + \bar{c}T^{\frac{\delta-1}{2}}\right\} \mathbb{P}\left\{b_{i,j} \leq \underline{c} + \bar{c}T^{\frac{\delta-1}{2}}\right\} + \mathbb{P}\left\{b_{i,j} \geq \underline{c} + \bar{c}T^{\frac{\delta-1}{2}}\right\}\right]$$

$$\leq \sum_{j=1}^{k_1} \sum_{\substack{i=1 \\ \beta_{i,j} \neq 0}}^{n} \left[\mathbb{P}\left\{|\tilde{z}_j^T| \geq \sqrt{T}\left(|\beta_j| - \frac{\gamma_T}{T}\left(\underline{c} + \bar{c}T^{\frac{\delta-1}{2}}\right)\right)\right\} + o(e^{-T^\delta})\right]$$

$$\leq (1 + o(1)) \sum_{j=1}^{k_1} \sum_{\substack{i=1 \\ \beta_{i,j} \neq 0}}^{n} \left[1 - \Phi\left((1 + o(1)\frac{1}{M_2}\sqrt{T}\beta_j)\right) + o\left(e^{-T^\delta}\right)\right] = o\left(e^{-T^\delta}\right)$$

Similarly, for $\tilde{b} = (\tilde{b}_{k_1+1}, ..., \tilde{b}_k) = vecinv(|C_{us}^T C_{ss}^T W_s|)$, $b \in \mathbb{R}^{n \times (k-k_1)}$, $\exists\ 0 < M_3 < \infty$, $\underline{g} \in (0, M_3)$ and $\bar{g} \in (0, M_3)$ s.t. $\mathbb{P}\left\{b_{i,j} \leq \underline{g} + \bar{g}T^{\frac{\delta-1}{2}}\right\} \geq 1 - o(e^{-T^\delta})$. Further, by Lemma 2.1, for $j = (k_1+1)...k$, $\mathbb{P}\left\{w_j^a \geq c_0 T^{\frac{1-\delta}{2}}\right\} \geq 1 - o(e)^{-T^\delta}$ for $\delta \in (0, 2d)$. Therefore, since $\gamma_T = \gamma_0 T^d$, for T large enough

$$\mathbb{P}\left\{S_u^c\right\} \leq \sum_{j=k_1+1}^{k} \sum_{i=1}^{n} \mathbb{P}\left\{|\tilde{q}_{i,j}^T| \geq \frac{\gamma_T}{\sqrt{T}}\left(w_j^a - \tilde{b}_{i,j}\right)\right\}$$

$$\leq \sum_{j=k_1+1}^{k} \sum_{i=1}^{n} \mathbb{P}\left\{|\tilde{q}_{i,j}^T| \geq \frac{\gamma_T}{\sqrt{T}}\left(w_j^a - \tilde{b}_{i,j}\right) \middle| w_j^a \geq c_0^{-1} T^{\frac{1-\delta}{2}}, \tilde{b}_{i,j} \leq \underline{g} + \bar{g} T^{\frac{\delta-1}{2}}\right\} \times$$

$$\times \mathbb{P}\left\{w_j^a \geq c_0^{-1} T^{\frac{1-\delta}{2}}, \tilde{b}_{i,j} \leq \underline{g} + \bar{g} T^{\frac{\delta-1}{2}}\right\} + \sum_{j=k_1+1}^{k} \sum_{i=1}^{n} \left[\mathbb{P}\left\{w_j^a \leq c_0^{-1} T^{\frac{1-\delta}{2}}\right\} + \mathbb{P}\left\{\tilde{b}_{i,j} \geq \underline{g} + \bar{g} T^{\frac{\delta-1}{2}}\right\}\right]$$

$$\leq \sum_{j=k_1+1}^{k} \sum_{i=1}^{n} \mathbb{P}\left\{|\tilde{q}_{i,j}^T| \geq \frac{\gamma_T}{\sqrt{T}}\left(c_o^{-1} T^{\frac{1-\delta}{2}} - \underline{g} + \bar{g} T^{\frac{\delta-1}{2}}\right)\right\} + o\left(e^{-T^{\delta}}\right)$$

$$\leq (1+o(1)) \sum_{j=k_1+1}^{k} \sum_{i=1}^{n} \left(1 - \Phi\left((1+o(1)) \frac{1}{c_0 M_3} T^{d-\delta/2}\right)\right) + o\left(e^{-T^{\delta}}\right) = o\left(e^{-T^{\min(\delta, 2d-\delta)}}\right)$$

Finally, it is easy to see that for $\forall d \in (0, 1/2)$, $\min\limits_{\delta \in (0,2d)} (\delta, \min(2d-\delta, \delta))$ is obtained when $\delta = d$. Therefore,

$$\mathbb{P}\left\{\hat{\beta}_{agl} =_s \beta\right\} \geq 1 - o(e^{-T^d}),$$

that is as the sample size increases, the probability of getting sign-consistent estimates from the adaptive group lasso approaches 1 at an exponential rate. ■

2.A.2 Adaptive Ridge Estimation

For notational ease, I sketch the proof for a simplified model, however, all the results go through for the setting in 2.16.

Consider the case of a linear model, where some of the regressors are vectors of zeros (which corresponds to the case of an unspanned factor, having zero covariance with the portfolio returns, and hence, zero columns in place β_j and c_j for some $j = k_1, .., k$). One approach to estimate such a model (the *oracle* one), would be to identify such factors ex ante and estimate the model parameters, using only the subset of regressors:

$$\hat{\lambda}_{or} = (\hat{X}_s' \hat{X}_s)^{-1} \hat{X}_s' \hat{Y}_s$$

where $\hat{Y}_s$ and $vec(\hat{X}_s)$ are $n \times 1$ and $nk_1 \times 1$ vectors of random variables, s.t. $vec(\hat{X}_s) \xrightarrow[T]{p} vec(X_s)$, $\hat{Y}_s \xrightarrow[T]{p} Y_s$, where $\xrightarrow[T]{p}$ stands for convergence in probability when $T \to \infty$, and

$$\sqrt{T}\left(\begin{array}{c} vec(\hat{X}_s - X_s) \\ \hat{Y}_s - Y_s \end{array}\right) \xrightarrow{d} Z, \quad Z = \left(\begin{array}{c} Z_x \\ Z_y \end{array}\right) \sim N\left(\left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right], \left[\begin{array}{cc} \Sigma_x & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_y \end{array}\right]\right)$$

This is exactly the setting of Adrian, Crump, and Moench (2013), where factor exposures are $\sqrt{T}$-consistently estimated and jointly follow an asymptotically normal distribution with the corresponding variance-covariance matrix. Note, that if X_s is full rank, then $\hat{\lambda} \xrightarrow[T]{p} \lambda_{0,s} = [X'_s X_s]^{-1} X'_s Y_s$ and

$$\sqrt{T}(\hat{\lambda}_{or} - \lambda_{0,s}) \xrightarrow{d} (X'_s X_s)^{-1} \sqrt{T}\left[\hat{X}_s \hat{Y}_s - \hat{X}'_s \hat{X}_s \lambda_{0,s}\right] = Z_\lambda,$$

After some manipulation one can also show that

$$Z_\lambda = (X'_s X_s)^{-1}\left[X'_s Z_y + [vecinv(Z_x)]'(Y_s - X_s\lambda_{0,s}) - X'_s vecinv(Z_x)\lambda_{0,s}\right]$$

Further, $Z_\lambda \sim N(0, \Sigma_{\lambda_0})$, where $\Sigma_\lambda = Var(Z_\lambda)$. The variance of the resulting estimator comes from the following components (and interaction between them):

- $(X'_s X_s)^{-1} X'_s Z_y$, the usual variation in $\hat{Y}$, which plays a role similar to the disturbance term in the classical linear regression.
- $(X'_s X_s)^{-1} X'_s vecinv(Z_x)\lambda_{0,s}$, the error-in-variables problem, stemming from the fact that $\hat{X}$ is observed only with an error. The origin of this component is similar to that of Shanken (1992) correction, arising in the cross-sectional Fama-MacBeth regressions.
- $(X'_s X_s)^{-1}[vecinv(Z_x)]'(Y_s - X_s\lambda_{0,s})$, coming from the fact that the original relationship between Y_s, X_s and λ_0 might not hold exactly. If the equality is exact (e.g. as in restrictions 2.12, this term disappears.

An alternative approach is to follow an adaptive group lasso estimation, followed by a ridge regression, introduced in Equation (2.16):

$$\hat{\lambda}^r = \underset{\lambda \in B \subset \mathbb{R}^k}{\arg\min}\,(\hat{Y}_s^{agl} - \hat{X}^{adl}\lambda)'(\hat{Y}_s^{agl} - \hat{X}^{adl}\lambda) + \sum_{i=1}^{k} p_i \lambda_i^2 \tag{2.18}$$

where $\hat{Y}^{adl}$ and $\hat{X}^{adl}$ are adaptive group lasso parameter estimates s.t. s.t. $X = [X_s, X_u]$, $X_u = \mathbf{0}_{n\times(k-k_1)}$, $vec(\hat{X}_s^{adl}) \xrightarrow[T]{p} X_s$, $vec(\hat{X}_u^{adl}) \xrightarrow[T]{p} \mathbf{0}_{n\times(k-k_1)}$, $\hat{Y}^{adl} \xrightarrow[T]{p} Y_s$, $p_i = \mathbb{1}\{X_i = \mathbf{0}_n\}$ and

$$\sqrt{T}\left(\begin{array}{c} vec(\hat{X}_s^{adl} - X_s) \\ \hat{Y}_s^{adl} - Y_s \\ vec(\hat{X}_u^{adl}) \end{array}\right) \xrightarrow{d} \left(\begin{array}{c} Z \\ \mathbf{0} \end{array}\right), \quad Z = \left(\begin{array}{c} Z_x \\ Z_y \end{array}\right) \sim N\left(\left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right], \left[\begin{array}{cc} \Sigma_x & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_y \end{array}\right]\right)$$

Note that by Proposition 2.2, $p_i \underset{T}{\overset{p}{\rightarrow}} 0$ for $i = 1..k_1$ and $p_i \underset{T}{\overset{p}{\rightarrow}} 1$ for $i = (k_1+1)...k$. Therefore,

$$(\hat{Y}_s^{agl} - \hat{X}^{adl}\lambda)'(\hat{Y}_s^{agl} - \hat{X}^{adl}\lambda) + \sum_{i=1}^{k} p_i \lambda_i^2 \underset{T}{\overset{p}{\rightarrow}} (Y_s - X_s\lambda_s)'(Y_s - X_s\lambda_s) + \sum_{i=k_1+1}^{k} \lambda_i^2 \tag{2.19}$$

This is a strictly convex function of λ, therefore by convexity lemma of Pollard (1991),

$$\hat{\lambda}^r = \begin{pmatrix} \hat{\lambda}_s^r \\ \hat{\lambda}_u^r \end{pmatrix} \underset{T}{\overset{p}{\rightarrow}} \lambda_0 = \begin{pmatrix} \lambda_{0,s} \\ \mathbf{0}_{(\mathbf{k}-\mathbf{k_1})\times \mathbf{1}} \end{pmatrix}$$

The asymptotic normality of the estimator follows from noting that the problem in Equation (2.18) can be written as follows:

$$\begin{aligned}\hat{u}^r = \underset{u\in B\subset\mathbb{R}^k}{\arg\min}\, T\Bigg[&\left(\lambda_0 + \tfrac{u}{\sqrt{T}}\right)' \hat{X}^{adl\prime}\hat{X}^{adl}\left(\lambda_0 + \tfrac{u}{\sqrt{T}}\right) - 2\left(\lambda_0 + \tfrac{u}{\sqrt{T}}\right)' \hat{X}^{adl\prime}\hat{Y}^{adl} + \textstyle\sum_{i=1}^{k} p_i \left(\lambda_{0,i} + \tfrac{u_i}{\sqrt{T}}\right)^2 \\ &-\lambda_0'\hat{X}^{adl\prime}\hat{X}^{adl}\lambda_0 + 2\lambda_0'\hat{X}^{adl\prime}\hat{Y}^{adl} - \sum_{i=1}^{k} p_i\lambda_{0,i}^2\Bigg] \qquad (2.20)\\ = \underset{u\in B\subset\mathbb{R}^k}{\arg\min}\, & u'\hat{X}^{adl\prime}\hat{X}^{adl}u - 2u'\sqrt{T}\left(\hat{X}^{adl}\hat{Y}^{adl} - \hat{X}^{adl\prime}\hat{X}^{adl}\lambda_0\right) + T\sum_{i=1}^{k} p_i\left[\left(\lambda_{0,i} + \frac{u_i}{\sqrt{T}}\right)^2 - \lambda_{0,i}^2\right],\end{aligned}$$

where $u = \sqrt{T}(\lambda - \lambda_0)$. Note, that by Proposition 2.2:
a) $Tp_i\left[\left(\lambda_{0,i} + \frac{u_i}{\sqrt{T}}\right)^2 - \lambda_{0,i}^2\right] \overset{p}{\rightarrow} 0$ for $i = 1..k_1$, and
b) $Tp_i\left(\lambda_{0,i} + \frac{u_i}{\sqrt{T}}\right)^2 \underset{T}{\overset{p}{\rightarrow}} u_i^2$ for $i = (k_1+1)..k$.

Further, note that

$$\begin{pmatrix} \hat{X}_s^{adl\prime}\hat{X}_s^{adl} & \hat{X}_s^{adl\prime}\hat{X}_u^{adl} \\ \hat{X}_u^{adl\prime}\hat{X}_s^{adl} & \hat{X}_u^{adl\prime}\hat{X}_u^{adl} \end{pmatrix} \overset{p}{\rightarrow} \begin{pmatrix} X_s'X_s & \mathbf{0}_{k_1\times(k-k_1)} \\ \mathbf{0}_{(k-k_1)\times k_1} & \mathbf{0}_{(k-k_1)\times(k-k_1)} \end{pmatrix}$$

$$\begin{aligned}\sqrt{T}\left(\hat{X}^{adl}\hat{Y}^{adl} - \hat{X}^{adl\prime}\hat{X}^{adl}\lambda_0\right) &= \sqrt{T}\left(\hat{X}^{adl\prime}(\hat{Y}^{adl} - Y) + \hat{X}^{adl\prime}Y - \hat{X}^{adl\prime}\hat{X}^{adl}\lambda_0\right) \qquad (2.21)\\ = \hat{X}^{adl\prime}\sqrt{T}(\hat{Y}^{adl} - Y) &+ \sqrt{T}(\hat{X}_{adl} - X)'(Y - X\lambda_0) + \sqrt{T}X'(Y - X\lambda_0) - X'\sqrt{T}(\hat{X}^{adl} - X)\lambda_0 \\ &\overset{d}{\rightarrow} \begin{pmatrix} X_s'Z_y + [vecinvZ_x]'(Y_s - X_s\lambda_0) - X_s'vecinv(Z_x)\lambda_{0,s} \\ \mathbf{0}_{(k-k_1)\times 1} \end{pmatrix}\end{aligned}$$

Therefore,

$$\begin{pmatrix} \hat{u}_s^r \\ \hat{u}_u^r \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (X_s'X_s)^{-1}\left[X_s'Z_y + (vecinvZ_x)'(Y_s - X_s\lambda_0) - X_s'vecinv(Z_x)\lambda_{0,s}\right] \\ \mathbf{0}_{(k-k_1)\times 1} \end{pmatrix}$$

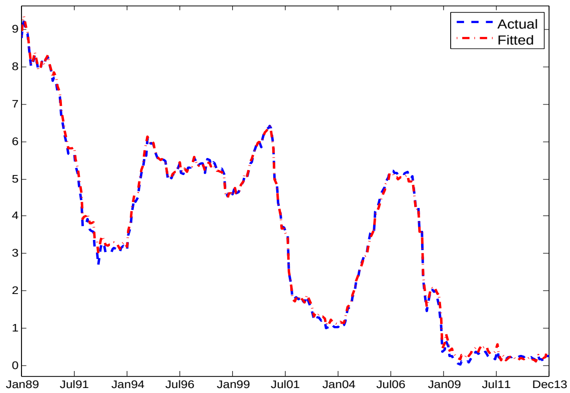
Proposition 2.3 immediately follows, if one notices that controlling for the effective sample size, the setting described above is the exact analogue of the ridge regression in Equation (2.16). The only distinction arises when recovering $\lambda_{1,ss}$, where in addition to the setting above, adaptive group lasso is also applied to the corresponding components of $\hat{Y}$, before it is vectorised:

$$\sqrt{T}\begin{pmatrix} vec(\hat{X}_s^{adl} - X_s) \\ vec(\hat{Y}_s^{adl} - Y_s) \\ vec(\hat{X}_u^{adl}) \\ vec(\hat{Y}_u^{adl}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_x \\ Z_y \\ \mathbf{0}_{n(k-k_1)\times 1} \\ \mathbf{0}_{n(k-k_1)\times 1} \end{pmatrix}, \quad Z = \begin{pmatrix} Z_x \\ Z_y \end{pmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}' & \Sigma_y \end{bmatrix}\right)$$

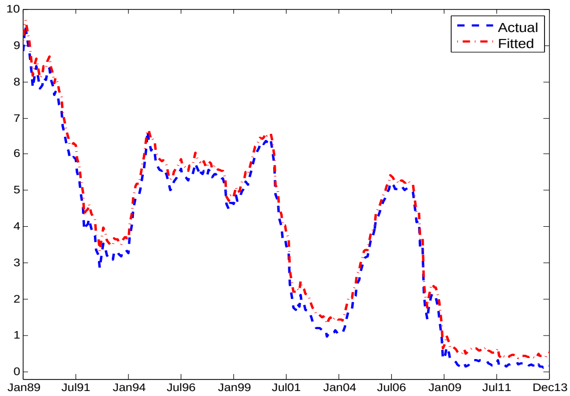
Indicator $(1-p_i)(1-p_j)$ is therefore used to identify the $k_1 \times k_1$ submatrix of risk premia, corresponding to the spanned factors, similar to the way p_i was used to eliminate the impact of unspanned variables in Equation (2.19). After vectorisation, one can derive the asymptotic distribution of $\hat{\lambda}^r$, following the same dimension reduction techniques outlined in Equations (2.19 – 2.21).

2.B Graphs and Tables

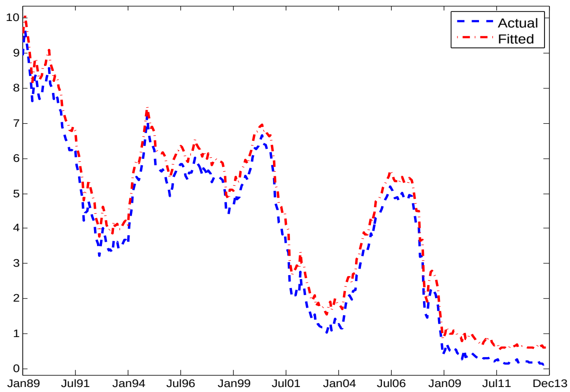
Figure 2.B.1: Typical model-implied and historical yields.



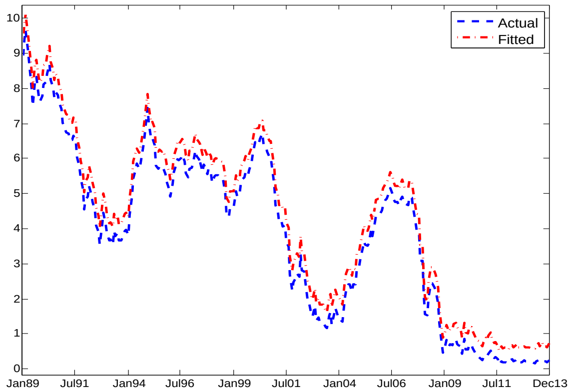
(a) 3 months



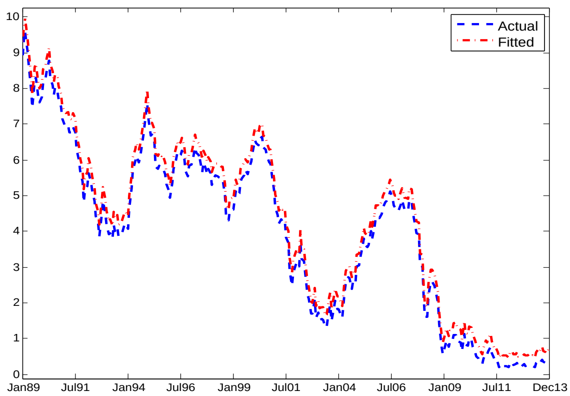
(b) 6 months



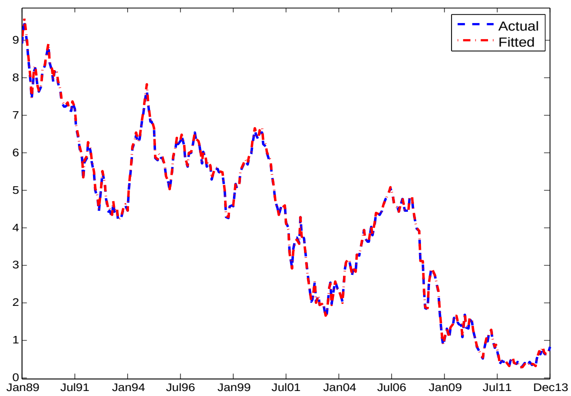
(c) 1 year



(d) 18 months

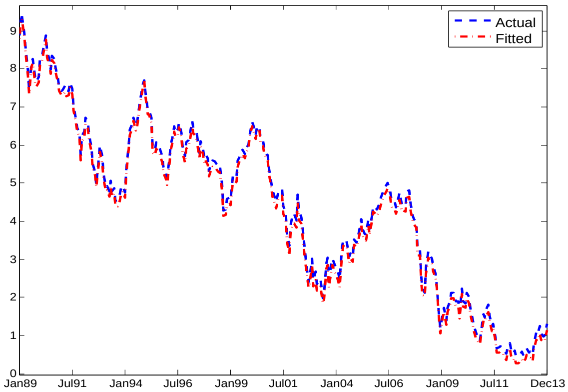


(e) 2 years

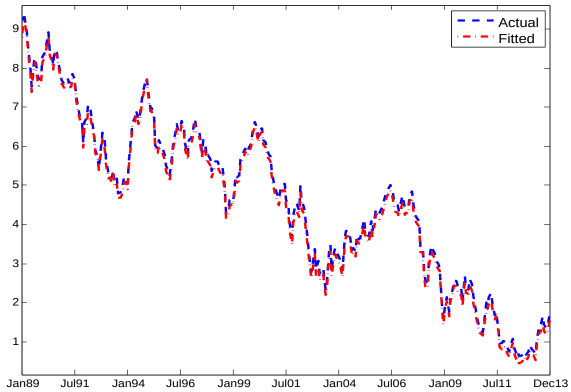


(f) 3 years

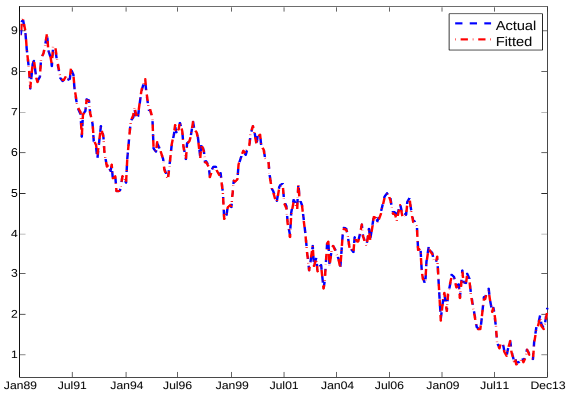
Figure 3.B.1: Typical model-implied and historical yields. (Cont.)



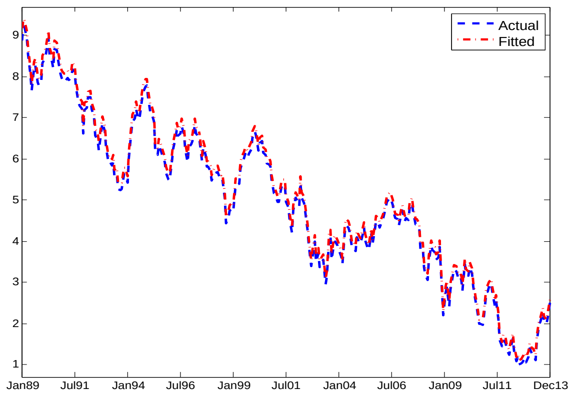
(a) 4 years



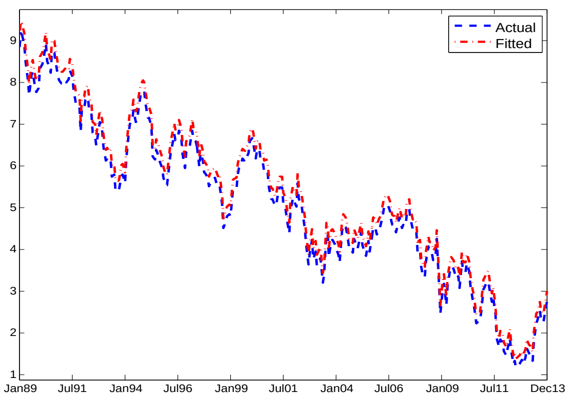
(b) 5 years



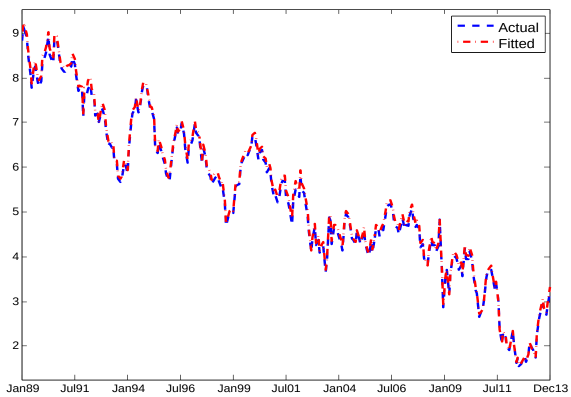
(c) 6 years



(d) 7 years



(e) 8 years



(f) 10 years

Note. The graphs present fitted and historical yields of Treasuries with various maturities, using the monthly observations for the time period of 1989:01-2013:12. The yields are fitted using 3 principal components, PCE Core inflation and CFNA index as factors; risk premia and other parameters are estimated following the regression-based approach of Adrian, Crump, and Moench (2013).

Table 2.B.1: Average bias of the risk premia estimates

Panel A: T=60

ACM (2013)					ARES				
-0.0742	0.0859	0.0823	-0.1235	-0.0069	-0.0751	0.0879	0.0678	-0.1235	-0.0069
0.017	-0.1046	0.0091	0.0031	0.0016	0.0169	-0.1044	0.0118	0.0031	0.0016
0.0085	0.0137	-0.0999	-0.0068	-0.0013	0.009	0.0157	-0.0912	-0.0068	-0.0013
0	0	0	0	0	-0.0024	0.0032	-0.013	0.0136	0
0	0	0	0	0	-0.0051	0.0076	-0.004	0.0004	0.0002

Panel B: T=120

ACM (2013)					ARES				
-0.0296	0.0395	0.0661	-0.0547	0.0004	-0.0304	0.0417	0.0533	-0.0547	0.0004
0.007	-0.0471	-0.0033	-0.0306	0.0012	0.007	-0.0472	0.0016	-0.0305	0.0012
0.004	0.007	-0.048	-0.0015	-0.001	0.0041	0.0079	-0.0406	-0.0015	-0.001
0	0	0	0	0	0	0.0001	-0.0002	-0.0004	0
0	0	0	0	0	0	0	0	0	0

Panel C: T=300

ACM (2013)					ARES				
-0.011	0.0132	0.0185	-0.039	0.0011	-0.0117	0.0154	0.0065	-0.039	0.0011
0.0026	-0.0156	-0.0052	0.0136	0.0002	0.0027	-0.0161	0.0014	0.0136	0.0002
0.0015	0.0026	-0.0162	0.008	-0.0014	0.0015	0.0029	-0.0094	0.008	-0.0014
0	0	0	0	0	0	0	0	0.0002	0
0	0	0	0	0	0	0	0	0	0

Panel D: T=600

ACM (2013)					ARES				
-0.0051	0.0073	0.0033	0.001	0.0003	-0.0057	0.0094	-0.0081	0.001	0.0003
0.0013	-0.007	-0.001	-0.008	0.0005	0.0014	-0.0077	0.0064	-0.008	0.0005
0.0008	0.001	-0.0066	-0.003	-0.0005	0.0007	0.0011	0.0001	-0.003	-0.0005
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Note. The table documents average bias in risk premia estimates produced by the regression-based approach of Adrian, Crump, and Moench (2013) and ARES, based on 2500 simulations of affine model described in Section 2.8 that includes 3 principal components and 2 unspanned factors. The data-generating process can include 60, 120, 300 or 600 monthly observations.

Table 2.B.2: Mean Squared Error (MSE) of the risk premia estimates

Panel A: T=60									
ACM (2013)					ARES				
0.0786	0.2122	0.5792	2.6107	0.1803	0.079	0.2151	0.5874	2.6107	0.1803
0.0347	0.0972	0.2673	1.1578	0.0818	0.0354	0.0984	0.2739	1.1578	0.0818
0.0137	0.0408	0.1186	0.5259	0.0364	0.0143	0.0427	0.1211	0.5259	0.0364
0	0	0	0	0	0.1211	0.1338	0.4341	0.7789	0.0017
0	0	0	0	0	0.2532	0.3821	0.2009	0.0183	0.0104

Panel B: T=120									
ACM (2013)					ARES				
0.0329	0.1054	0.3432	1.8487	0.1163	0.0331	0.1071	0.3483	1.8487	0.1163
0.0144	0.0464	0.1551	0.7993	0.0527	0.0146	0.0469	0.1589	0.7993	0.0527
0.0064	0.0211	0.0717	0.3708	0.0235	0.0066	0.0217	0.0729	0.3708	0.0235
0	0	0	0	0	0.001	0.0046	0.0119	0.076	0.0003
0	0	0	0	0	0	0	0	0	0

Panel C: T=300									
ACM (2013)					ARES				
0.0134	0.0468	0.184	1.1428	0.0679	0.0134	0.0475	0.1868	1.1428	0.0679
0.0056	0.0211	0.0821	0.505	0.0293	0.0056	0.0213	0.0838	0.505	0.0293
0.0025	0.009	0.0383	0.2264	0.014	0.0025	0.0092	0.0387	0.2264	0.014
0	0	0	0	0	0	0.0001	0.0002	0.0109	0
0	0	0	0	0	0	0	0	0	0

Panel D: T=600									
ACM (2013)					ARES				
0.0071	0.0272	0.1276	0.8191	0.0471	0.0071	0.0276	0.1294	0.8191	0.0471
0.0032	0.0124	0.0578	0.351	0.0211	0.0032	0.0125	0.0589	0.351	0.0211
0.0014	0.0055	0.0265	0.1595	0.0095	0.0014	0.0056	0.0268	0.1595	0.0095
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Note. The table documents average Mean Squared Error of the risk premia estimates produced by the regression-based approach of Adrian, Crump, and Moench (2013) and ARES. Results are based on 2500 simulations of the affine model described in Section 2.8 that includes 3 principal components and 2 unspanned factors. The data-generating process can include 60, 120, 300 or 600 monthly observations.

Chapter 3

Consumption Risk of Bonds and Stocks

3.1 Introduction

The central insight of consumption based macro-finance models is that equilibrium prices of financial assets should be determined by their equilibrium risk to households' marginal utilities and, in particular, current and future marginal utilities of consumption: agents are expected to demand a premium for holding assets that are more likely to yield low returns when the marginal utility of consumption is high i.e. when consumption (current and expected) is low. Nevertheless, in the data the contemporaneous covariance of asset returns and consumption growth is small and not disperse cross-sectionally, making it challenging to rationalised both average risk premia (e.g., Mehra and Prescott (1985), Weil (1989)) and their wide cross-sectional dispersion (e.g., Hansen and Singleton (1983), Mankiw and Shapiro (1986), Breeden, Gibbons, and Litzenberger (1989), Campbell (1996)).[1]

In this chapter, we document that consumption growth reacts slowly, but significantly, to bond and stock returns common innovations. These slow consumption adjustment shocks account for about a quarter of the time series variation of aggregate consumption growth, and its innovations explain most of the time series variation of stock returns (on average about 79%), and a significant, but small, share of the time series variation of bond returns,

[1]Recently, Julliard and Ghosh (2012) show that pricing kernels based on consumption growth alone cannot explain either the equity premium puzzle, or the cross-section of asset returns, even after taking into account the possibility of rare disasters.

and generate substantial predictability for future consumption growth.

Since consumption responds with a lag to changes in wealth, the contemporaneous covariance of consumption and wealth understates and mismeasures the true risk of an asset, rendering empirically measured risk premia hard to rationalise. On the contrary, slow consumption adjustment (SCA) risk, measured by the cumulated response of consumption growth to asset return innovations, can *jointly* explain the average term structure of interest rates and the cross-section of a broad set of stock returns (including industry portfolios and Fama-French size and book to market portfolio).

To assess the role of SCA risk in a robust manner, and using post-war data on a large cross section of both stock and US treasury returns, we perform our empirical analysis following two very different approaches and identification strategies.

First, we consider a flexible parametric setting in which consumption growth is modelled as being the sum of two independent processes: a (potentially, since parameters are estimated) long memory moving average component that (potentially) co-moves with asset returns and a transitory component orthogonal to financial assets. Innovations to asset return are in turn modelled as depending (potentially) on the long memory component of consumption plus an orthogonal component.

Empirically, we find that: *a*) consumption reacts very slowly (i.e. over a period of two to four years), but significantly, to asset returns innovations, and these innovations account for about 27% of the time series variation of consumption growth; *b*) returns on portfolios of stocks load significantly on the SCA component, with a pattern that closely mimics the value and size pricing anomalies, and this component tends to explain between 36% and 95% of their time series variation; *c*) returns on US treasury bonds load significantly on the SCA component, with loadings increasing with the time to maturity, but this component explains no more than 3.5% of their time series variations (an additional latent variable, independent from both consumption and stock returns, seems to drive most of the time series variation of bonds); *e*) SCA risk, measured as the loading of asset returns on the SCA component, can explain between 57% and 90% of the joint cross-section of stocks and bond returns.[1]

Second, not to take an ex-ante stand on a parametric model of consumption dynamics, we consider a broad class of consumption-based equilibrium models (see, e.g., Ghosh, Julliard,

[1] In our baseline specification we consider a cross section of 46 asset given by 12 industry portfolios, 25 size and book-to-market portfolios, and 9 bond portfolio, but the results appear robust to alternative specifications.

and Taylor (2013)) in which the stochastic discount factor can be factorized into a component that depends on consumption growth and an additional, model specific, component. In this setting, following Parker and Julliard (2005), we show that a pricing kernel can be constructed by measuring asset risk via the covariance between an asset return and the change in marginal utility over several quarters following the return. Using this measure, we demonstrate that the SCA risk is priced in the cross-section of bond holding returns, as well as the joint cross-section of stocks and bonds. Moreover, we show that the slow consumption adjustment risk creates a 'fanning out' pattern in consumption betas, leading to both more *pronounced* and *dispersed* covariance with the stochastic discount factor. As a result, the model captures 85% of the cross-sectional variation in bonds returns, and 37-94% of the joint cross-sectional variation in stocks and bonds.

Interestingly, our findings are consistent (both qualitatively and quantitatively) with the consumption dynamics postulated by the Long Run Risk (LRR) literature (see e.g. Bansal and Yaron (2004), Hansen, Heaton, Lee, and Roussanov (2007), Bansal, Kiku, and Yaron (2012)), but are also supportive of a broader class of consumption based asset pricing models.

Our analysis builds upon the finding of Parker and Julliard (2005) that consumption risk measured by the covariance of an assets return and consumption growth cumulated over many quarters following the return – that is, measured as slow consumption adjustment risk – can explain a large fraction of the variation in average returns across the 25 Fama-French portfolios and, more broadly, on the empirical evidence linking slow movements in consumption and asset returns (see, e.g., Daniel and Marshall (1997), Bansal, Dittmar, and Lundblad (2005), Jagannathan and Wang (2007), Hansen, Heaton, and Li (2008), Malloy, Moskowitz, and Vissing-Jorgensen (2009)). We expand upon this framework by both *i*) identifying the SCA risk component from, and quantifying its relevance for, the time series properties of consumption and asset returns, and *ii*) by showing that this component can price jointly different classes of assets and tends to act as a driving factor of the term structure of interest rates. We also show that an additional, non-spanned (i.e. that does not seem to require a risk premium), factor is also required to rationalise the time series behaviour of bonds, and that this factor tends to behave as a slope type component.[1]

More broadly, our work is connected to the large literature on the co-pricing of stocks

[1]This last finding is consistent with Chernov and Mueller (2012) that identify an unspanned latent factor driving in bond yields.

and bonds.[1] In particular, our focus on the role of macroeconomic risk is related to a series of works that combine the affine asset pricing framework with a parsimonious mix of macro variables and bond factors for the joint pricing of bonds and stocks. In particular: Bekaert and Grenadier (1999) and Bekaert, Engstrom, and Grenadier (2010), that presents a linear model for the simultaneous pricing of stock and bond returns that jointly accommodate the mean and volatility of equity and long term bond risk premia; Brennan, Wang, and Xia (2004), that assumes that the investment opportunity set is completely described by two state variables given by the real interest rate and the maximum Sharpe ratio, and the state variables (estimated using US Treasury bond yields and inflation data) are shown to be related to the equity premium, the dividend yield, and the Fama-French size and book-to-market portfolios; Lettau and Wachter (2011), that focus on matching an upward sloping bond yield term structure and a downward sloping equity yield curve via an affine model that incorporates persistent shocks to the aggregate dividend, inflation, risk-free rate, and price of risk processes; Koijen, Lustig, and Nieuwerburgh (2010), that develops an affine model in which three factors –the level of interest rates, the Cochrane and Piazzesi (2005) factor,[2] and the dividend-price ratio– have explanatory power for the cross-section of bonds and stock returns, while the latter two factors have explanatory power for the time series of these assets; Ang and Ulrich (2012), that considers an affine model in which returns to bonds (real and nominal) and stocks, are decomposed into five components meant to capture the real short rate dynamics as well as term premium, inflation related components (a nominal premium, an expected inflation as well as an inflation risk component) as well as a real cash flow risk element.

The reminder of the chapter is organized as follows. Section 3.2 formally defines the concept of slow consumption adjustment risk in a broad class of consumption based asset pricing models. Sections 3.3 presents the econometric methodology, while a description of the data is reported in Section 3.4. Our empirical findings are reported in Section 3.5 while Section 3.6 concludes. Additional methodological details, as well as robustness checks and

[1] E.g.: Fama and French (1993) expands the original set of Fama and French (1992) stock market factors (meant to capture the overall market return, as well as the value and the size premia), with two bond factors (the excess return on a long bond and a default spread), meant to capture term and default premia; Mamaysky (2002) built upon the affine term structure framework canonically used in term structure modelling (see, e.g., Duffie and Kan (1996a)) by adding affine dividend yields to help pricing jointly bonds and stocks.

[2] Cochrane and Piazzesi (2005) find that a single factor (a single tent-shaped linear combination of forward rates), predicts excess returns on one- to five-year maturity bonds. This factor tends to be high in recessions, but forecasts future expansion, i.e. this factor seems to incorporate good news about future consumption.

additional empirical evidence, are reported in the Appendix.

3.2 The Slow Consumption Adjustment Risk of Asset Returns

Representative agent based consumption asset pricing models with either CRRA, Epstein and Zin (1989), or habit based preferences, as well as several models of complementary in the utility function, and models with either departures from rational expectations, or robust control, or ambiguity aversion, and even some models with solvency constraints,[1] all imply a consumption Euler equation of the form

$$C_t^{-\phi} = \mathbb{E}_t \left[C_{t+1}^{-\phi} \tilde{\psi}_{t+1} R_{j,t+1} \right] \tag{3.1}$$

for any gross asset return j including the risk free rate R^f_{t+1}, and where $\mathbb{E}_t$ is the rational expectation operator conditional on information up to time t, C_t denotes flow consumption, $\tilde{\psi}_{t+1}$ depends on the particular form of preferences (and expectation formation mechanism) considered, and the ϕ parameter is a function of the underlying preference parameters.[2] Rearranging terms, moving to unconditional expectations, and using the definition of co-variance, we can rewrite the above equation as a model of expected returns

$$\mathbb{E}\left[R^e_{t+1}\right] = -\frac{Cov\left(M_{t+1}; R^e_{t+1}\right)}{\mathbb{E}\left[M_{t+1}\right]}. \tag{3.2}$$

where $M_{t+1} := \left(C_{t+1}/C_t\right)^{-\phi} \tilde{\psi}_{t+1}$ represents the stochastic discount factor between time t and $t+1$ and $R^e \in \mathbb{R}^N$ denotes a vector of excess returns. Log-linearizing the above relationship, expected returns can be expressed as

$$\mathbb{E}\left[R^e_{t+1}\right] = \left[\phi Cov\left(\Delta c_{t,t+1}; r^e_{t+1}\right) - Cov\left(\log \tilde{\psi}_{t+1}; r^e_{t+1}\right)\right] \lambda \tag{3.3}$$

[1] See, e.g.: Bansal and Yaron (2004); Abel (1990), Campbell and Cochrane (1999), Constantinides (1990), Menzly, Santos, and Veronesi (2004); Piazzesi, Schneider, and Tuzel (2007), Yogo (2006); Basak and Yan (2010), Hansen and Sargent (2010); Chetty and Szeidl (2015); Ulrich (2010); Lustig and Nieuwerburgh (2005).

[2] E.g., ϕ would denote relative risk aversion in the CRRA framework, while it would be a function of both risk aversion and elasticity of intertemporal substitution with Epstein and Zin (1989) recursive utility.

where $\Delta c_{t,t+1} := \ln\left(C_{t+1}/C_t\right)$, $r^e \in \mathbb{R}^N$ denotes log excess returns, and λ is a positive scalar. Since, in the data, the covariance between one period consumption growth and asset returns is small and has a much smaller cross-sectional dispersion than average excess returns, the first term of the above equation is not sufficient for pricing a cross-section of asset returns, and most of the modelling effort in the literature has been devoted to identifying a $\tilde{\psi}$ component that can help rationalise observed returns.

Note that Equation (3.1) above implies that

$$C_t^{-\phi} = \mathbb{E}_t\left[C_{t+1+S}^{-\phi}\psi_{t+1+S}\right]$$

where $\psi_{t+1+S} := R^f_{t,t+1+S}\prod_{j=0}^{S}\tilde{\psi}_{t+1+j}$. Hence, the Euler equation

$$\mathbf{0}_N = \mathbb{E}\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\phi}\tilde{\psi}_{t+1}R^e_{t+1}\right] \tag{3.4}$$

where $\mathbf{0}_N$ denotes and N-dimensional vector of zeros, can be equivalently rewritten as

$$\mathbf{0}_N = \mathbb{E}\left[\left(\frac{C_{t+1+S}}{C_t}\right)^{-\phi}\psi_{t+1+S}R^e_{t+1}\right]. \tag{3.5}$$

Once again, using the definition of covariance, we can rewrite the above equation as a model of expected returns

$$\mathbb{E}\left[R^e_{t+1}\right] = -\frac{Cov\left(M^S_{t+1}; R^e_{t+1}\right)}{\mathbb{E}\left[M^S_{t+1}\right]}. \tag{3.6}$$

where $M^S_{t+1} := \left(C_{t+1+S}/C_t\right)^{-\phi}\psi_{t+1+S}$. That is, under the null of the model being correctly specified, there is an entire family of SDFs that can be equivalently used for asset pricing: M^S_{t+1} for every $S \geq 0$. Log-linearizing the above expression, we have the linear factor model

$$\mathbb{E}\left[R^e_{t+1}\right] = \left[\phi Cov\left(\Delta c_{t,t+1+S}; r^e_{t+1}\right) - Cov\left(\log\psi_{t+1+S}; r^e_{t+1}\right)\right]\lambda_S \tag{3.7}$$

where $\Delta c_{t,t+1+S} := \ln\left(C_{t+1+S}/C_t\right)$ and λ_S is a positive scalar.

But why measure risk, and price returns, using the slow consumption adjustment framework as in equations (3.5)-(3.7) instead of the contemporaneous risk as in equations (3.2)-(3.4)? First, it is a well-known fact that consumption displays excessive smoothness in response to the wealth shocks (Flavin (1981), Hall and Mishkin (1982)), which can be caused by various adjustment costs (Gabaix and Laibson (2001)) and asynchronous consumption/investment decisions (Lynch (1996)). Moreover, the problem could be further exacerbated if the agent has a nonseparable utility function, potentially including labour or other state variables that are also costly to adjust, and hence leading to further staggering in the consumption adjustment in response to wealth innovations. Second, if there is measurement error in consumption, then using a one-period growth rate does not reflect the true pricing impact of the SDF. Indeed, in a recent paper Kroencke (2013) demonstrates that one of the reasons for the failure of the standard consumption-based model to solve equity premium and risk-free rate puzzles, is that NIPA consumption data is filtered to eliminate the impact of the measurement error. The unfiltered data, in turn, produces substantially better results. The fourth quarter to fourth quarter consumption growth of Jagannathan and Wang (2007), as well as the ultimate consumption risk of Parker and Julliard (2005), are related to the reconstructed unfiltered time series of consumption growth, and therefore provide a better measure for the overall consumption risk.

To model parametrically the–potential–slow reaction of consumption to financial market shocks, we postulate that the consumption growth process can be decomposed in two terms: a white noise disturbance, w_c with variance σ_c^2, that is independent from financial market shocks, plus a (covariance stationary) autocorrelated process–the slow consumption adjustment component–that depends on current and past stocks to asset returns. In order not to have to take an ex ante stand on the particular time series structure of the slow adjustment component, we work with its (potentially infinite) moving average representation. That is we model the consumption growth process as:

$$\Delta c_{t-1,t} = \mu_c + \sum_{j=0}^{\bar{S}} \rho_j f_{t-j} + w_t^c; \tag{3.8}$$

where $\bar{S}$ is a positive integer (potentially equal to $+\infty$), the ρ_j coefficients are square summable, and most importantly f_t, a white noise process normalised to have unit variance, is the fundamental innovation upon which all asset returns loads contemporaneously i.e. given a vector of log excess returns, r^e, we have

$$\underset{N\times 1}{r_t^e} = \underset{N\times 1}{\mu_r} + \underset{N\times 1}{\rho^r} f_t + \underset{N\times 1}{w_t^r} \tag{3.9}$$

where μ_τ is a vector of expected values, ρ^r contains the asset specific loadings on the common risk factor, w_t^r is a vector of white noise shocks with diagonal covariance matrix Σ_r (the diagonality assumption can be relaxed as explained below and in Appendix 3.A), that are meant to capture asset specific idiosyncratic shocks.

The dynamic system in equations (3.8)-(3.9) can be reformulated as a state-space model, and Bayesian posterior inference can be conducted to estimate both the unknown parameters (μ_c, μ_r, $\{\rho_j\}_{j=0}^{\bar{S}}$, ρ^r, σ_c^2, Σ_r) and the time series of the unobservable common factor of consumption and asset returns ($\{f_t\}_{t=1}^T$). This estimation procedure is described in detail in the next section and Appendix 3.A.

Note that, since $\Delta c_{t-1,t+S} \equiv \sum_{j=0}^{S} \Delta c_{t-1+j,t+j} \equiv \ln\left(C_{t+S}/C_{t-1}\right)$, from the one period consumption growth process in equation (3.8) we can recover the dynamic of cumulated consumption growth with a simple rotation since

$$[\Delta c_{t-1,t}, \Delta c_{t-1,t+1}, ..., \Delta c_{t-1,t+S}]' \equiv \Gamma\, [\Delta c_{t-1,t}, \Delta c_{t,t+1}, ..., \Delta c_{t-1+S,t+S}]'$$

where Γ is a lower triangular square matrix of ones (of dimension S). From this last expression it is easy to see that the ρ_j coefficients identify the impulse response function of slow consumption adjustment to the fundamental asset market shock f_t as

$$\frac{\partial \mathbb{E}\left[\Delta c_{t-1,t+S}\right]}{\partial f_t} = \sum_{j=0}^{S} \rho_j \tag{3.10}$$

where $\rho_{j>\bar{S}} := 0$. Moreover, the consumption betas of the factor model of asset returns in equation (3.7) are fully characterised by the loadings of the dynamic system on the factor f_t since

$$Cov\left(\Delta c_{t-1,t+S}; r_t^e\right) \equiv \sum_{j=0}^{S} \rho_j \rho^r. \tag{3.11}$$

That is, the time series estimates of the latent factor loadings ($\hat{\rho}_j$ and $\hat{\rho}^r$) can be used to assess whether the slow consumption adjustment component has explanatory power for the cross-section of risk premia (via, for instance, simple cross-sectional regressions of returns on these estimated covariances).

The formulation in Equations (3.8)-(3.9) can be generalize to allow for a bonds specific latent factor (g_t) to which consumption, potentially, reacts slowly over time. This is an appealing extension since the factor f_t, as shown in the empirical section, explains most of the time series variability of stocks, a quarter of the one of consumption growth, but a small share of the time series variation of bonds. The dynamic system in this case becomes:

$$\Delta c_{t-1,t} = \mu_c + \sum_{j=0}^{\bar{S}} \rho_j f_{t-j} + \sum_{j=0}^{\bar{S}} \theta_j g_{t-j} + w_t^c; \tag{3.12}$$

$$\underset{N\times 1}{r_t^e} = \underset{N\times 1}{\mu_r} + \underset{N\times 1}{\rho^r} f_t + \left[\underset{N_b\times 1}{\theta'^b}, \quad \mathbf{0}'_{N-N_b} \right]' g_t + \underset{N\times 1}{w_t^r}; \tag{3.13}$$

where N_b is the number of bonds and they are ordered first in the vector r_t^e, $\theta^b \in \mathbb{R}^{N_b}$ contains the bond loadings on the factor g_t–a white noise process with variance normalized to one. Note that in this case the implied covariance of consumption and returns becomes:

$$Cov\left(\Delta c_{t-1,t+S}; r_t^e\right) \equiv \sum_{j=0}^{S} \rho_j \rho^r + \left[\theta'^b, \quad \mathbf{0}'_{N-N_b} \right]' \sum_{j=0}^{S} \theta_j. \tag{3.14}$$

3.3 Econometric Methodology

Our empirical analysis is based on both parametric and nonparametric inference, ensuring the results are robust to the methodology employed. The main approach (Section 3.3.1) consists in rewriting the model in Equations (3.8)-(3.9) in state-space form and employ standard Bayesian filtering techniques to recover the unobservable latent consumption factor (f_t) and other model parameters. Since the model is tightly parametrised, with the factor loadings driving not only the time series, but also the cross-sectional relationships between asset returns, this in turn allows us to assess model performance in both time series and cross-sectional dimensions, using variance decomposition and Fama-MacBeth (1973) cross-sectional regressions.

In addition, we also use the standard semi-parametric techniques (e.g. GMM and Empirical Likelihood estimation) to assess whether ultimate consumption risk of Parker and Julliard (2005) can successfully capture the cross-section of stock and bond returns. Section 3.3.2 provides further details on the moment construction, parameter estimation and tests used for inference.

3.3.1 Parametric Inference

We can rewrite the dynamic model in Equations (3.8)-(3.9) in state-space form, assuming Gaussian innovations, as

$$z_t = F z_{t-1} + v_t, \;\; v_t \sim \mathcal{N}\left(\mathbf{0}_{\bar{S}+1}; \Psi\right); \tag{3.15}$$

$$y_t = \mu + H z_t + w_t, \;\; w_t \sim \mathcal{N}\left(\mathbf{0}_{N+1}; \Sigma\right). \tag{3.16}$$

where $y_t := [\Delta c_t, r_t^{e\prime}]$, $z_t := [f_t, ..., f_{t-\bar{S}}]'$, $\mu := [\mu_c, \mu_r']'$, $v_t := \left[f_t, \mathbf{0}_{\bar{S}}'\right]'$, $w_t := [w_t^c, w_t'^r]'$,

$$\Psi := \underbrace{\begin{bmatrix} 1 & \mathbf{0}_{\bar{S}}' \\ \mathbf{0}_{\bar{S}} & \mathbf{0}_{\bar{S}\times\bar{S}} \end{bmatrix}}_{(\bar{S}+1)\times(\bar{S}+1)}, \quad F := \underbrace{\begin{bmatrix} \mathbf{0}_{\bar{S}}' & 0 \\ I_{\bar{S}} & \mathbf{0}_{\bar{S}} \end{bmatrix}}_{(\bar{S}+1)\times(\bar{S}+1)}, \tag{3.17}$$

$$\Sigma := \underbrace{\begin{bmatrix} \sigma_c^2 & \mathbf{0}_N' \\ \mathbf{0}_N & \Sigma_r \end{bmatrix}}_{(N+1)\times(N+1)}, \quad H := \underbrace{\begin{bmatrix} \rho_0 & \rho_1 & \dots & \rho_{\bar{S}} \\ \rho^r & \mathbf{0}_N & \dots & \mathbf{0}_N \end{bmatrix}}_{(N+1)\times(\bar{S}+1)}. \tag{3.18}$$

and $I_{\bar{S}}$ and $\mathbf{0}_{\bar{S}\times\bar{S}}$ denote, respectively, an identity matrix and a matrix of zeros of dimension $\bar{S}$.

Similarly, the dynamic system in Equations (3.12)-(3.13) can be represented in the state-space form (3.15)-(3.16) with: $z_t := [f_t, ..., f_{t-\bar{S}}, g_t, ..., g_{t-\bar{S}}]'$; $v_t := \left[f_t, \mathbf{0}_{\bar{S}}', g_t, \mathbf{0}_{\bar{S}}',\right]' \sim \mathcal{N}\left(\mathbf{0}_{\bar{S}+1}; \Psi\right)$; Ψ and F block diagonal with blocks repeated twice and given, respectively, by the two matrices in equation (3.17); and with space equation coefficients given by

$$H := \underbrace{\begin{bmatrix} \rho_0 & \dots & \dots & \rho_{\bar{S}} & \theta_0 & \dots & \dots & \theta_{\bar{S}} \\ \rho_1^r & 0 & \dots & 0 & \theta_1^b & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{N_b}^r & 0 & \dots & 0 & \theta_{N_b}^b & 0 & \dots & 0 \\ \dots & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \rho_N^r & 0 & \dots & 0 & \dots & 0 & \dots & 0 \end{bmatrix}}_{(N+1)\times 2(\bar{S}+1)}. \tag{3.19}$$

The above state-space system implies the following conditional likelihood for the data:

$$y_t | \mathcal{I}_{t-1}, \mu, H, \Psi, \Sigma, z_t \sim \mathcal{N}(\mu + H z_t; \Sigma) \tag{3.20}$$

where $\mathcal{I}_{t-1}$ denotes the history of the state and space variables until time $t-1$. Hence, under a diffuse (Jeffreys') prior and conditional on the history of z_t and y_t, and given the diagonal structure of Σ, we have the standard Normal-inverse-Gamma posterior distribution for the parameters of the model (see e.g. Bauwens, Lubrano, and Richard (1999)). Moreover, the posterior distribution of the unobservable factors z_t conditional on the data and the parameters, can be constructed using a standard Kalman filter and smoother approach (see, e.g., Primiceri (2005)).

Using equation (3.7), the above specification for the dynamics of consumption and asset returns implies, in the presence of only one latent factor (f_t) common to both assets and consumption

$$\mathbb{E}[R_t^e] = \alpha + \left(\sum_{j=0}^{S} \rho_j \rho^r\right) \lambda_f \tag{3.21}$$

where λ_f is a positive scalar variable that captures the price of risk associated with the slow consumption adjustment risk, and $\alpha \in \mathbb{R}^N$. If consumption fully captures the risk of asset returns, the above expression should hold with $\alpha = \mathbf{0}_N$, otherwise α should capture the covariance between the omitted risk factors and asset returns.

Similarly, if we also allow for a bond specific latent factor (g_t), the implied cross-sectional model of returns is

$$\mathbb{E}[R_t^e] = \alpha + \left(\sum_{j=0}^{S} \rho_j \rho^r\right) \lambda_f + \left[\begin{array}{cc} \theta'^b, & \mathbf{0}'_{N-N_b} \end{array}\right]' \sum_{j=0}^{S} \theta_j \lambda_g \tag{3.22}$$

with the additional testable restriction $\lambda_f = \lambda_g$.

Equation (3.21) (and similarly Equation (3.21)), conditional on the data and the parameters of the state-space model, defines a standard cross-sectional regression, hence the parameters α, λ_f and λ_g can be estimated via standard Fama and MacBeth (1973) cross-sectional regressions. This implies that, not only we can compute posterior means and confidence bands for both the coefficients of the state space model and for the unobservable factor's time series, but we can also compute means and confidence bands for the Fama and MacBeth (1973) estimates of the cross sectional regressions defined in Equations (3.21) and

(3.22). That is, we can jointly test the ability of the slow consumption adjustment risk of explaining both the time series and the cross-section of asset returns with a simple Gibbs sampling approach described in detail in Appendix 3.A.

3.3.2 Semi-parametric Inference

We start with the pricing restriction in Euler Equation (3.5):

$$\mathbf{0} = \mathbb{E}\left[M_{t+1}^{S} R_{t+1}^{e}\right]$$

where $M_{t+1}^{S} := \left(C_{t+1+S}/C_t\right)^{-\phi} \psi_{t+1+S}$ and $S \geq 0$.

The fact that the stochastic discount factor can be decomposed into the product of the consumption growth over several consecutive periods (C_{t+1+S}/C_t) and an additional, potentially unobservable, component, makes the above setting particularly appealing for the application of Empirical Likelihood -based techniques (for an excellent overview, see Kitamura (2006)) as discussed in Ghosh, Julliard, and Taylor (2013).

Consider the following transformation of the Euler equation:

$$\begin{aligned}\mathbf{0} = \mathbb{E}\left[M_t^S R_t^e\right] &\equiv \int \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\phi} \psi_{t+S} R_t^e dP = \int \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\phi} \frac{\psi_{t+S}}{\bar{\psi}} R_t^e dP \\ &= \int \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\phi} R_t^e d\Psi = \mathbb{E}^{\Psi}\left[\left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\phi} R_t^e\right] \end{aligned} \tag{3.23}$$

where P is the unconditional physical probability measure, $\bar{\psi} = \mathbb{E}\left[\psi_{t+S}\right]$, Ψ is another probability measure, related to the physical one through the Radon-Nikodym derivative[1] $\frac{d\Psi}{dP} = \frac{\psi_{t+S}}{\bar{\psi}}$.

Empirical Likelihood provides a natural framework for recovering parameter estimates and probability measure Ψ defined by Equation (3.23), by minimising Kullback-Leibler Information Criterion (KLIC):

$$(\hat{\Psi}, \hat{\phi}) = \arg\min_{\Psi,\phi} D(P||\Psi) \equiv \arg\min_{\Psi} \int \ln \frac{dP}{d\Psi} dP \quad \text{s.t.} \quad \mathbf{0} = \mathbb{E}^{\Psi}\left[\left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\phi} R_t^e\right] \tag{3.24}$$

Equation (3.24) provides a nonparametric maximum likelihood estimation of the probability

[1]We assume absolute continuity of both P and Ψ.

measure, induced by the unobservable components of the SDF, and has been used in various applications, including the recovery of the risk-neutral probability density (Stutzer (1995)). For more information on the rationale behind this change of measure, see Ghosh, Julliard, and Taylor (2013).

Following Csiszar (1975) duality approach, one can easily show that:

$$\hat{\Psi}_t = \frac{1}{T\left(1 + \hat{\lambda}(\theta)' \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\hat{\phi}} R_t^e\right)} \quad \forall t = 1..T, \tag{3.25}$$

where $\hat{\phi}$ and $\hat{\lambda} \equiv \hat{\lambda}(\hat{\phi}) \in \mathbb{R}^n$ are the solution to the dual optimisation problem:

$$\hat{\phi} = \underset{\phi \in \mathbb{R}}{\arg\min} \; -\sum_{t=1}^{T} \ln\left(1 + \hat{\lambda}(\phi)' \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\phi} R_t^e\right) \tag{3.26}$$

$$\hat{\lambda}(\phi) = \underset{\lambda \in \mathbb{R}^n}{\arg\min} \; -\sum_{t=1}^{T} \ln\left(1 + \lambda(\phi)' \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-\phi} R_t^e\right) \tag{3.27}$$

The dual problem is usually solved via the combination of internal and external loops (Kitamura (2001)): first, for each ϕ find the optimal values of the Langrange multipliers λ, as in Equation (3.27); then minimize the value of the dual objective function w.r.t. $\phi(\hat{\lambda})$, following Equation (3.26).

Empirical likelihood estimator is known not only for its nonparametric likelihood interpretation, but also for its convenient asymptotic representation and properties. It belongs to the family of Generalised Empirical Likelihood estimators (Newey and Smith (2004)), with other notable members including the Exponentially Tilted Estimator (ET, Kitamura and Stutzer (1997)) and the Continuously Updated GMM (CU-GMM, Hansen, Heaton, and Yaron (1996)). While the whole family enjoys the same asymptotic distribution of the parameter estimates, achieves the semiparametric efficiency bound of Chamberlain (1987), and shares the standard battery of tests for moment equalities (e.g. J-test), the empirical likelihood estimator is also higher-order efficient (Newey and Smith (2004), Anatolyev (2005)).

We can also capture the average pricing error of the model implied by Equation (3.5) simply by introducing additional parameters in the following way:

$$\mathbf{0} = \mathbb{E}\left[M_{t+1}^{S}\left(R_{t+1}^{e} - \alpha\right)\right], \tag{3.28}$$

where α stands for the average rate of return that is not cross-sectionally captured through the covariance between M_{t+1}^{S} and R_{t+1}^{e}, since Equation (3.28) implies

$$\mathbb{E}\left[R_{t+1}^{e}\right] = \alpha - \frac{Cov\left(M_{t+1}^{S}, R_{t+1}^{e}\right)}{\mathbb{E}[M_{t+1}^{S}]}. \tag{3.29}$$

Parameter estimation proceeds in exactly the same way, following the procedure outlined in Equations (3.24)-(3.27). We consider several versions of Equation (3.28): $\alpha = 0$ (correct model specification); average pricing errors; error specific to a particular asset class ($\alpha_b \neq \alpha_s$); and a common level of mispricing for both stocks and bonds ($\alpha_b = \alpha_s$).

For each model we also report the cross-sectional adjusted R-squared

$$R_{adj}^{2} = 1 - \tfrac{n-2}{n-1}\,\widehat{Var}_c\left(\tfrac{1}{T}R_{i,t+1} - \hat{\alpha} - \frac{\widehat{Cov}\left((C_{t+1+S}/C_t)^{-\hat{\phi}}, R_{t+1}^{e}\right)}{\mathbb{E}[(C_{t+1+S}/C_t)^{-\hat{\phi}}}\right) \Bigg/ \widehat{Var}_c\left(\tfrac{1}{T}R_{i,t+1}\right) \tag{3.30}$$

where $\widehat{Var}_c$ is the sample cross-sectional variance and $\widehat{Cov}$ is the sample time series covariance.

Finally, for the sake of completeness we also use two-stage Generalised Method of Moments (GMM, Hansen (1982)) to estimate consumption-based asset pricing models on the cross-section of stock and bond returns, and report its results alongside those for EL. While the estimator-implied probabilities no longer have the convenient nonparametric maximum likelihood interpretation (unlike those in Equation (3.25)), if one restricts the class of admissible SDF to the external habit models, asset pricing implications and inference based on the *ultimate consumption risk* only, remain valid. Under fairly general conditions, this result is a direct consequence of Proposition 1 in Parker and Julliard (2003), implying that GMM estimates of risk aversion retain consistency and asymptotical normality, and do not require an explicit knowledge of the habit function, if one relies on the ultimate consumption risk in the estimation.

3.4 Data Desciption

Bond holding returns are calculated on a quarterly basis using the zero coupon yield data constructed by Gurkaynak and Wright (2007)[1] from fitting the Nelson-Siegel-Svensson curves daily since June 1961, and excess returns are computed subtracting the return on a three-month Treasury bill. We consider the set of the following maturities: 6 months, 1, 2, 3, 4, 5, 6, 7, and 10 years, which gives us a set of 9 bond portfolios.

We consider several portfolios of stock returns. The baseline specification relies, in addition to the bond portfolios, on the 25 size and book-to-market Fama-French portolios (Fama and French (1992)), and 12 industry portfolios, available from Kenneth French data library. We consider monthly returns from July, 1961 to December, 2013, and accumulate them to form quarterly returns, matching the frequency of consumption data. Excess returns are then formed by subtracting the correponding return on the three-month Treasury bill.

Consumption flow is measured as real (chain-weighted) consumption expenditure on non-durable goods per capita available from the National Income and Product Accounts (NIPA). We use the end-of-period timing convention and assume that all of the expenditure occurs at the end of the period between t and $t+1$. We make this (common) choice because under this convention the entire period covered by time t consumption is part of the information set of the representative agent before time $t+1$ returns are realised. All the returns are made real using the corresponding consumption deflator.

Overall, this gives us consumption growth and matching real excess quarterly holding returns on a number of portfolios, from the forth quarter of 1961 to the end of 2013.

3.5 Empirical Evidence

While our model allows for a potentially infinite number of lags for the consumption process, in order to proceed with the actual estimation one has to choose a particular value of $\bar{S}$. For the rest of the section we use $S = 15$ for a number of reasons.

First, we rely on the previous results of Parker and Julliard (2003), who demonstrate that most of the pricing ability of the ultimate consumption risk is contained within the time span of 15 quarters. They define a proxy for the signal-to-noise ratio, taking into account both

[1] The data is regularly updated and available at:
http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html

the time-series and cross-sectional variation of the data, and find that the maximum (as well as the best overall fit) is obtained around $S = 11$.

Second, Equation (3.8) implies a certain autocorrelation structure of the nondurable consumption growth through the combination of the common factor lags and its loadings. Hence, the value of S should be high enough to capture most of the time series autocorrelation in the consumption growth. Figure 3.C.1 in the Appendix presents the sample autocorrelation coefficients and the results of Ljung-Box (1978) and Box-Pierce (1970) tests. Since most of the dependence occurs within the first 15 lags, this value also becomes a natural choice for the lag truncation.

Further, intuitively most of the pricing impact from the consumption adjustment is probably taking place within the business cycle frequency, consistent with a number of recent empirical studies (e.g. Bandi and Tamoni (2015)). Therefore, $S = 15$ is a rather conservative choice, since it provides a 4 year window to capture most of the interaction between the ultimate consumption and returns.

Finally, our results remain robust to including additional lags.

3.5.1 Parametric Approach

We start our analysis by examining the time-series properties of a one (common) factor model implied by Equations (3.8)–(3.9). We then turn to the 2-factor specification described by Equations (3.12)–(3.13). Finally, we present the cross-sectional properties of the model and demonstrate that the slow consumption adjustment risk is a priced factor, explaining a significant proportion of the cross-sectional variation in returns.

3.5.1.1 Time Series Properties of Stocks and Bonds

Our baseline cross-section consists of 9 bond portfolios, 25 Fama-French portfolios sorted by size and book-to-market, and 12 industry portfolios. We estimate the model in Equations (3.8)–(3.9) using the inference procedure outlined in Section 3.3.1. Figures 3.1 and 3.2 present stock and bond loadings on the common factor, along with the 68% and 90% confidence bounds.

All the portfolios in Figure 3.1 display significant and positive exposures to the common factor. However, even more interesting is a widely recognisable pattern in the factor loadings: decreasing from the smallest to the largest decile on size, with a similar effect for book-to-

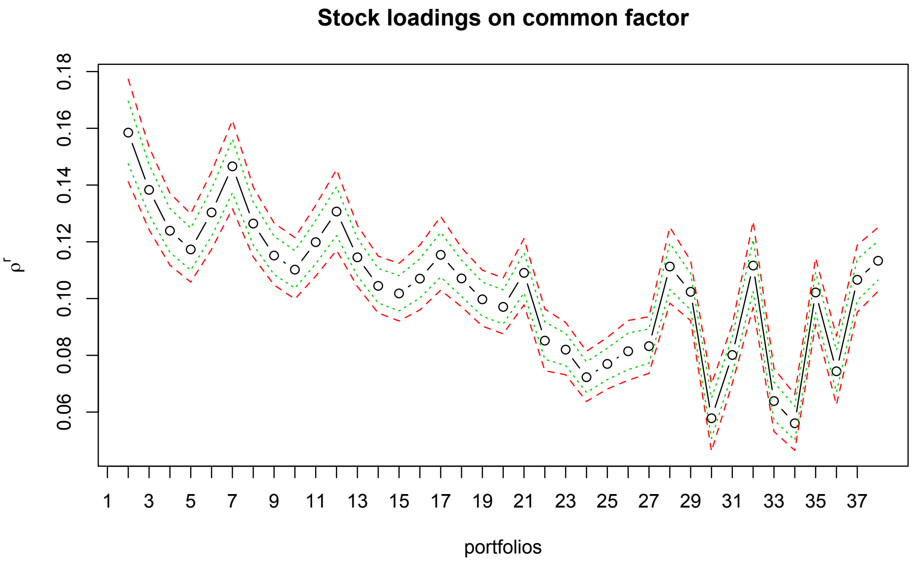


Figure 3.1: Common factor loadings (ρ^r) of the stock portfolios in the one-factor model.

Note. The graph presents posterior means (continuous line with circles) and centred posterior 90% (dashed line) and 68% (dotted line) coverage regions. Ordering of portfolios: 25 Fama and French (1992) size and book-to-market sorted portfolios (e.g. portfolio 2 is the smallest decile of size and the second smaller decile of book-to-market ratio), and 12 industry portfolios.

market sorting. This is in line with the size and value anomalies and, in addition, provides some preliminary evidence that the SCA risk plays an important role in explaining the cross-sectional dispersion of stocks returns. These findings also remain unchanged, when a second, bond-specific factor is added into the model (see Figure 3.3, lower panel).

In a single factor model, bond loadings, however, are not as prominent (Figure 3.2). While there is some evidence in favour of their increase with the bond maturity, the magnitude is still considerably smaller than that of the stocks.

Figures 3.2 (upper panel) and 3.4 highlight the importance of adding a bond-specific factor into the model. While the cross-section of bonds reveals a very pronounced maturity-driven pattern of loadings on the bond-specific factor, g_t, its presence also allows to highlight the effect of the consumption-related component. Compared to a one factor specification, these loadings are still not as high as those of the stocks, however, they are contained within very tight confidence bounds, are significantly different from zero (except for the 6 months return), and generally increase with maturity.

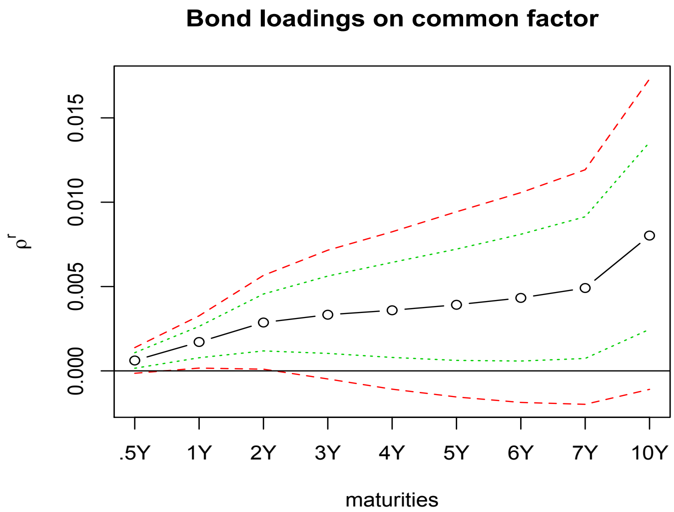


Figure 3.2: Bond loadings (ρ^r) on the common factor (f_t).

Note. The graph presents posterior means (continuous line with circles) and centred posterior 90% (dashed line) and 68% (dotted line) coverage regions.

To summarise, not only (nearly) all the assets in the mixed cross-section of stocks and bonds have a significant positive exposure to the innovations in the ultimate consumption growth, the pattern of those loadings reflects well-known features of the data: size and value anomalies for stocks, and positive slope of the yield curve for bonds.

One of the possible concerns could be that we inadvertently capture a factor that heavily loads on one of the principal components of the cross-section of asset returns and thus mechanically has rather high factor loadings (Lewellen, Nagel, and Shanken (2010)). However, this is not the case. While there is indeed some correlation with the principal components of the cross-sections, composed of different assets (see Table 3.1), the common factor does not heavily correlate with any of them in particular, but rather displays a certain degree of spread in loadings. For example, it is related to the first, third and fourth principal components of the joint cross-section of stocks and bonds. Therefore, we conclude that our results are not driven by a particular implied factor structure of a given cross-section, but rather reflect a more general feature of the data.

The economic magnitude of asset exposure to the SCA risk can in turn be assessed by the standard variance decomposition techniques. Figure 3.5 summarises our results. The common factor explains on average 79% of the time series variation in the stock returns,

Table 3.1: Correlation of Slow Consumption Adjustment with Principal Components

	Correlation of:									
	$\sum_{j=0}^{\bar{S}} \hat{\rho}_j \hat{f}_{t-j}$					$\sum_{j=0}^{\bar{S}} \hat{\theta}_j \hat{g}_{t-j}$				
PCA of:	I	II	III	IV	V	I	II	III	IV	V
r^e	-.37	.01	-.13	-.17	.03	-.03	-.32	-.01	-.54	.04
r^e_{bonds}	.11	-.12	.10	.15	-.03	.64	-.10	.01	.06	-.08
r^e_{stocks}	.38	.08	-.11	.01	-.01					

ranging from 36% to nearly 95% for individual portfolios. Moreover, this level of fit in our model is produced by a single consumption-based factor, as opposed to some of the alternative successful specifications, relying on 3 and sometimes 4 explanatory variables.

The same common factor accounts for a small (about 1.5%), but significant proportion of the time series variation in bond returns as well. The bond-specific factor, in turn, manages to capture most of the residual time series in variation in returns. While the model captures just about 55% of the variation in the 6-month bond returns, its performance rapidly improves with maturity and results in a nearly perfect fit for the time horizon of 2 years and more.

3.5.1.2 Consumption Process and its Properties

Slow consumption adjustment explains a significant proportion of the time series variation in consumption growth. As Figure 3.5 demonstrates, the common factor is responsible for roughly 27% of the variation in the one-period nondurable consumption growth, 33% of the two-period consumption growth, and so on, followed by a slow decline towards just above 5% for the 15-period growth. The bond-specific factor amounts for an additional 5% of the explanatory power. While these numbers may not seem as impressive as those for the cross-section of stocks, the pattern is highly persistent and significant, confirming a common factor structure between nondurable consumption growth and asset returns. Further, it also allows to use the model in Equations (3.12)–(3.13) for predictive purposes.

The upper right panel in Figure 3.5 displays the outcome of the predictive regression for the one-period consumption growth, should one rely on the factor loadings from Equation 3.12. Ultimate consumption risk predicts about 27% of the time series variation in the next period consumption and 18% of the consumption growth 2 quarters from now. Interestingly,

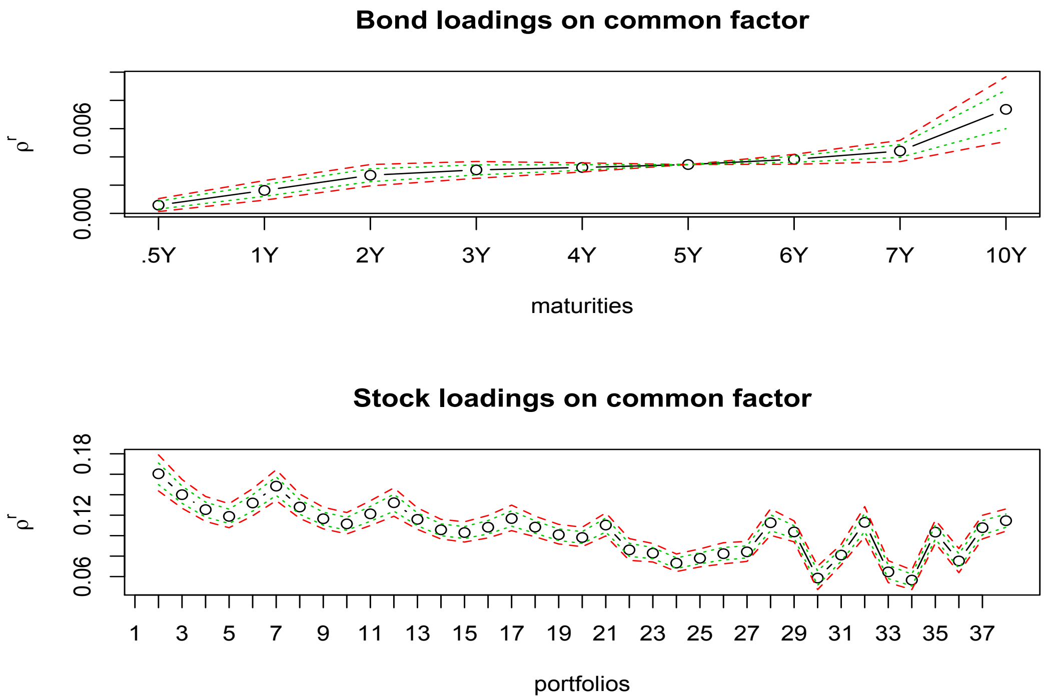


Figure 3.3: Bond and stock loadings on the common factor (f_t).

Note. Upper panel: loadings of bonds (ρ^r) on common factor (f_t). Lower panel: loadings of stock portfolios (ρ^r) on common factor (f_t). The graph presents posterior means (continuous line with circles) and centred posterior 90% (dashed line) and 68% (dotted line) coverage regions. Ordering of the portfolios: 25 Fama and French (1992) size and book-to-market sorted portfolios (e.g. portfolio 2 is the smallest decile of size and the second smaller decile of book-to-market ratio), and 12 industry portfolios.

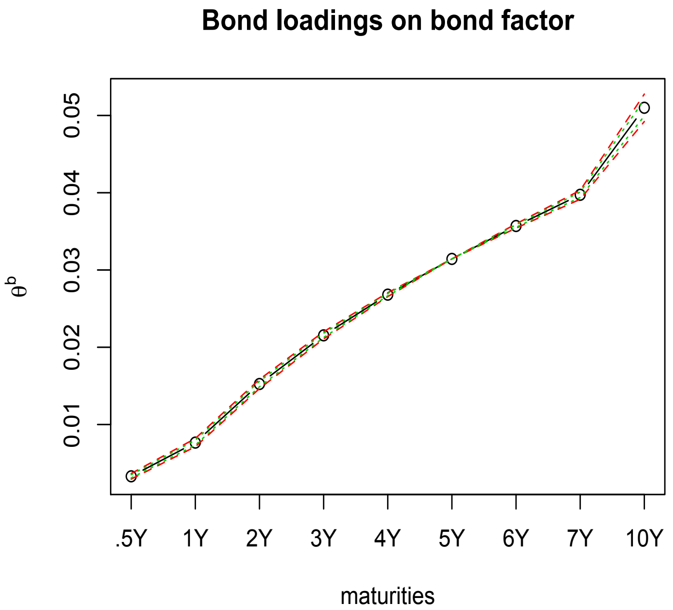
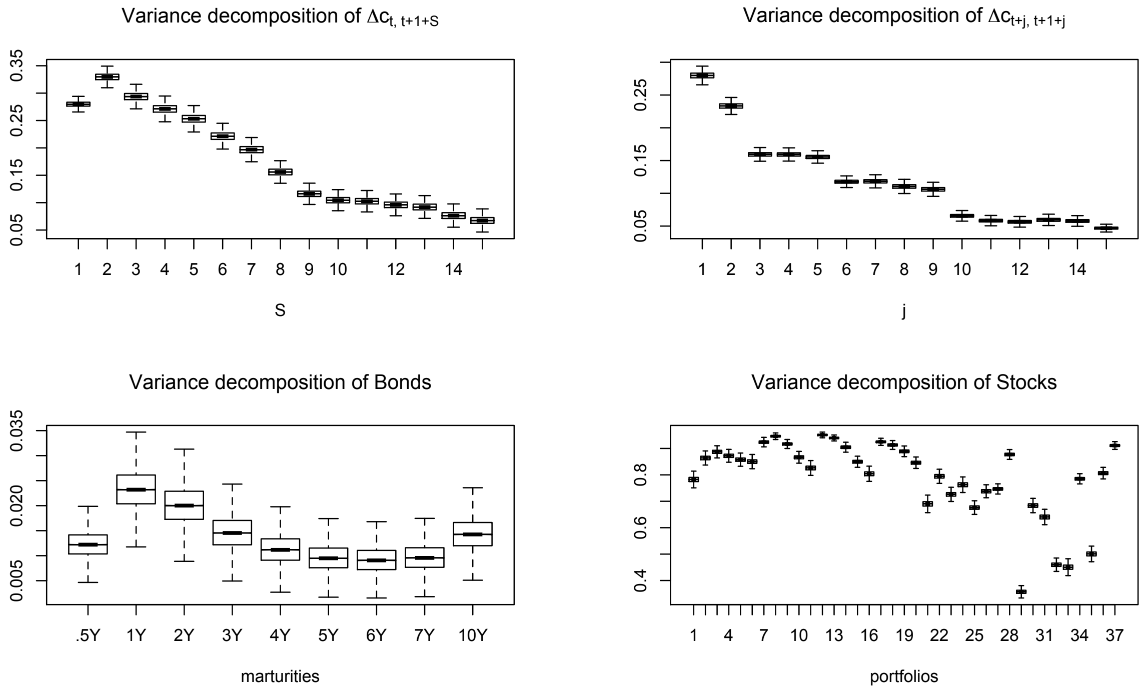
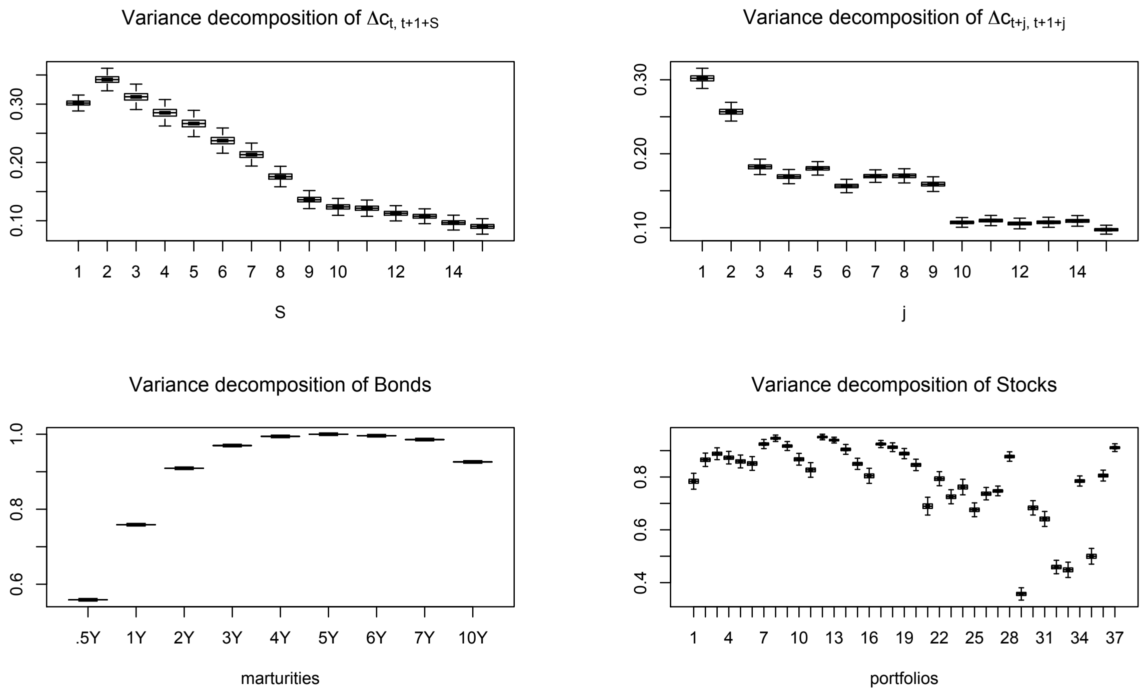


Figure 3.4: Bond loadings (θ^b) on the bond-specific factor (g_t).

Note. The graph presents posterior means (continuous line with circles) and centred posterior 90% (dashed line) and 68% (dotted line) coverage regions.



(a) Box-plots of percentage of time series variances explained by the common component f_t.



(b) Box-plots of percentage of time series variance explained jointly by the common component f_t and the bond component g_t.

Figure 3.5: Variance decomposition box-plots of asset returns and consumption growth

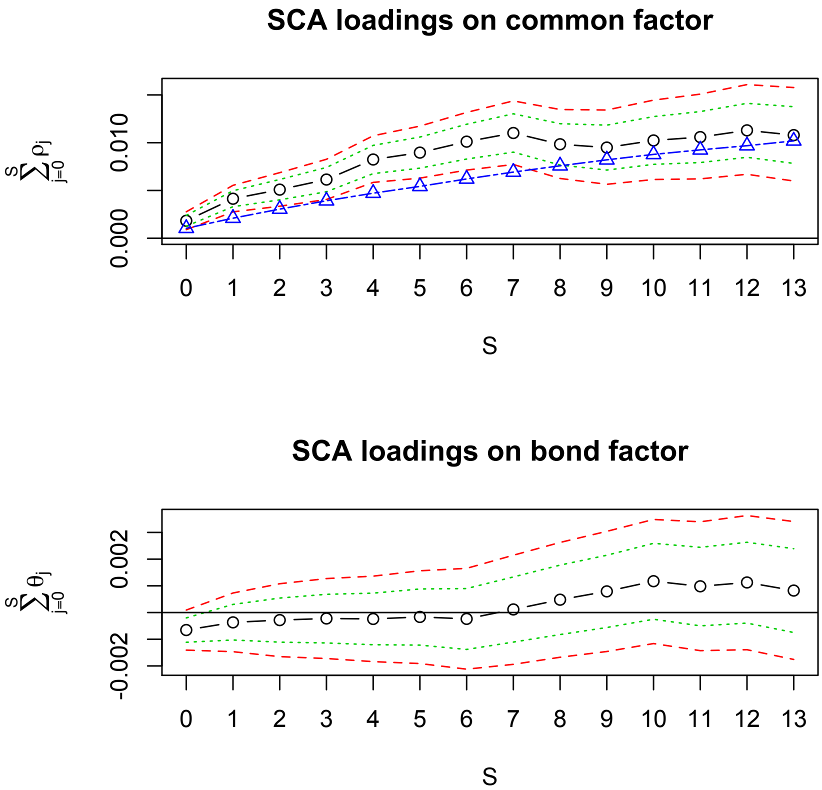


Figure 3.6: Slow consumption adjustment ($\Delta c_{t,t+1+S}$) response to shocks.

Note. Upper panel: SCA response to common factor (f_t) shocks, lower panel: SCA response to bond only factor (g_t) shocks. The graph presents posterior means (continuous line with circles) and centred posterior 90% (dashed line) and 68% (dotted line) coverage regions. Triangles denote Bansal and Yaron (2004) implied values.

the model retains significant predictive power (albeit, much lower) even for the one-period consumption that will occur nearly 4 years from now. A bond-specific factor increases the quality of predictive regressions by roughly another 5%.

The consumption growth process in Equation (3.12) is similar to the moving average decomposition, which allows us to model the dynamics of the slow consumption adjustment ($\Delta c_{t,t+1+S}$) in response to a common and/or a bond-specific shock. Figure 3.6 depicts SCA loadings on the factors as a function of the horizon S. If $S = 0$, the case of a standard consumption-based asset pricing model, SCA virtually does not load on the common factor. This is expected, since the factor manifests itself at a lower frequency. Indeed, as S increases, the impact of the common factor becomes more and more pronounced, levelling off at around $S = 11$. Interestingly, the pattern of the loadings observed in our two-factor model, is very similar to the one implied by the moving average representation of the consumption process in Bansal and Yaron (2004)[1]. In short, our setting reveals a similar degree of persistency and response rates, as their consumption process. The pricing of stocks and bonds, however, differs, because we consider a more flexible, reduced form model that nevertheless uncovers a very similar consumption-related pattern in the data as the one implied by the long-run risk model.

As a robustness check, we recover the long-run impact of common innovations to financial market returns and nondurable consumption using a simple bivariate SVAR model for the market excess return and consumption growth. We achieve identification via long-run restrictions on the impulse response functions á la Blanchard and Quah (1989). In particular, we distinguish a fundamental long-run shock, that can have a long-run impact on both market return and consumption, and a transitory shock that is restricted not to have a long-run impact on asset prices.

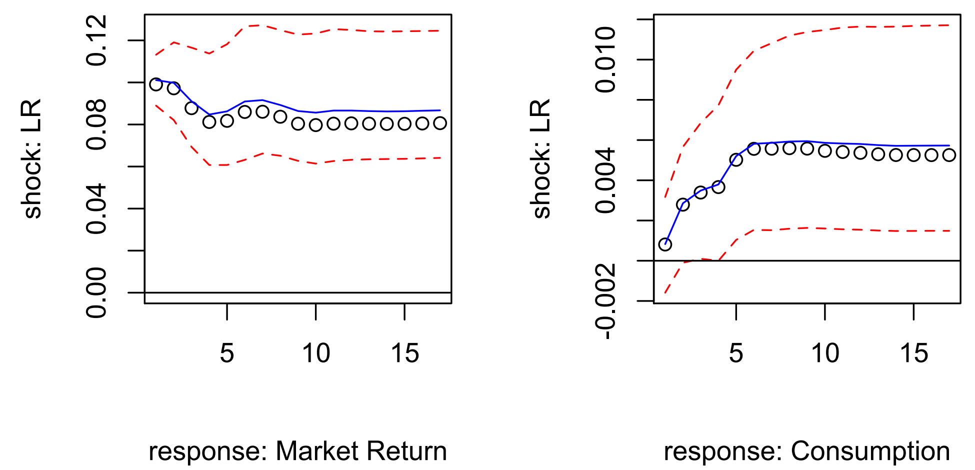


Figure 3.7: Cumulated response functions to a long-run shock

Note. The shock identified via a VAR and imposed long-run restrictions. Left panel depicts the cumulated response function for the market return, while the right one - for consumption growth. The graphs include posterior median (continuous line), mean (circles), and centred 95% coverage region (dotted lines).

Figure 3.7 displays the cumulated impulse response functions to a long-run fundamental shock, that is allowed to have a potemtially permanent impact on both the market excess

[1]For more details on the construction of the MA representation, see Appendix 3.B

return and nondurable consumption. In line with our previous reasoning, the latter response to a shock (right panel) is very similar to the one we observed from the SCA loadings on the common factor (Figure 3.6), while the response of the market returns (left panel), is consistent with an immediate and complete reaction of asset returns to the long-run shock as in our state-space model in Equations (3.8)-(3.9).

All these observations confirm that within the stream of nondurable consumption flow there is a rather persistent slow-moving component, accounting for 28% of the one-period time series variation in consumption growth, with innovations of that factor driving most of the contemporaneous changes in stocks returns and a small, but significant proportion in bonds. Next, we investigate whether this risk is actually *priced* in the cross-section of assets.

3.5.1.3 The Price of Consumption Risk

Recovering factor loadings in Equations (3.12)–(3.13) also produces a cross-section of average returns on the set of portfolios. Figure 3.8 displays the scatterplot of the average vs. fitted excess returns for the baseline mixed cross-section of 46 assets. While the subset of bond returns demonstrates an almost perfect fit (lower left corner of the plot), the variation in the cross-section of stocks is also well-captured.

Further, as Equation (3.22) demonstrates, model-implied factor loadings of the asset returns determine their full exposure of the SCA risk and thus allow not only to assess the cross-sectional fit of the model, but also to test whether the slow consumption adjustment is indeed a *priced* risk factor, and whether the common and bond factors share the same value of the risk premium.

Following the critique of Lewellen, Nagel, and Shanken (2010), we are using a mixed cross-section of assets to ensure that there is no dominating implied factor structure of the returns. Indeed, if that was the case, it could lead to spuriously high significance levels, quality of fit, and significantly complicate the overall model assessment. However, as Table 3.1 indicated, the slow consumption adjustment factor does not heavily load on any of the main principal components of the returns. Further, we provide confidence bounds for the cross-sectional measure of fit to ensure the point estimates reflect the actual pricing ability of the model. Finally, since both stocks and bonds have significant loadings on the common factor (and in the case of bonds, also on the bond-specific one), we do not face the problem of *irrelevant*, or *spurious* factors (Kan and Zhang (1999b)), that could also lead to the

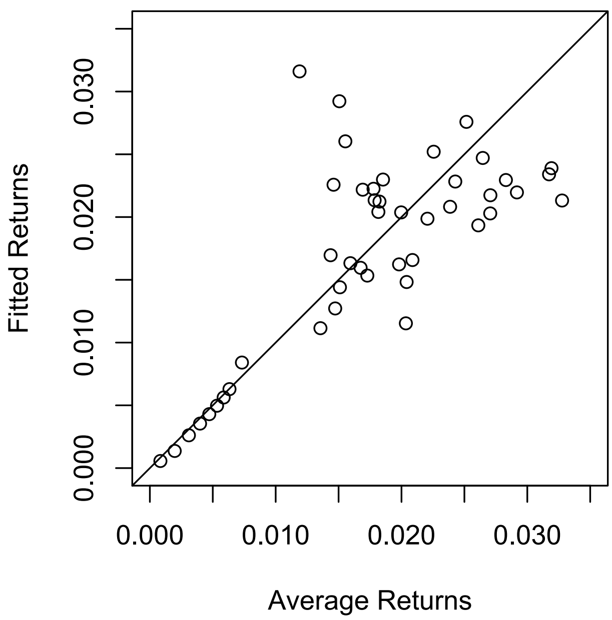


Figure 3.8: SCA risk: Average and Fitted Excess returns.

Note. Fitted versus average returns using the consumption betas implied by the latent factor specification in Equations (3.12)–(3.13).

unjustifiably high significance levels.

Table 3.2 summarizes the cross-sectional pricing performance of our parametric model of consumption on a mixed cross-section of 9 bond portfolios, 25 Fama-French portfolios sorted by size and book-to-market, and 12 industry portfolios. For each of the specifications, we recover the full posterior distribution of the factor loadings, and estimate the associated risk premia using Fama-MacBeth (1973) cross-sectional regressions. Regardless of the specification, there is strong support in favour of the slow consumption adjustment being a priced risk in the composite cross-section of assets with the risk premia of about 14-20% per quarter.

The average pricing error is about 0.005% per quarter, and the cross-sectional R^2 varies from 57% to 91%, depending on whether the intercept is included in the model. While allowing for a common intercept in the estimation substantially lowers cross-sectional fit, 95% posterior coverage bounds remain very tight, providing a reliable indicator of the model

Table 3.2: Cross-Sectional Regressions with State-Space Loadings

Row:	α	λ_f	λ_g	$\lambda_f = \lambda_g$	$\bar{R}^2$
		One latent factor specification			
(1)	.0056 [0.0051, .0062]	14.77 [8.89, 26.01]			.57 [.54, .60]
(2)		20.00 [12.05, 35.16]			.90 [.89, .91]
		Two latent factor specification			
(3)	.0057 [.0052, .0061]	14.97 [8.72, 27.45]			.57 [.54, .60]
(4)		20.30 [11.85, 37.18]			0.90 [0.89, 0.91]
(5)	.0069 [−539.5, 497.7]	13.79 [7.96, 25.49]	−1.44 [−539.5, 497.7]		.56 [0.53, 0.59]
(6)		20.27 [11.83, 37.12]	19.57 [−1140, 1218]		.91 [.90, .92]
(7)	.0053 [.0042, 0.0064]			15.24 [8.80, 28.40]	.57 [.53, .60]
(8)				20.29 [11.85, 37.19]	.90 [.89, .91]

Note. The table presents posterior means and centred 95% posterior coverage (in square brackets) of the Fama and MacBeth (1973) cross-sectional regression of excess returns on $\sum_{j=0}^{S} \rho_j \rho^r$ (with associated coefficient λ_f) and $\left[\ \theta'^b, \ \ \mathbf{0}'_{N-N_b} \ \right]' \sum_{j=0}^{S} \theta_j$ (with associated coefficient λ_g). The column labeled $\lambda_f = \lambda_g$ reports restricted estimates. Cross-section of assets: 25 Fama and French (1992) size and book-to-market portfolio; 12 industry sorted portfolios; 9 bond portfolios.

performance.

While the risk premium on the common factor is strongly identified and seems to play an important role in explaining the cross-section of both stock and bond returns, the bond factor loadings do not provide an equally large spread for recovering its pricing impact with the same degree of accuracy. As a result, the risk premium appears to be insignificant, unless its value is restricted to that of the common factor. To summarise, the bond-specific factor is *unspanned*, in the sense that while it is essential for explaining most of the time series variation in bond returns and producing a correct slope of the yield curve, it does not have any cross-sectional impact on bond returns.

3.5.2 Semi-parametric approach

Since the relevance of slow consumption adjustment risk for the cross-section of stocks has already been highlighted by Parker and Julliard (2005), we first focus on the cross-section of bonds only, and provide empirical evidence that the SCA risk is important for explaining their cross-section of returns. We then turn to analysing the model performance for pricing a composite set of bonds and stocks.

Table 3.3 summarizes the performance of the consumption-based asset pricing model on the cross-section of bond returns for various values of S of the ultimate consumption measure of Parker and Julliard (2005). While EL estimation remains valid in the presence of the multiplicative unobservable part of the stochastic discount factor, evaluating GMM output requires a certain degree of caution, since in this case, to the best of our knowledge, the same robustness is achieved only within the class of external habit models (see Proposition 1 of Parker and Julliard (2003)). Nevertheless, for the sake of completeness we report both sets of results.

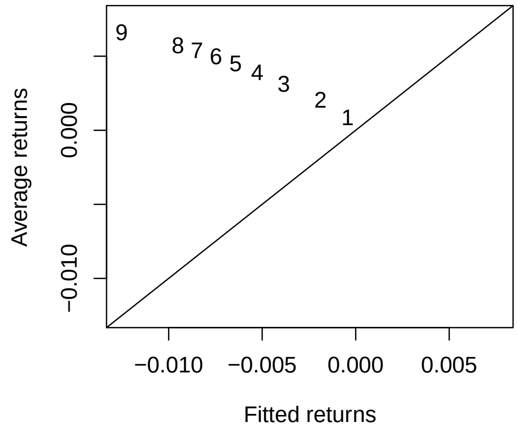
The $S = 0$ case corresponds to the standard consumption-based asset pricing model, where the spread of the returns is driven only by their contemporaneous correlation with the consumption growth. Both EL and GMM output reflect the well-known failure of the classical model to capture the cross-section of bond returns: according to the J-test, the model is rejected in the data, and the cross-sectional adjusted R-squared is negative. Increasing the span of consumption growth to 2 or more quarters drastically changes the picture: J-test no longer rejects the model, and the level of cross-sectional fit increases up to 85% for $S = 12$, for example.

Further, the estimates of the power coefficient ϕ (which in the case of additively separable CRRA utility corresponds to the Arrow-Pratt relative risk-aversion coefficient) not only appear to be much smaller (hence more in line with the economic theory), but also more precisely estimated. The large standard error associated with this parameter for the standard consumption-based model ($S = 0$) is due to the fact that the level and spread of the contemporaneous correlation between asset returns and consumption growth is rather low. This in turn leads to substantial uncertainty in parameter estimation. As the number of quarters used to measure consumption risk increase, the link between bond returns and the slow moving component of the consumption becomes more pronounced, resulting in lower standard errors, better quality of fit, and the overall ability of the model to match the cross-section of bond returns. In fact, model-implied average excess returns are very close to the actual ones, in drastic contrast to the standard consumption-based asset pricing model. This is shown in Figure 3.9 which presents fitted and actual average excess returns on the cross-section of 9 bond portfolios for several values of the consumption horizon S.

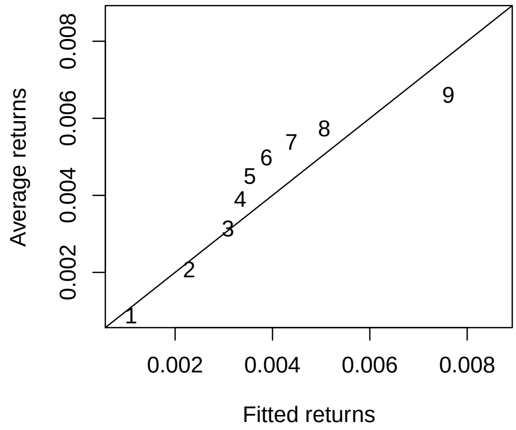
Table 3.3: Cross-Section of Bond Returns and Ultimate Consumption Risk

	Empirical Likelihood				Generalised Method of Moments			
Horizon S (Quarters)	R^2_{adj}(%) (1)	α (2)	ϕ (3)	J-test (4)	R^2_{adj}(%) (5)	α (6)	ϕ (7)	J-test (8)
0	-837	0.0007	100	13.0888	-10	0.0000	4	19.5597
		(0.0003)	(28.5)	[0.0700]		(0.0002)	(73.5)	[0.0066]
1	-167	0.0009	88	7.6457	-35	0.0005	42	11.5448
		(0.0004)	(24.8)	[0.3649]		(0.0005)	(47.3)	[0.1166]
2	70	0.0030	120	2.8778	43	0.0009	50	4.6351
		(0.0005)	(21.8)	[0.8961]		(0.0011)	(52.6)	[0.7044]
3	39	0.0010	70	4.5187	61	0.0006	35	5.1968
		(0.0004)	(16.2)	[0.7185]		(0.0005)	(20.5)	[0.6360]
4	69	0.0008	55	3.4531	48	0.0004	33	3.2207
		(0.0003)	(13.4)	[0.8402]		(0.0004)	(16.1)	[0.8639]
5	5	0.0008	45	6.8134	38	0.0004	27	6.0294
		(0.0003)	(10.5)	[0.4486]		(0.0003)	(13.0)	[0.5363]
6	3	0.0008	42	8.9256	42	0.0002	23	6.8397
		(0.0003)	(10.0)	[0.2580]		(0.0003)	(11.5)	[0.4458]
7	64	0.0004	33	9.8236	64	0.0001	22	6.4740
		(0.0003)	(9.9)	[0.1988]		(0.0003)	(10.7)	[0.4856]
8	70	0.0006	35	9.6027	69	0.0003	24	6.5862
		(0.0003)	(10.1)	[0.2122]		(0.0003)	(12.3)	[0.4732]
9	53	0.0008	55	8.2778	67	0.0004	26	6.8314
		(0.0003)	(10.5)	[0.3087]		(0.0003)	(14.7)	[0.4466]
10	77	0.0008	38	10.2472	73	0.0004	25	6.8649
		(0.0002)	(12.3)	[0.1750]		(0.0003)	(18.4)	[0.4431]
11	77	0.0008	44	8.2683	72	0.0006	26	7.7110
		(0.0002)	(14.3)	[0.3095]		(0.0003)	(23.7)	[0.3588]
12	85	0.0008	78	6.1561	88	0.0008	34	6.8054
		(0.0002)	(16.3)	[0.5216]		(0.0003)	(26.5)	[0.4494]
13	69	0.0007	85	5.8494	89	0.0007	37	6.0817
		(0.0002)	(17.5)	[0.5574]		(0.0003)	(28.7)	[0.5302]
14	88	0.0006	72	8.0283	90	0.0007	41	6.7445
		(0.0002)	(19.6)	[0.3301]		(0.0004)	(30.3)	[0.4560]
15	77	0.0006	70	7.3656	69	0.0008	46	7.2723
		(0.0002)	(22.1)	[0.3918]		(0.0005)	(36.4)	[0.4011]

Note. The table reports the pricing of 9 excess bond holding returns for various values of the horizon S, and allowing for an intercept. Standard errors are reported in parentheses and p-values in brackets. Estimation is done using EL and two-stage GMM.



(a) S=0



(b) S=12

Figure 3.9: Slow consumption adjustment factor and the cross-section of bond returns

Note. The figures show average and fitted returns on the cross-section of 9 bond portfolios (1961Q1-2013Q4), sorted by maturity. The model is estimated by Empirical Likelihood for various values of consumption horizon S. $S = 0$ corresponds to the standard consumption-based asset pricing model; $S = 12$ corresponds to the use of ultimate consumption risk, where the cross-section of returns is driven by the their correlation with the consumption growth over 13 quarters, starting from the contemporaneous one.

The contemporaneous correlation between bond returns and consumption growth (Panel A, $S = 0$) is so low that not only it results is rather poor fit, but actually reverses the order of the portfolios: i.e. the fitted average return from holding long-term bonds is smaller than that of the short term ones. And again, once the horizon used to measure consumption risk is increased, the quality of fit substantially improves, leading to an R-squared of 85% for $S = 12$ (see Panel on the right).

The ability of slow consumption adjustment risk to capture a large proportion of the cross-sectional variation in returns is not a feature of the bond market alone: it works equally well on the joint cross-section of stocks and bonds, providing a simple and parsimonious one factor model for co-pricing securities in both asset classes.

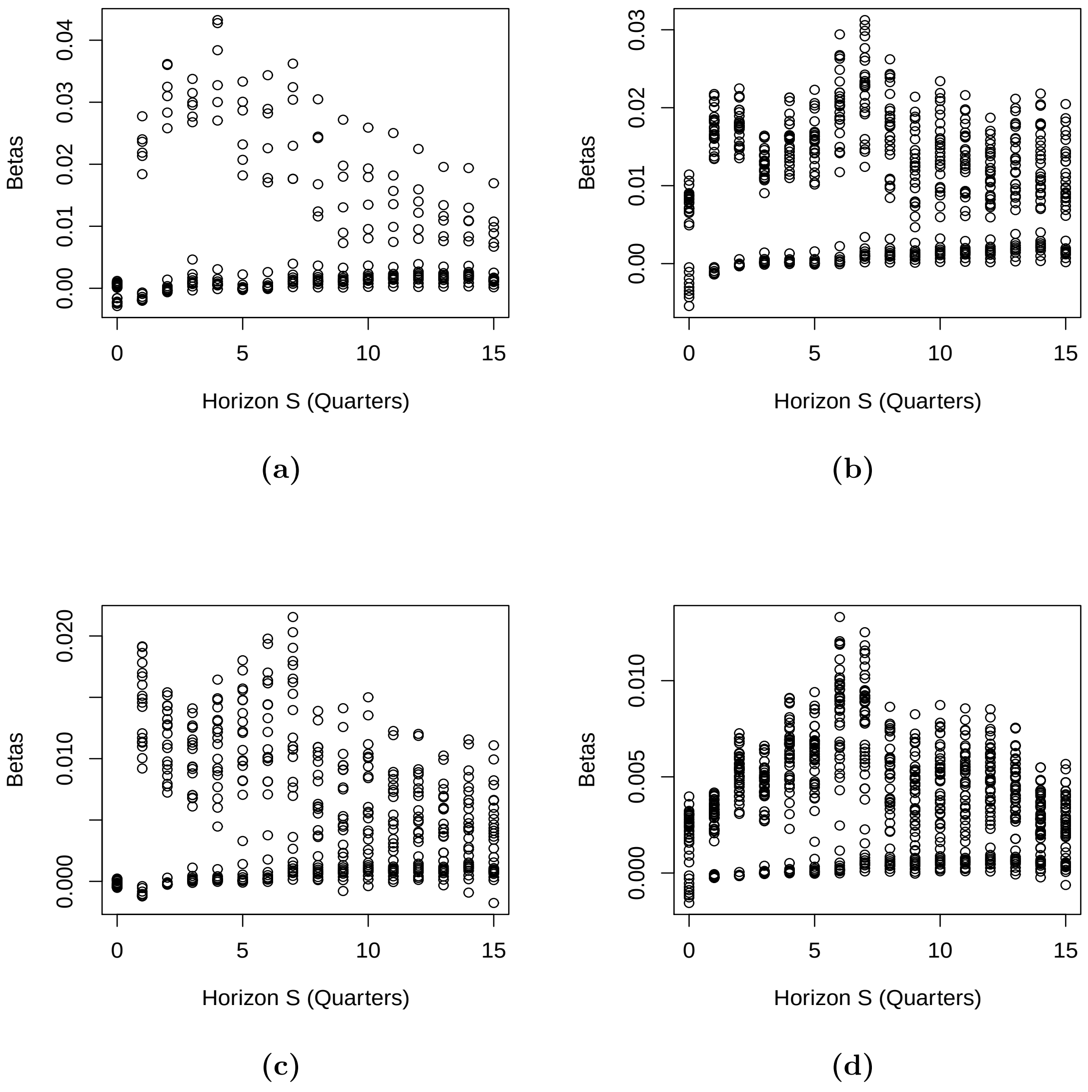
Table 3.4 summarises the model performance with various joint cross-sections of stocks and bonds for different consumption horizons S. Compared to the standard case of $S = 0$, slow consumption adjustment substantially improves model performance in a number of ways. While a simple consumption-based asset pricing model is rejected by the J-test on all the cross-section of stocks, the test values are dramatically improved over the range of $S =$

Table 3.4: Expected Excess Returns and Consumption Risk, 1967:Q3-2013:Q4

	Empirical Likelihood			Generalised Method of Moments		
Horizon S	$R^2_{adj}(\%)$	ϕ	J-test	$R^2_{adj}(\%)$	ϕ	J-test
(Quarters)	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 9 Bonds and Fama-French 6 portfolios						
0	-13	-7	36.8568	70	60	36.3730
		(26.3)	[0.0013]		(27.7)	[0.0016]
10	95	23	7.275274	89	30	28.3589
		(6.0)	[0.9495]		(6.8)	[0.0194]
11	94	23	6.389318	94	32	29.3379
		(6.5)	[0.9724]		(8.5)	[0.0145]
12	91	22	5.864083	96	35	30.5354
		(6.3)	[0.9819]		(9.4)	[0.0101]
Panel B: 9 Bonds and Fama-French 25 portfolios						
0	46	41	56.7788	64	73	157.2452
		(17.8)	[0.0084]		(15.0)	[0.0000]
10	75	20	24.3141	24	41	31.8799
		(3.8)	[0.8899]		(6.1)	[0.5719]
11	76	20	21.2727	45	21	26.3571
		(3.7)	[0.9563]		(6.3)	[0.8224]
12	70	18	20.9430	49	22	22.3989
		(3.4)	[0.9612]		(7.7)	[0.9364]
Panel C: 9 Bonds, Fama-French 6, and Industry 12 portfolios						
0	-6	-3	59.7497	59	68	156.2215
		(21.2)	[0.0003]		(15.0)	[0.0000]
10	54	13	24.2148	-68	40	22.9235
		(3.9)	[0.6184]		(7.0)	[0.6891]
11	51	12	24.2189	-38	42	22.0777
		(3.7)	[0.6181]		(7.0)	[0.7334]
12	52	12	22.1532	-3	45	22.0186
		(3.5)	[0.7295]		(6.4)	[0.7364]
Panel D: 9 Bonds, Fama-French 25, and Industry 12 portfolios						
0	22	19	82.6606	50	86	213.7053
		(15.1)	[0.0007]		(14.2)	[0.0000]
10	37	8	52.2543	-48	42	48.9612
		(2.5)	[0.2440]		(5.3)	[0.3551]
11	37	8	49.6145	-7	44	47.8821
		(2.3)	[0.3312]		(5.6)	[0.3963]
12	36	8	47.4384	16	48	41.7552
		(2.2)	[0.4138]		(6.3)	[0.6506]

Note. The table reports the pricing of excess returns of stocks and bonds, allowing for no intercept. Standard errors are reported in parentheses and p-values in brackets. Estimation is done using EL and GMM.

Figure 3.10: Cross-sectional spread of exposure to slow consumption adjustment risk



Note. Panels present the spread of normalised betas for the various sets of assets and horizon S (0-15): (a) 9 bonds and 6 Fama-French portfolios, (b) 9 bonds and 25 Fama-French portfolios, (c) 9 bonds, 12 Industry and 6 Fama-French portfolios, (d) 9 bonds, 12 Industry and 25 Fama-French portfolios. All the parameters were estimated by Empirical Likelihood.

$10-12$: in fact, based on Empirical Likelihood Estimation, the model is no longer rejected in any of the cross-sections. Combined with the improved values of the power parameter

(ϕ), the accuracy of its estimation (lower standard errors), and a substantial increase in the cross-sectional quality of fit, measured by the R^2, Table 3.4 presents compelling evidence in favour of the slow consumption adjustment risk being an important driver for the cross-sections of both stocks and bonds. Appendix ?? provides similar empirical evidence for the alternative model specifications that also include a common or asset class-specific intercept as a proxy for model misspecification.

But why does the slow consumption adjustment risk provide a better fit for the cross-sectional spread in expected returns? The empirical evidence, presented in the previous section, suggests that both stocks and bonds tend to co-vary more with the consumption growth over the next few periods (captured by the common unobservable factor and the loadings on it). However, not only the SCA risk measure increases the average asset exposure to consumption growth, it also improves the *spread* of the latter. While the standard one-period consumption growth does not perform well in either dimensions, leading to the equity premium puzzle and a relatively poor cross-sectional fit, the SCA factor seems to achieve both objectives: it increases the amount of measured risk as well as its cross-sectional dispersion.

Figure 3.10 displays the dispersion of the model-implied scaled betas,[1] associated with the consumption growth over different horizon values and for different cross-sections of assets. As we move away from the standard case of $S = 0$, two observations immediately arise. First, there is a substantial improvement in the average asset exposure to consumption growth, which leads to lower and more accurate estimates of the risk aversion. However, it is the increase in the spread of betas, with a particular contribution from the stocks, which is most striking. The 'fanning out' effect, observed for the higher values of the consumption horizon S, further supports the hypothesis that the fundamental source of risk in the asset returns is related to the aggregate consumption growth, and should take into account its slow speed of adjustments to the common shocks.

Finally, the fact that there is a significant correlation between asset returns and consumption growth over the several periods (both in terms of its level and spread), also serves as an additional robustness check against a potential problem of *spurious factors* type (Kan and Zhang (1999b)), i.e. factors that are only weakly related to the asset returns and thus only *appear* to be driving the cross-section of asset returns.

[1] We define betas as the ratio between the asset covariance with the model-implied scaled SDF and its variance.

3.6 Conclusion

This paper provides empirical evidence that the slow consumption adjustment risk is an important driver for both stock and bond returns. A flexible parametric model with common factors driving asset dynamics and consumption identifies a slow varying component of consumption that responds to financial shocks. Both stocks and bonds load significantly load on SCA risk factor, generating a sizeable risk premium and a dispersion in returns, consistent with the size and value anomalies, as well as the positive slope of the yield curve. As a result, our model explains between 36% and 95% of the time series variation in returns and between 57% and 90% of the joint cross-sectional variation in stocks and bonds.

Moreover, we find that slow consumption adjustment innovations drive more than a quarter of the time series variation of consumption growth, indicating that financial market related shocks are first order drivers of consumption risk.

While generally consistent with the consumption dynamics postulated in the long run risk framework, these empirical findings nevertheless pose several important questions. Can the results be applied to other asset classes, such as currencies or commodities? What is the nature of the unspanned factor, driving most of the time series variation in bonds?

Appendix

3.A State Space Estimation and Generalisations

Let $\Pi' := [\mu, H]\,, x_t' := [1, z_t']$. Under a (diffuse) Jeffreys' prior the likelihood of the data in equation (3.20) implies the posterior distribution

$$\Pi' \big| \, \Sigma, \{z_t\}_{t=1}^{T}, \{y_t\}_{t=1}^{T} \sim \mathcal{N}\left(\hat{\Pi}'_{OLS}; \Sigma \otimes \left(x'x\right)^{-1}\right)$$

where x contains the stacked regressors, and the posterior distribution of each element on the main diagonal of Σ is given by

$$\sigma_j^2 \big| \, \{z_t\}_{t=1} \sim \text{Inv-}\Gamma\left(\left(T - m_j - 1\right)/2, T\hat{\sigma}^2_{j,OLS}/2\right)$$

where m_j is the number of estimated coefficients in the j-th equation. Moreover, F and Ψ have a Dirac posterior distribution at the points defined in equation (3.17). Therefore, the missing part necessary for taking draws via MCMC using a Gibbs sampler, is the conditional distributions of z_t. Since

$$\begin{array}{c} y_t \\ z_t \end{array} \Bigg| \, \mathcal{I}_{t-1}, H, \Psi, \Sigma \sim \mathcal{N}\left(\left[\begin{array}{c} \mu \\ F z_{t-1} \end{array}\right]; \left[\begin{array}{cc} \Omega & H \\ H' & \Psi \end{array}\right]\right),$$

where $\Omega := Var_{t-1}\left(y_t\right) = H\Psi H' + \Sigma$, this can be constructed, and values can be drawn, using a standard Kalman filter and smoother approach. Let

$$z_{t|\tau} := E\left[z_t | y^\tau, H, \Psi, \Sigma\right]; \quad V_{t|\tau} := Var\left(z_t | H, \Psi, \Sigma\right).$$

where y^τ denotes the history of y_t until τ. Then, given $z_{0|0}$ and $V_{0|0}$, the Kalman filer delivers:

$$\begin{aligned} z_{t|t-1} = F z'_{t-1|t-1}; \;\; V_{t|t-1} = F V_{t-1|t-1} F' + \Psi; \;\; K_t = V_{t|t-1} H' \left(H V_{t|t-1} H' + \Sigma\right)^{-1} \\ z_{t|t} = z_{t|t-1} + K_t \left(y_t - \mu - H z_{t|t-1}\right); \;\; V_{t|t} = V_{t|t-1} - K_t H V_{t|t-1}. \end{aligned}$$

The last elements of the recursion, $z_{T|T}$ and $V_{T|T}$, are the mean and variance of the normal distribution used to draw z_T. The draw of z_T and the output of the filter can then be used for the first step of the backward recursion, which delivers the $z_{T-1|T}$ and $V_{T-1|T}$ values necessary to make a draw for z_{T-1} from a gaussian distribution. The backward recursion can be continued until time zero, drawing each value of z_t in the process, with the following updating formulae for a generic time t recursion:

$$z_{t|t+1} = z_{t|t} + V_{t|t} F' V_{t+1|t}^{-1} \left(z_{t+1} - F z_{t|t} \right); \;\; V_{t|t+1} = V_{t|t} - V_{t|t} F' V_{t+1|t}^{-1} F V_{t|t}.$$

Hence parameters and states can be drawn via Gibbs sampler using the following algorithm:

1. Take a guess $\tilde{\Pi}'$ and $\tilde{\Sigma}^{-1}$ (e.g. freq. estimate), and use it to construct initial draws for μ and H. Using also F and Ψ, draw the z_t history using the Kalman recursion above with (Kalman step)
$$z_t \sim \mathcal{N}\left(z_{t|t+1}; z_{t|t+1} \right).$$
2. Conditioning on $\{z_t\}_{t=1}^{T}$ (drawn at the previous step) and $\{y_t\}_{t=1}^{T}$ run OLS imposing the zero restrictions and get $\hat{\Pi}'_{OLS}$ and $\hat{\Sigma}_{OLS}$, and draw $\tilde{\Pi}'$ and $\tilde{\Sigma}^{-1}$ from the N-i-Γ. Use the draws as the initial guess for the previous point of the algorithm (N-i-Γ step), and repeat.

Computing posterior confidence intervals for the cross-sectional performance of the model, conditional on the data, is relatively simple since, conditional on a draw of the time series parameters, estimates of the risk premia (λ's in equations (3.21) and (3.22)) are just a mapping obtainable via the linear projection of average returns on the asset loadings in H. Hence, to compute posterior confidence intervals for the cross-sectional analysis, we repeat the cross-sectional estimation for each posterior draw of the time series parameters, and report the posterior distribution of the cross-sectional statistics across these draws.

3.B The Moving Average Representation of The Long Run Risk Process

We we assume the same data generating process as in Bansal and Yaron (2004), with the only exception that we introduce a square-root process for the variance, as in Hansen, Heaton, Lee, and Roussanov (2007), that is:

$$\Delta c_{t,t+1} = \mu + x_t + \sigma_t \eta_{t+1}; \;\; x_{t+1} = \rho x_t + \phi_e \sigma_t e_{t+1}; \;\; \sigma_{t+1}^2 = \sigma^2 (1 - \nu_1) + \nu_1 \sigma_t^2 + \sigma_w \sigma_t w_{t+1},$$

where η_t, e_t, w_t, $\sim$ iid $\mathcal{N}(0,1)$. The calibrated *monthly* parameter values are: $\mu = 0.0015$, $\rho = 0.979$, $\phi_e = 0.044$, $\sigma = 0.0078$, $\nu_1 = 0.987$, $\sigma_w = 0.00029487$. To extract the quarterly frequency moving average representation of the process, we proceed in two steps. First, we simulate a long sample (five million observations) from the above system treating the given parameter values as the truth. Second, we aggregate the simulated data into quarterly observation and we use them to estimate, via MLE, the moving average representation of consumption growth in equation (3.8).

3.C Additional Empirical Results

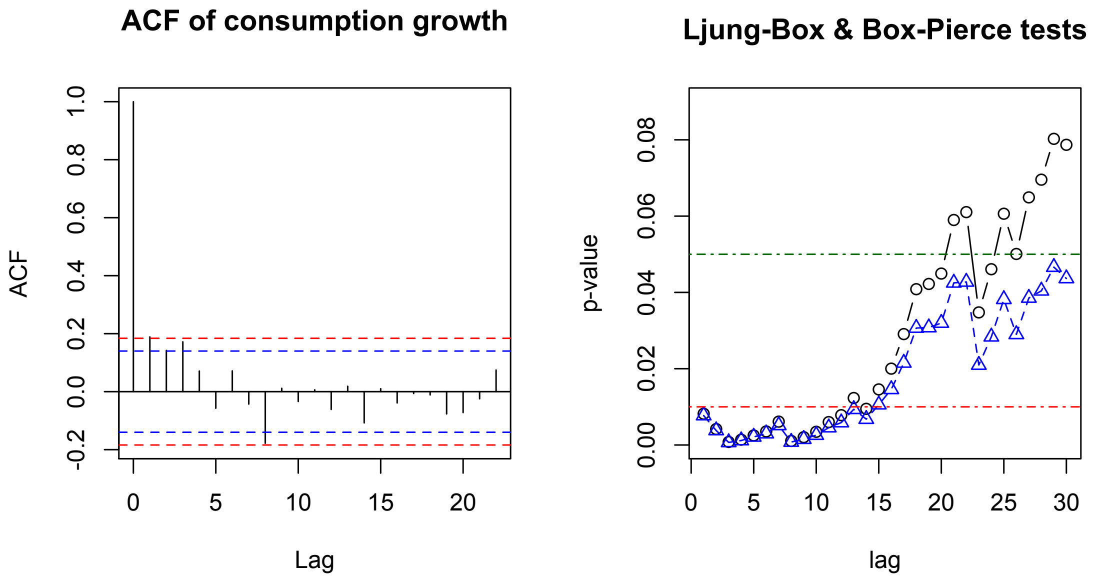


Figure 3.C.1: Autocorrelation structure of consumption growth.

Note. Left panel: autocorrelation function of consumption growth ($\Delta c_{t,t+1+S}$) with 95% and 99% confidence bands. Right panel: $p-$values of Ljung and Box (1978) (triangles) and Box and Pierce (1970) (circles) tests.

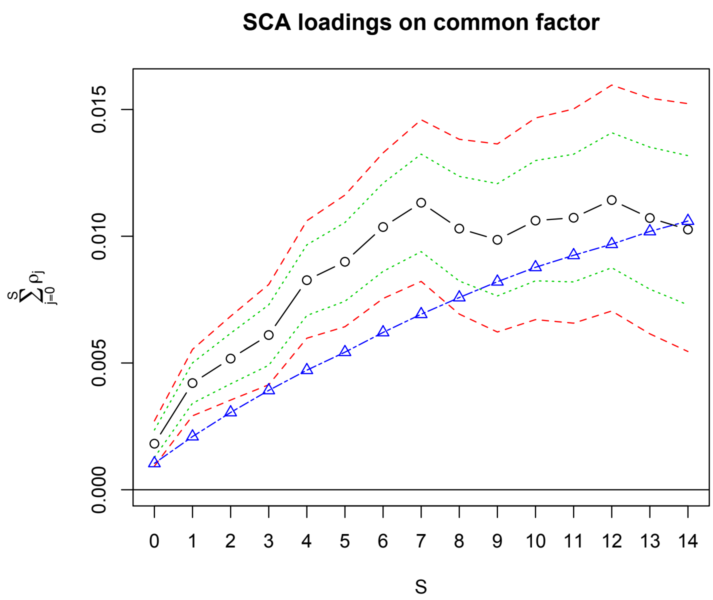


Figure 3.C.2: Slow Consumption Adjustment response to the common factor (f_t) shock.

Note. Posterior means (continuous line with circles) and centred posterior 90% (dashed line) and 68% (dotted line) coverage regions. Triangles denote Bansal and Yaron (2004) implied values.

Table 3.C.1: Expected Excess Returns and Consumption Risk, 1967:Q3-2013:Q4

	Empirical Likelihood					Generalised Method of Moments				
Horizon S (Quarters)	$R^2_{adj}(\%)$ (1)	α_b (2)	α_s (3)	ϕ (4)	J-test (5)	$R^2_{adj}(\%)$ (6)	α_b (7)	α_s (8)	ϕ (9)	J-test (10)
Panel A: 9 Bonds and Fama-French 6 portfolios										
0	-6	0.0001	0.0162	-74	59.7497	71	0.0003	0.0137	33	38.3181
		(0.0002)	(0.0045)	(21.2)	[0.0003]		(0.0003)	(0.0069)	(42.1)	[0.0001]
10	54	0.0004	0.0105	22	24.2148	29	0.0006	0.0132	28	17.3421
		(0.0003)	(0.0046)	(3.9)	[0.6184]		(0.0004)	(0.0046)	(6.9)	[0.1372]
11	51	0.0005	0.0099	24	24.2189	44	0.0008	0.0131	30	17.6300
		(0.0003)	(0.0047)	(3.7)	[0.6181]		(0.0003)	(0.0048)	(8.3)	[0.1274]
12	52	0.0005	0.0093	22	22.1532	53	0.0009	0.0136	32	18.6997
		(0.0003)	(0.0049)	(3.5)	[0.7295]		(0.0003)	(0.0050)	(9.0)	[0.0960]
Panel B: 9 Bonds and Fama-French 25 portfolios										
0	50	-0.0006	0.0130	60	62.3266	61	0.0011	0.0125	50	226.2077
		(0.0002)	(0.0045)	(21.2)	[0.0007]		(0.0001)	(0.0038)	(15.3)	[0.0000]
10	72	-0.0002	0.0104	19	23.1802	26	0.0019	0.0063	37	44.4437
		(0.0003)	(0.0038)	(3.9)	[0.8425]		(0.0003)	(0.0018)	(5.9)	[0.0558]
11	79	-0.0002	0.0096	18	20.8589	56	0.0020	0.0052	39	33.7601
		(0.0002)	(0.0039)	(3.9)	[0.9156]		(0.0003)	(0.0020)	(6.5)	[0.3355]
12	78	-0.0001	0.0096	17	20.4496	64	0.0018	0.0065	42	28.8556
		(0.0002)	(0.0040)	(3.7)	[0.9257]		(0.0002)	(0.0015)	(7.2)	[0.5768]
Panel C: 9 Bonds, Fama-French 6, and Industry 12 portfolios										
0	64	0.0000	0.0119	-14	59.4323	-33	0.0006	0.0239	31	124.6547
		(0.0002)	(0.0041)	(22.3)	[0.0001]		(0.0001)	(0.0027)	(18.4)	[0.0000]
10	72	0.0003	0.0131	14	21.9269	-77	0.0016	0.0140	32	44.9201
		(0.0003)	(0.0039)	(4.3)	[0.5836]		(0.0003)	(0.0024)	(6.3)	[0.0060]
11	70	0.0004	0.0119	11	24.8752	-53	0.0018	0.0140	34	37.2377
		(0.0002)	(0.0040)	(3.9)	[0.4126]		(0.0003)	(0.0023)	(7.0)	[0.0414]
12	72	0.0004	0.0115	10	24.4976	2	0.0019	0.0107	38	29.5539
		(0.0002)	(0.0041)	(3.7)	[0.4335]		(0.0003)	(0.0024)	(7.7)	[0.2000]
Panel D: 9 Bonds, Fama-French 25, and Industry 12 portfolios										
0	54	0.0005	0.0124	23	78.2258	36	0.0007	0.0146	58	269.4971
		(0.0002)	(0.0036)	(16.7)	[0.0008]		(0.0002)	(0.0027)	(13.4)	[0.0000]
10	61	-0.0002	0.0114	6	55.7091	-29	0.0018	0.0093	36	71.4739
		(0.0003)	(0.0034)	(2.5)	[0.0926]		(0.0003)	(0.0013)	(4.7)	[0.0041]
11	62	-0.0002	0.0112	6	53.6016	8	0.0020	0.0090	37	60.1299
		(0.0002)	(0.0034)	(2.4)	[0.1289]		(0.0003)	(0.0013)	(4.8)	[0.0430]
12	62	-0.0002	0.0111	6	51.8898	25	0.0019	0.0082	42	47.2360
		(0.0002)	(0.0034)	(2.2)	[0.1659]		(0.0003)	(0.0012)	(5.4)	[0.3036]

Note. The table reports the pricing of excess returns of stocks and bonds, allowing for separate asset class-specific intercepts. Standard errors are reported in parentheses and p-values in brackets. Estimation is done using EL and GMM.

Table 3.C.2: Expected Excess Returns and Consumption Risk, 1967:Q3-2013:Q4

	Empirical Likelihood				Generalised Method of Moments			
Horizon S	$R^2_{adj}(\%)$	α	ϕ	ELR-test	$R^2_{adj}(\%)$	α	ϕ	J-test
(Quarters)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: B Bonds and Fama-French 6 portfolios								
0	-30	0.0002	-16	30.1955	73	0.0007	73	35.0646
		(0.0002)	(25.0)	[0.0044]		(0.0003)	(27.1)	[0.0008]
10	94	0.0008	23	11.5946	91	0.0009	29	24.9738
		(0.0003)	(6.0)	[0.5611]		(0.0004)	(6.8)	[0.0233]
11	95	0.0006	24	10.4758	94	0.0011	32	24.4029
		(0.0003)	(6.7)	[0.6546]		(0.0003)	(8.5)	[0.0276]
12	92	0.0005	23	11.1154	96	0.0012	34	25.2110
		(0.0003)	(6.5)	[0.6011]		(0.0003)	(9.3)	[0.0217]
Panel A: B Bonds and Fama-French 25 portfolios								
0	54	-0.0004	52	78.6597	60	0.0018	61	321.3738
		(0.0002)	(18.2)	[0.0000]		(0.0001)	(15.3)	[0.0000]
10	74	-0.0001	19	68.5008	38	0.0025	38	48.0606
		(0.0002)	(3.7)	[0.0002]		(0.0003)	(5.7)	[0.0340]
11	76	0.0000	20	67.9188	62	0.0024	40	35.1659
		(0.0002)	(3.7)	[0.0002]		(0.0003)	(6.0)	[0.3205]
12	70	0.0000	18	71.0791	67	0.0029	44	30.5687
		(0.0002)	(3.4)	[0.0001]		(0.0002)	(7.4)	[0.5390]
Panel C: 9 Bonds, Fama-French 6, and Industry 12 portfolios								
0	-6	0.0001	-6	63.2328	61	0.0017	55	273.0204
		(0.0002)	(21.9)	[0.0002]		(0.0002)	(15.2)	[0.0000]
10	56	0.0009	14	56.6896	-24	0.0037	35	51.9830
		(0.0003)	(4.0)	[0.0003]		(0.0003)	(6.4)	[0.0012]
11	51	0.0009	12	58.4329	-9	0.0042	37	38.4378
		(0.0002)	(3.7)	[0.0002]		(0.0003)	(6.9)	[0.0419]
12	52	0.0009	12	58.0225	10	0.0039	41	29.1776
		(0.0002)	(3.6)	[0.0002]		(0.0003)	(6.5)	[0.2566]
Panel D: 9 Bonds, Fama-French 25, and Industry 12 portfolios								
0	26	-0.0003	22	146.685	54	0.0016	69	356.9325
		(0.0002)	(15.2)	[0.0000]		(0.0002)	(13.5)	[0.0000]
10	38	-0.0002	8	141.4802	-25	0.0039	39	77.4115
		(0.0002)	(2.5)	[0.0000]		(0.0003)	(4.8)	[0.0014]
11	38	-0.0002	8	140.6384	16	0.0041	39	66.0979
		(0.0002)	(2.3)	[0.0000]		(0.0003)	(4.9)	[0.0172]
12	37	-0.0002	8	140.8904	29	0.0041	43	51.9741
		(0.0002)	(2.2)	[0.0000]		(0.0003)	(5.6)	[0.1912]

Note. The table reports the pricing of excess returns of stocks and bonds, allowing for a common intercept. Standard errors are reported in parentheses and p-values in brackets. Estimation is done using EL and GMM.

Table 3.C.3: Expected Excess Returns and Consumption Risk, 1967:Q3-2013:Q4

	Empirical Likelihood				Generalised Method of Moments			
Horizon S	$R^2_{adj}(\%)$	α	ϕ	ELR-test	$R^2_{adj}(\%)$	α	ϕ	J-test
(Quarters)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	-30	0.0002	-16	30.1955	73	0.0007	73	35.0646
		(0.0002)	(25.0059)	[0.0044]		(0.0003)	(27.1)	[0.0008]
1	50	0.0005	55	19.8352	64	0.0006	50	26.7987
		(0.0003)	(16.5936)	[0.0994]		(0.0003)	(16.4)	[0.0133]
2	3	0.0008	50	15.9515	39	0.0008	45	20.5230
		(0.0004)	(11.4430)	[0.2517]		(0.0004)	(11.3)	[0.0829]
3	27	0.0007	45	14.3198	55	0.0007	40	20.5313
		(0.0004)	(9.3960)	[0.3517]		(0.0004)	(9.4)	[0.0827]
4	-33	0.0004	40	12.6842	16	0.0004	36	18.8278
		(0.0003)	(7.8412)	[0.4725]		(0.0003)	(7.7)	[0.1285]
5	58	0.0004	29	11.8102	42	0.0005	31	19.6120
		(0.0003)	(6.5887)	[0.5433]		(0.0003)	(6.7)	[0.1053]
6	67	0.0004	27	12.0794	53	0.0005	29	20.0162
		(0.0003)	(6.0256)	[0.5211]		(0.0003)	(6.2)	[0.0948]
7	61	0.0002	26	11.9012	43	0.0004	28	22.5791
		(0.0003)	(5.8619)	[0.5358]		(0.0003)	(6.0)	[0.0470]
8	89	0.0003	25	12.3113	74	0.0006	29	23.9049
		(0.0003)	(5.8866)	[0.5023]		(0.0003)	(6.3)	[0.0320]
9	95	0.0003	25	13.0312	92	0.0009	29	24.9160
		(0.0003)	(5.9862)	[0.4454]		(0.0003)	(6.4)	[0.0237]
10	94	0.0008	23	11.5946	91	0.0009	29	24.9738
		(0.0003)	(5.9595)	[0.5611]		(0.0004)	(6.8)	[0.0233]
11	95	0.0006	24	10.4758	94	0.0011	32	24.4029
		(0.0003)	(6.7275)	[0.6546]		(0.0003)	(8.5)	[0.0276]
12	92	0.0005	23	11.1154	96	0.0012	34	25.2110
		(0.0003)	(6.5436)	[0.6011]		(0.0003)	(9.3)	[0.0217]
13	86	0.0004	22	11.8978	96	0.0012	35	26.5862
		(0.0003)	(6.3313)	[0.5360]		(0.0003)	(9.6)	[0.0142]
14	85	0.0004	23	11.7044	97	0.0013	42	18.5716
		(0.0003)	(6.5983)	[0.5520]		(0.0005)	(13.2)	[0.1370]
15	79	0.0005	21	13.4734	96	0.0021	43	32.4073
		(0.0003)	(6.1575)	[0.4120]		(0.0004)	(12.7)	[0.0021]

Note. The table reports the pricing of 9 excess bond holding returns and 6 Fama-French portfolios, sorted on size and book-to-market. We report the results for various values of the horizon parameters S and allow for a common intercept. Standard errors are reported in parentheses and p-values in brackets. Estimation is done using EL and GMM.

References

Abel, A. B. (1990): "Asset Prices Under Habit Formation and Catching Up with the Joneses," *American Economic Review*, 80, 38–42. 123

Adrian, T., R. K. Crump, and E. Moench (2013): "Pricing the Term Structure with Linear Regressions," *Journal of Financial Economics*, 110(1), 103–27. 1, 89, 90, 92, 93, 95, 96, 100, 101, 103, 112, 116, 117, 118

Anatolyev, S. (2005): "GMM, GEL, Serial Correlation, and Asymptotic Bias," *Econometrica*, 73, 983–1002. 131

Andrews, D., and P. Guggenberger (2009): "Incorrect Asymptotic Size of Subsampling Procedures Based on Post-consistent Model Selection Estimators," *Journal of Econometrics*, 152(1), 19–27. 6

Andrews, D. W. K. (1991): "Asmptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59(2), 307–345. 45, 48, 54

——— (1994): "Empirical Process Methods in Econometrics," *Econometrica*, 4, 2247–2294. 85, 87

Ang, A., and M. Piazzesi (2002): "A No-arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745–87. 89

Ang, A., M. Piazzesi, and M. Wei (2006): "What Does the Yield Curve Tell Us About GDP Growth?," 131, 359–403. 88, 90

Ang, A., and M. Ulrich (2012): "Nominal Bonds, Real Bonds, and Equity," Manuscript, Columbia University. 122

Asness, C., A. Frazzini, and L. H. Pedersen (2014): "Quality Minus Junk," Working Paper. 3, 43

Ball, R., and P. Brown (1968): "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting Research*, 6, 159–78. 4

BANDI, F., AND A. TAMONI (2015): "Business-Cycle Consumption Risk and Asset Prices," available at SSRN: http://ssrn.com/abstract=2337973. 7, 134

BANSAL, R., R. F. DITTMAR, AND C. T. LUNDBLAD (2005): "Consumption, Dividends, and the Cross Section of Equity Returns," *The Journal of Finance*, 60(4), pp. 1639–1672. 121

BANSAL, R., D. KIKU, AND A. YARON (2012): "Risks For the Long Run: Estimation with Time Aggregation," NBER Working Papers 18305, National Bureau of Economic Research, Inc. 121

BANSAL, R., AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509. 121, 123, 140, 141, 153, 155

BASAK, S., AND H. YAN (2010): "Equilibrium Asset Prices and Investor Behaviour in the Presence of Money Illusion," *Review of Economic Studies*, 77(3), 914–936. 123

BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxofrd University Press, Oxford. 129

BEKAERT, G., E. ENGSTROM, AND S. R. GRENADIER (2010): "Stock and bond returns with Moody Investors," *Journal of Empirical Finance*, 17(5), 867–894. 122

BEKAERT, G., AND S. R. GRENADIER (1999): "Stock and Bond Pricing in an Affine Economy," NBER Working Papers 7346, National Bureau of Economic Research, Inc. 122

BERK, R., L. BROWN, A. BUJA, K. ZHANG, AND L. ZHAO (2013): "Valid Post-selection Inference," *Annals of Statistics*, 41(2), 802–37. 3

BERNAND, V., AND J. THOMAS (1990): "Evidence that Stock Prices do not Fully Reflect the Implications of Current Earnings for Future Earnings," *Journal of Accounting and Economics*, 13, 305–40. 4

BLANCHARD, O. J., AND D. QUAH (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79, 655–73. 141

BOONS, M., AND A. TAMONI (2014): "Horizon-Specific Macroeconomic Risks and the Cross-Section of Expected Returns," Working Paper. 7

BOX, G. E. P., AND D. A. PIERCE (1970): "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65(332), pp. 1509–1526. 155

Brave, S. (2009): "The Chicago Fed National Activity Index and Business Cycles," *Chicago Fed Letter*, (268). 102

Breeden, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–96. 50

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989): "Empirical Test of the Consumption-Oriented CAPM," *The Journal of Finance*, 44(2), 231–262. 119

Breiman, L. (1996): "Heuristics of Instability and Stabilization in Model Selection," *Annals of Statistics*, 24(6), 97–126. 3

Brennan, M. J., A. W. Wang, and Y. Xia (2004): "Estimation and Test of a Simple Model of Intertemporal Capital Asset Pricing," *The Journal of Finance*, 59(4), 1743–1776. 122

Burnside, C. (2010): "Identification and Inference in Linear Stochastic Discount Factor Models," NBER Working Paper 16634. 2, 4, 7, 19

Campbell, J. Y. (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 298–345. 119

——— (2003): "Consumption-Based Asset Pricing," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, chap. 13. North-Holland, Amsterdam. 51

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–51. 123

Campbell, J. Y., and R. J. Shiller (1991): "Yield Spreads and Interest Rate Movements: a Birds Eye View," *Review of Economic Studies*, 58, 495–514. 88

Caner, M. (2009): "LASSO Type GMM Estimator," *Econometric Theory*, 25, 1–23. 6, 24, 85

Caner, M., and M. Fan (2014): "Hybrid GEL Estimators: Instrument Selection with Adaptive Lasso," Working Paper. 6, 85

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82. 43, 49

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–34. 131

Chan, L., N. Jegadeesh, and J. Lakonishok (1996): "Momentum Strategies," *Journal of Finance*, 51, 1681–713. 4

Chatterjee, A., and S. N. Lahiri (2010): "Asymptotic Properties of the Residual Bootstrap for Lasso Estimators," *Proceedings of the American Mathematical Society*, 138, 4497–4509. 21, 22

——— (2011): "Bootstrapping Lasso Estimators," *Journal of the American Statistical Association*, 106, 608–25. 6, 21

——— (2013): "Rates of Convergence of the Adaptive Lasso Estimators to the Oracle Distribution and Higher Order Refinements by the Bootstrap," *Annals of Statistics*, 41(3), 1232–59. 21

Chernov, M., and P. Mueller (2012): "The Term Structure of Inflation Expectations," *Journal of Financial Economics*, 106, 367–94. 89, 121

Chetty, R., and A. Szeidl (2015): "Consumption Commitments and Habit Formation," Working Paper. 123

Cochrane, J., and M. Piazzesi (2005): "Bond Risk Premia," *American Economic Review*, 94, 138–60. 88, 122

Constantinides, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98(2), 519–43. 123

Cooper, M., H. Gulen, and M. Schill (2008): "Asset Growth and the Cross-section of Stock Returns," *Journal of Finance*, 63, 1609–52. 4

Cox, J., J. Ingersoll, and S. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407. 88

Cragg, J. G., and S. G. Donald (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–50. 2

Csiszar, I. (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," *Annals of Probability*, 3, 146–158. 131

Daniel, K. D., and D. Marshall (1997): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle at Long Horizons," *Macroeconomic Dynamics*, 1(2), 452–84. 52, 121

Duffie, D., and R. Kan (1996a): "A Yield-Factor Model Of Interest Rates," *Mathematical Finance*, 6(4), 379–406. 122

——— (1996b): "Simulated Moments Estimation of Markov Models of Asset Prices," *Mathematical Finance*, 6, 379–406. 88

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–968. 123

Fairfield, P., S. Whisenant, and T. Yohn (2003): "Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing," *The Accounting Review*, 78, 353–71. 4

Fama, E. F., and R. R. Bliss (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–92. 88

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465. 1, 43, 122, 133, 135, 138, 144

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *The Journal of Financial Economics*, 33, 3–56. 1, 3, 122

——— (2006): "Profitability, Investment, and Average returns," *Journal of Financial Economics*, 82, 491–518. 4

——— (2015): "Dissecting Anomalies with a Five-Factor Model," *Review of Financial Studies*, forthcoming. 4

Fama, E. F., and J. MacBeth (1973): "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636. 129, 144

Favero, C. A., L. Niu, and L. Sala (2012): "Term Structure Forecasting: No-Arbitrage Restrictions versus Large Information Set," *Journal of Forecasting*, 31, 124–56. 89

Fisher, J. D. M. (2000): "Forecasting Inflation with a Lot of Data," No. 151. Federal Reserve Bank of Chicago. 102

Flavin, M. (1981): "The Adjustment of Consumption to Changing Expectations about Future Income," *Journal of Political Economy*, 89, 974–1009. 125

Friedman, J., T. Hastie, H. Hoffling, and R. Tibshirani (2007): "Pathwise Coordinate Descent," *The Annals of Applied Statistics*, 1(2), 302–32. 15

Friedman, J., T. Hastie, and R. Tibshirani (2010): "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33, 1–22. 37

Gabaix, X., and D. Laibson (2001): "The 6D bias and the equity premium puzzle," in *N.B.E.R. Macroeconomics Annual 2001*, ed. by B. Bernanke, and K. Rogoff, pp. 257–311. Cambridge: MIT Press. 125

Ghosh, A., C. Julliard, and A. Taylor (2013): "What is the Consumption-CAPM missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models," London School of Economics Manuscript. 120, 130, 131

Gospodinov, N., R. M. Kan, and C. Robotti (2014a): "Misspecification-Robust Inference in Linear Asset Pricing Models with Irrelevant Risk Factors," *Review of Financial Studies*, 27, 2139–2170. 2, 7, 11, 19, 20, 30, 37, 38, 39, 40, 42, 45, 48, 54

——— (2014b): "Spurious Inference in Unidentified Asset-Pricing Models," Working Paper. 4, 12, 33

Guggenberger, P. (2010): "The Impact of a Hausman Pretest on the Size of a Hypothesis Test: the Panel Data Case," *Journal of Econometrics*, 156(2), 337–43. 3

Gurkaynak, Refet, B. S., and J. H. Wright (2007): "The US treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54(8), 2291–2304. 102, 133

Hall, R. E., and F. S. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50, 461–481. 125

Hamilton, J. D., and D. H. Kim (2002): "A Reexamination of the Predictability of Economic Activity Using the Yield Spread," *Journal of Money, Credit, and Banking*, 34, 340–60. 88

Hamilton, J. D., and C. Wu (2012): "Identification and Estimation of Affine Term Structure Models," (168), 315–31. 104

Hansen, L. P. (1982): "Large Sample Properties of Method of Moments Estimators," *Econometrica*, 50, 1029–1054. 132

Hansen, L. P., J. Heaton, J. Lee, and N. Roussanov (2007): "Intertemporal Substitution and Risk Aversion," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 61. Elsevier. 121, 153

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), 262–80. 131

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116(2), 260–302. 121

Hansen, L. P., and R. Jagannathan (1997): "Asssessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52, 557–590. 12

Hansen, L. P., and T. J. Sargent (2010): "Fragile Beliefs and the Price of Uncertainty," *Quantitative Economics*, 1(1), 129–162. 123

Hansen, L. P., and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–86. 51

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–68. 51, 119

Harvey, C., Y. Liu, and H. Zhu (2013): "...and the cross-section of expected returns," Working Paper, Duke University. 1

Hastie, T., R. Tibshirani, and J. Friedman (2011): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition)*, Springer Series in Statistics. Springer-Verlag, New York. 36

Haugen, R., and N. Baker (1996): "Commonality in the Determinants of Expected Stock Returns," *Journal of Financial Economics*, 41, 401–439. 4

Hou, K., C. Xue, and L. Zhang (2014): "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, p. forthcoming. 4, 44, 49

Jagannathan, R., and Y. Wang (2007): "Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns," *The Journal of Finance*, 62(4), pp. 1623–1661. 1, 52, 121, 125

Jagannathan, R., and Z. Wang (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *The Journal of Finance*, 51, 3–53. 1, 44, 52

——— (1998): "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression," *The Journal of Finance*, 53, 1285–1309. 2, 17, 20

Joslin, S., M. Priebsch, and K. J. Singleton (2012): "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Unpublished working paper*. 89

Joslin, S., K. J. Singleton, and H. Zhu (2011): "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, 24, 926–70. 104

Julliard, C., and A. Ghosh (2012): "Can Rare Events Explain the Equity Premium Puzzle?," *Review of Financial Studies*, 25(10), 3037–3076. 119

Kan, R. M., and C. Zhang (1999a): "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54, 103–27. 11, 20

——— (1999b): "Two-pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 204–35. 2, 5, 12, 142, 150

Kitamura, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, 69, 1661–1672. 131

——— (2006): "Empirical Likelihood Methods in Econometrics: Theory and Practice," Cowles Foundation Discussion Papers 1569, Cowles Foundation, Yale University. 130

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative To Generalized Method Of Moments Estimation," *Econometrica*, 65(4), 861–874. 131

Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149, 149–73. 2, 5, 12, 16, 17, 20

Kleibergen, F., and R. Paap (2006): "Generalised Reduced Rank Test Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126. 2, 7

Kleibergen, F., and Z. Zhan (2013): "Unexplained Factors and their Effects on Second Pass R-squareds and t-tests," Working Paper. 2, 4, 12, 28, 30, 33

Knight, K., and W. Fu (2000): "Asymptotics for Lasso-type Estimators," *The Annals of Statistics*, 28(5), 1356–78. 80, 84, 85, 87

Koijen, R. S., H. Lustig, and S. V. Nieuwerburgh (2010): "The Cross-Section and Time-Series of Stock and Bond Returns," NBER Working Papers 15688, National Bureau of Economic Research, Inc. 122

Kroencke, T. A. (2013): "Asset Pricing Without Garbage," available at SSRN: http://ssrn.com/abstract=2327055. 52, 125

Lettau, M., and S. Ludvigson (2001a): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56(3), 815–49. 51

——— (2001b): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1286. 1, 6, 44, 45, 52

Lettau, M., and J. A. Wachter (2011): "The term structures of equity and interest rates," *Journal of Financial Economics*, 101(1), 90–113. 122

Lewellen, J., S. Nagel, and J. Shanken (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96(2), 175–194. 6, 28, 136, 142

Li, D., and L. Zhang (2010): "Does q-theory with Investment Frictions Explain Anomalies in the Cross-section of Returns?," *Journal of Financial Economics*, 210, 297–314. 44

Liao, Z. (2013): "Adaptive GMM Shrinkage Estimation with Consistent Moment Selection," *Econometric Theory*, 29, 1–48. 6, 13, 24, 85

LINTNER, J. (1965): "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20, 587–615. 1

LITTERMAN, R., AND J. A. SCHEINKMAN (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, pp. 54–61. 88

LIU, L. X., T. M. WHITED, AND L. ZHANG (2009): "Investment-based Expected Stock Returns," *Journal of Political Economy*, 112, 1105–39. 44

LJUNG, G. M., AND G. E. P. BOX (1978): "On a measure of lack of fit in time series models," *Biometrika*, 65(2), 297–303. 155

LUCAS, JR., R. E. (1976): "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets*, vol. 1 of *Carnegie Rochester Conference Series on Public Policy*, pp. 19–46. North-Holland, Amsterdam. 50

LUDVIGSON, S. (2013): "Advances in Consumption-Based Asset Pricing: Empirical Tests," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, vol. 2, chap. 12. North-Holland, Amsterdam. 51

LUDVIGSON, S., AND S. NG (2009): "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22, 5027–67. 89

LUSTIG, H., N. ROUSSANOV, AND A. VERDELHAN (2011): "Common Risk Factors in Currency Markets," *Review of Financial Studies*, (11), 3731–77. 1

LUSTIG, H., AND A. VERDELHAN (2007): "The Cross-Section of Foreign Currency Risk Premia and US Consumption Growth Risk," *American Economic Review*, (1), 89–117. 1

LUSTIG, H. N., AND S. G. V. NIEUWERBURGH (2005): "Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective," *The Journal of Finance*, 60(3), pp. 1167–1219. 123

LYNCH, A. W. (1996): "Decision Frequency and Synchronization Across Agents: Implications for Aggregate Consumption and Equity Return," *Journal of Finance*, 51(4), 1479–97. 52, 125

MALLOY, C. J., T. J. MOSKOWITZ, AND A. VISSING-JORGENSEN (2009): "Long-Run Stockholder Consumption Risk and Asset Returns," *The Journal of Finance*, 64(6), 2427–2479. 121

MAMAYSKY, H. (2002): "A Model For Pricing Stocks and Bonds," Yale School of Management Working Papers ysm279, Yale School of Management. 122

Mankiw, N. G., and M. D. Shapiro (1986): "Risk and Return: Consumption Beta Versus Market Beta," *Review of Economics and Statistics*, 68, 452–59. 51, 119

Mehra, R., and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(2), 145–61. 51, 119

Menzly, L., T. Santos, and P. Veronesi (2004): "Understanding Predictability," *Journal of Political Economy*, 112(1), 1–47. 123

Nardi, Y., and A. Rinaldo (2008): "On the Asymptotic Properties of the Group Lasso Estimator for Linear Models," *Electronic Journal of Statistics*, 2, 605–33. 100

Newey, W., and R. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255. 131

Novy-Marx, R. (2013): "The Other Side of Value: The Gross Profitability Premium," *Journal of Financial Economics*, 108(1), 1–28. 4

——— (2014): "Predicting Anomaly Performance with Politics, the Weather, Global Warming, Sunspots, and the Stars," *Journal of Financial Economics*, forthcoming. 1

Parker, J. A., and C. Julliard (2003): "Consumption Risk and Cross-Sectional Returns," NBER Working Papers 9538, National Bureau of Economic Research, Inc. 132, 133, 145

——— (2005): "Consumption Risk and the Cross-Section of Expected Returns," *Journal of Political Economy*, 113(1). 1, 121, 125, 127, 144, 145

Piazzesi, M. (2010): "Affine Term Structure Models," in *Ait-Sahalia, Y., Hansen L.P. (Eds.) Handbook of Financial Econometrics*, vol. 1, pp. 691–758. Elsevier, Amsterdam, Netherlands. 89

Piazzesi, M., M. Schneider, and S. Tuzel (2007): "Housing, consumption and asset pricing," *Journal of Financial Economics*, 83(3), 531–569. 123

Politis, D. N., and J. P. Romano (1994): "The Stationary Bootstrap," *The Journal of American Statistical Association*, 89(428), 1303–1313. 45, 48, 54

Polk, C., and P. Sapienza (2009): "The Stock Market and Corporate Investment: A Test of Catering Theory," *Review of Financial Studies*, 22, 187–217. 4

Pollard, D. (1991): "Asymptotics for Least Absolute Deviations Regression Estimators," *Econometric Theory*, 7, 186–99. 77, 81, 113

——— (1994): "On the Asymptotics of Constrained M-estimation," *Annals of Statistics*, 22, 1993–2010. 80, 84

PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852. 129

SANTOS, T., AND P. VERONESI (2004): "Labor Income and Predictable Stock Returns," *Review of Financial Studies*, Forthcoming. 52

SHANKEN, J. (1985): "Multivariate Tests of the Zero-beta CAPM," *Journal of Financial Economics*, 14, 327–47. 27

——— (1992): "On the Estimation of Beta-Pricing Models," *The Review of Financial Studies*, 5, 1–33. 17, 24, 81, 112

SHARPE, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19(3), 425–42. 1

STOCK, J., AND M. WATSON (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44(2), 293–335. 102

STOCK, J., M. WATSON, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments.," *Journal of Business and Economic Statistics*, 20, 518 29. 5

STOCK, J. H., AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Economic Models: Essays in Honour of Thomas Rothenberg*, ed. by D. Andrews, and J. Stock, pp. 80–108. Cambridge, UK: Cambridge University Press. 3

STUTZER, M. (1995): "A Bayesian approach to diagnosis of asset pricing models," *Journal of Econometrics*, 68(2), 367 – 397. 131

TIBSHIRANI, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B*, 58, 267–88. 6, 13, 22

TITMAN, S., K. WEI, AND F. XIE (2004): "Capital Investments and Stock Returns," *Journal of Financial and Quantitative Analysis*, 39, 677–700. 4

TSENG, P. (1988): "Coordinate Ascent for Maximizing Nondifferentiable Concave Functions," *Technical Report LIDS-P 1840*. 15

——— (2001): "Convergence of Block Coordinate Descent Method for Nondifferentiable Maximation," *Journal of Optimization Theory Applications*, 109, 474–94. 15

ULRICH, M. (2010): "Observable Long-Run Ambiguity and Long-Run Risk," Columbia University Manuscript. 123

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–88. 88

Weil, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421. 51, 119

Xing, Y. (2008): "Interpreting the Value Effect through the q-theory: An Empirical Investigation," *Review of Financial Studies*, 21, 1767–95. 4

Yogo, M. (2006): "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance*, 61(2), 539–580. 1, 44, 51, 53, 123

Zhao, P., and B. Yu (2006): "On Model Selection Consistency of Lasso," *Journal of Machine Learning Research*, (7), 2541–63. 100

Zhiang, L., and Z. Zhan (2013): "An Empirical Comparison of Non-traded and Traded Factors in Asset Pricing," Working paper. 4

Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of Americal Statistical Association*, 101, 1418–29. 6, 13