Improved Tests for Spatial Autoregressions

Francesca Rossi

Dissertation submitted for the degree of

Doctor of Philosophy in Economics at

The London School of Economics and Political Science

18/07/2011

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.

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

Francesca Rossi

Abstract

Econometric modelling and statistical inference are considerably complicated by the possibility of correlation across data data recorded at different locations in space. A major branch of the spatial econometrics literature has focused on testing the null hypothesis of spatial independence in Spatial Autoregressions (SAR) and the asymptotic properties of standard test statistics have been widely considered. However, finite sample properties of such tests have received relatively little consideration. Indeed, spatial datasets are likely to be small or moderately-sized and thus the derivation of finite sample corrections appears to be a crucially important task in order to obtain reliable tests. In this project we consider finite sample corrections based on formal Edgeworth expansions for the cumulative distribution function of some relevant test statistics.

In Chapter 1 we provide the background for the results derived in this thesis. Specifically, we describe SAR models together with some established results in first order asymptotic theory for tests for independence in such models and give a brief account on Edgeworth expansions. In Chapters 2 and 3 we present refined procedures for testing nullity of the spatial parameter in pure SAR based on ordinary least squares and Gaussian maximum likelihood, respectively. In both cases, the Edgeworth-corrected tests are compared with those obtained by a bootstrap procedure, which is supposed to have similar properties. The practical performance of new tests is assessed with Monte Carlo simulations and two empirical examples. In Chapter 4 we propose finite sample corrections for Lagrange Multiplier statistics, which are computationally particularly convenient as the estimation of the spatial parameter is not required. Monte Carlo simulations and the numerical implementation of Imhof's procedure confirm that the corrected tests outperform standard ones. In Chapter 5 the slightly more general model known as "mixed" SAR is considered. We derive suitable finite sample corrections for standard tests when the parameters are estimated by ordinary least squares and instrumental variables. A Monte Carlo study again confirms that the new tests outperform ones based on the central limit theorem approximation in small and moderately-sized samples.

Al nonno

Acknowledgments

First of all, I am greatly indebted to my advisor, Professor Peter Robinson, for his invaluable support and guidance.

Many thanks are due to Abhisek Banerjee, Zsofia Barany, Ziad Daoud, Abhimanyu Gupta, Javier Hidalgo, Tatiana Komarova, Jungyoon Lee, Oliver Linton, Elena Manzoni, Myung Hwan Seo, Marcia Schafgans and Sorawoot Tang Srisuma for plenty of helpful comments and discussions.

I also thank all my friends and fellow researchers at the LSE, in particular the participants at the work in progress seminars, for discussions and comments on my work.

Financial support from the Bank of Italy, Fondazione Luigi Einaudi and ESRC Grant RES-062-23-0036 is gratefully acknowledged.

I owe a lot to my family, and in particular to my parents, for their support over the past few years, and to Antonio, for being so patient and supportive. All this would not have been possible without them.

I also thank all the friends who made me feel at home in London.

Contents

Abstract **3**

Acknowledgments **5**

Contents **6**

List of Figures **8**

List of Tables **9**

1 Introduction **11**
- 1.1 Spatial Autoregressions 11
- 1.2 First order statistics 14
- 1.3 Edgeworth expansions 19
- 1.4 Finite sample issues and contribution of this thesis 23
- 1.5 Some definitions and Assumptions 25

2 Improved OLS Test Statistics for Pure SAR **28**
- 2.1 Test against a one-sided alternative: Edgeworth-corrected critical values and corrected statistic 28
- 2.2 Test against a two-sided alternative: Edgeworth-corrected critical values and corrected statistic 32
- 2.3 Corrected critical values and corrected statistic for pure SAR with a location parameter 34
- 2.4 Test against a local alternative 35
- 2.5 Bootstrap correction and simulation results 39
- A Appendix 49

3 Improved Test Statistics based on MLE for Pure SAR **63**
- 3.1 Test against a one-sided alternative: Edgeworth-corrected critical values and corrected statistic 64
- 3.2 Bootstrap correction and Monte Carlo results 66
- 3.3 Empirical evidence: the geography of happiness 70
- 3.4 Empirical evidence: the distribution of crimes in Italian provinces . . . 73
- A Appendix 75
  - A.1 Proof of Theorem 3.1 75
  - A.2 Auxiliary results 81

4 Finite Sample Corrections for the LM Test in SAR Models **92**
- 4.1 Edgeworth-corrected LM tests for independence in pure SAR 92

4.2 Improved LM tests in regressions where the disturbances are spatially correlated . . . 95
4.3 Alternative correction . . . 98
4.4 Bootstrap correction and simulation results . . . 101
4.5 The exact distribution . . . 105
A Appendix . . . 109

5 Refined Tests for Mixed SAR Models 119
5.1 Refined tests based on OLS estimates . . . 119
5.2 Refined tests based on IV estimates . . . 124
5.3 Finite sample corrections for testing β. . . . 128
5.4 Bootstrap and Monte Carlo results . . . 130
A Appendix . . . 135

References 149

List of Figures

2.1 Simulated pdf of $a\hat{\lambda}$ under H_0 in (1.2.1) 43
2.2 Simulated pdf of $g(a\hat{\lambda})$ under H_0 in (1.2.1) 43
3.1 Simulated pdf of $\tilde{a}\tilde{\lambda}$ under H_0 in (1.2.1) 68
3.2 Simulated pdf of $\tilde{g}(\tilde{a}\tilde{\lambda})$ under H_0 in (1.2.1) 69

List of Tables

2.1 Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.1.1) when λ in model (1.2.5) is estimated by OLS and the sequence h is "divergent". The reported values have to be compared with the nominal 0.05. 42

2.2 Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.1.1) when λ in model (1.2.5)is estimated by OLS and the sequence h is "bounded". The reported values have to be compared with the nominal 0.05. 42

2.3 Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.2.1) when λ in model (1.2.5) is estimated by OLS and the sequence h is "divergent". The reported values have to be compared with the nominal 0.05. 44

2.4 Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.2.1) when λ in model (1.2.5)is estimated by OLS and the sequence h is "bounded". The reported values have to be compared with the nominal 0.05. 44

2.5 Empirical powers of the tests of H_0 in (1.2.1) against H_1 in (2.5.3), with $\bar{\lambda} = 0.1, 0.5, 0.8$, when λ in model (1.2.5) is estimated by OLS and the sequence h is "divergent". α is set to 0.95. 45

2.6 Empirical powers of the tests of H_0 in (1.2.1) against H_1 in (2.5.3), with $\bar{\lambda} = 0.1, 0.5, 0.8$, when λ in model (1.2.5) is estimated by OLS and the sequence h is "bounded". α is set to 0.95. 45

2.7 Numerical values corresponding to (2.5.5) (second row) and (2.5.6) (third row), compared with the simulated values for the power of a test of (1.2.1) against (2.4.1) when λ_n in model (1.2.5) is estimated by OLS and the sequence h is "divergent". 46

2.8 Numerical values corresponding to (2.5.5) (second row) and (2.5.6) (third row), compared with the simulated values for the power of a test of (1.2.1) against (2.4.1) when λ_n in model (1.2.5) is estimated by OLS and the sequence h is "bounded". 47

2.9 Simulated values of the power of a test of (1.2.1) against (2.4.1) based on the standard and corrected statistics when λ_n in model (1.2.5) is estimated by OLS and the sequence h is "divergent". The values should be compared with the target 0.304. . 47

2.10 Simulated values of the power of a test of (1.2.1) against (2.4.1) based on the standard and corrected statistics when λ_n in model (1.2.5) is estimated by OLS and the sequence h is bounded. The values should be compared with the target 0.304. . . . 48

3.1 Empirical sizes of the tests of H_0 in (1.2.1) when λ in (1.2.5) is estimated by MLE and the sequence h is "divergent". The reported values have to be compared with the nominal 0.05. 67

3.2 Empirical sizes of the tests of H_0 in (1.2.1) when λ in (1.2.5) is estimated by MLE and the sequence h is "bounded". The reported values have to be compared with the nominal 0.05. 67

3.3 Empirical powers of the tests of H_0 in (1.2.1) against H_1 in (2.5.3) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ in (1.2.5) is estimated by MLE and the sequence h is "divergent". α is set to 0.95. 69

3.4 Empirical powers of the tests of H_0 in (1.2.1) against (2.5.3) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ in (1.2.5) is estimated by MLE and the sequence h is "bounded". α is set to 0.95. 70
3.5 Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.3.2) is estimated by OLS 72
3.6 Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.3.2) is estimated by MLE 72
3.7 Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.4.1) is estimated by OLS 74
3.8 Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.4.1) is estimated by MLE 74
4.1 Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.5) when the sequence h is "divergent". The reported values have to be compared with the nominal 0.05. 103
4.2 Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.5) when the sequence h is "bounded". The reported values have to be compared with the nominal 0.05. 103
4.3 Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.15) when the sequence h is "divergent". The reported values have to be compared with the nominal 0.05. 104
4.4 Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.15) when the sequence h is "bounded". The reported values have to be compared with the nominal 0.05. 104
4.5 Edgeworth-corrected and Imhof's α-quantiles of the cdf of T in when h is "divergent". 108
4.6 Edgeworth-corrected and Imhof's α-quantiles of the cdf of T when h is "bounded". 108
5.1 Empirical sizes of the tests of (1.2.1) against a one-sided alternative when λ and β in (1.1.3) are estimated by OLS. The reported values have to be compared with the nominal 0.05. 133
5.2 Empirical sizes of the tests of (1.2.1) against a one-sided alternative when λ and β in (1.1.3) are estimated by IV. The reported values have to be compared with the nominal 0.05. 133
5.3 Empirical sizes of the tests of (5.3.1) with $R = (0, 1)'$ against a one-sided alternative. λ and β in (1.1.3) are estimated by IV and $\lambda = 0.1, 0.7$. The reported values have to be compared with the nominal 0.05. 134
5.4 Empirical sizes of the tests of (5.3.1) with $R = (-1, 1)'$ against a one-sided alternative. λ and β in (1.1.3) are estimated by IV and $\lambda = 0.1, 0.7$. The reported values have to be compared with the nominal 0.05. 135

1 Introduction

This chapter provides the background to appreciate the specific contribution of this thesis to the existing spatial econometric literature and the main theoretical techniques we will employ. Specifically, in Section 1.1 we discuss some issues that arise with spatial data, together with a description of spatial autoregressions. In Section 1.2 we give an account of established first order asymptotic theory for testing for lack of correlation in spatial autoregressions. In Section 1.3 we report an overview on Edgeworth expansions and their application to derive improved tests, while in Section 1.4 we describe the motivation of this thesis, in view of some of the existing results discussed in Sections 1.2 and 1.3. Finally, in Section 1.5 we introduce some definitions and assumptions that will be used in Chapters 2-5.

1.1 Spatial Autoregressions

Econometricians face considerable challenges posed by the possibility of cross-sectional correlation, with respect to both modelling and statistical inference. Indeed, starting from the early work by Moran (1950), Cliff and Ord (1968, 1972) and, more recently, Cressie (1993), just to name a few, a large body of literature known as Spatial Econometrics has addressed issues entailed by potential correlation across data recorded at different locations in space. For recent reviews and discussions of the challenges and progresses in the spatial econometric literature, refer to Robinson (2008a) and Anselin (2010).

Much of the spatial statistical literature has focused on data recorded on a lattice, i.e. regularly-spaced observations on a $d-$dimensional space, where $d > 1$. In general, intervals between observations are constant within dimensions, but are allowed to vary across different dimensions. Some extensions of standard asymptotic theory for the time series setting ($d = 1$) to the case $d > 1$ are possible, as first noted by Whittle (1954). Indeed, Whittle (1954) demonstrates that, in general, multilateral models have a "half-plane" type of unilateral moving average representation which extends the well known Wold representation for time series data, and hence suggests estimates of the underlying unknown parameters based on an approximation for the Gaussian log-likelihood function. However, such half-plane representations might contain functions of the coefficients of the multilateral model that cannot be expressed in closed form. Furthermore, a serious source of complication arising with lattice data for $d > 1$ is the bias of the estimates entailed by the "edge effect". Techniques to overcome such bias are developed in Guyon (1982), Dahlhaus and Künsh (1987) and Robinson and Vidal Sanz (2006).

However, observations recorded on a lattice are very uncommon in economics. For example, in geographical settings there is irregular spacing when observations are recorded across cities, regions or countries. When the data are recorded at irregularly-spaced geographical locations, a generalisation of the established theory for irregularly-spaced time series is still possible. Specifically, some cases of irregularly-spaced time series can be described by an underlying continuous time process where spacing is generated by a point process. When the continuous time process is a first order stochastic differential equation with constant coefficients and driven by white noise, consistent and asymptotically normal estimates of the unknown parameters can be obtained from an approximated Gaussian log-likelihood (see Robinson (1977)). This framework can in principle be extended to more general models, but estimation and asymptotic theory become complicated.

In any case, "space" should be more generally intended as a network, which includes physical/geographical space as a very special case, and in turn correlation across observations may depend on some very general notion of economic distance (e.g. differences in household income) that does not necessarily have a geographical interpretation (see e.g. Conley and Ligon (2002) or Conley and Dupor (2003)). The economic distance between units (or economic agents) i and j is defined as the distance between u_i and u_j, where u_i and u_j are vectors of characteristics pertaining to agents i and j, respectively. The distance between u_i and u_j might be defined in an Euclidean sense. The aforementioned extensions of the theory for irregularly-spaced time series to spatial data are unsuitable in case there is no geographical aspect.

Spatial autoregressions (SAR) offer a useful, applicable framework for describing such data. In SAR models the notion of possible irregular spacing, applied to general economic distances, is embodied in an $n \times n$ weight matrix (n being sample size), denoted W_n, which needs to be chosen by the practitioner. Let w_{ij} be the $(i, j)-$th element of W_n. Conventionally, $w_{ii} = 0$ for $i = 1, ...n$, i.e. the spatial interaction of each unit (or economic agent) with itself is set to zero. Although in principle w_{ij} can be negative, in most practical applications W_n has non negative entries and is row normalized, so that elements of each row sum to 1. In view of such normalization, w_{ij} can be defined in terms of the inverse of an economic distance d_{ij} between units i and j, i.e.

$$w_{ij} = \frac{d_{ij}}{\sum_{s=1}^{n} d_{is}}, \tag{1.1.1}$$

where $d_{ij} \geq 0$ and possibly $d_{ij} \neq d_{ji}$, i.e. symmetry of the spatial interaction between units i and j (or economic agents i and j) is not imposed.

For instance, when the data are recorded across different regions or countries W_n can be chosen according to a contiguity criterion, i.e. $w_{ij} = 1$ if regions or countries share a border and $w_{ij} = 0$ otherwise. Eventually, the resulting matrix can then be the row normalized so that $\sum_{j=1}^{n} w_{ij} = 1$ for all i. Alternatively, w_{ij} can be defined as inverse of the geographical distance (e.g. measured in miles of kilometers) between locations i and j. In such case, the resulting W_n is not row normalized. In the empirical applications considered in this thesis, we will choose W_n according to the former specification, i.e. based on a contiguity criterion and then row normalized.

An example of W_n, introduced by Case (1991), that has been extensively used to illustrate theoretical results is

$$W_n = I_r \otimes B_m, \qquad B_m = \frac{1}{m-1}(l_m l_m' - I_m), \tag{1.1.2}$$

where $n = rm$, r being the number of districts and m the number of households in each district. In (1.1.2), $\otimes$ indicates the Kronecker product, l_m an $m-$ dimensional column of ones and I_r the $r \times r$ identity matrix. Henceforth, we retain the subscript to either l or I only when the dimension is other than n, i.e. throughout l and I denote an $n-$dimensional column of ones and the $n \times n$ identity matrix. Under (1.1.2), two households are neighbours if they belong to the same district, and each neighbour is given the same weight. Since W_n in (1.1.2) is symmetric and block diagonal, (1.1.2) is a convenient choice computationally and hence it has often been adopted in Monte Carlo simulations to illustrate theoretical results. Indeed, throughout this thesis we will employ (1.1.2) for our simulation studies. It should be stressed that although (1.1.2) has been introduced by Case (1991) in a geographical setting, the block diagonal structure of (1.1.2) can be used to describe more general situations where each unit (agent) is equally influenced by units (agents) with similar characteristics and is not affected by other units (agents) in the economy.

Although the choice of W_n plays a central role in deriving asymptotic theory for spatial data and is a crucial empirical issue, we should outline that in this thesis we will deal with tests for spatial independence and most of our results are derived under the null hypothesis of no spatial correlation. For this reason, our results would be valid even in case W_n is not correctly chosen. However, efficiency of tests is affected by the choice of W_n.

Let Y_n be an $n \times 1$ vector of observations, X_n an $n \times k$ matrix of exogenous regressors of full column rank which might include a column of ones, and ϵ_n an $n \times 1$ vector of independent and identically distributed (iid) random variables, with mean zero and unknown variance σ^2. We assume that, for some unknown scalar λ and some unknown $k \times 1$ vector β, the data follow a general SAR model, i.e.

$$Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n. \tag{1.1.3}$$

For notational simplicity, in the sequel we drop the n subscript, writing $\epsilon = \epsilon_n$, $Y = Y_n$, $X = X_n$, $W = W_n$, with the same convention for other $n-$dependent quantities.

Model (1.1.3) is a very parsimonious method of describing spatial dependence, conveniently depending only on economic distances rather than actual locations, which may be unknown or not relevant. For sake of clarity, it should be stressed that often in the spatial econometric literature "spatial independence" is used as a synonym for "lack of spatial correlation", though these concepts are in general identical only under

Gaussianity. Although a major drawback of SAR models is the ex ante specification of W, to which parameter estimates are sensitive, (1.1.3) has been widely used in practical applications. Relevant book-length descriptions of SAR model and its applications include Anselin (1988) and Arbia (2006). Even more importantly, (1.1.3) represents a convenient, widely-usable class of alternatives in testing the null hypothesis of lack of spatial correlation which, if true, considerably simplifies statistical inference. Much of the results in this thesis (Chapters 2, 3 and 4) are derived under the assumption that $\beta = 0$ a priori in (1.1.3), or that $X = l$, i.e. (1.1.3) only contains an intercept.

Although in this thesis we will only deal with SAR models, we acknowledge that an interesting alternative approach to describe spatial interactions based on economic distances has been formulated by Conley (1999). In Conley (1999), economic agents' observations are modelled as realizations of a random process at points of a Euclidean space and the distance between two agents in such Euclidean space reflects their proximity (economic distance). Under mixing conditions and in a random field setting, Conley (1999) derives asymptotic theory for various estimates. Conley (1999) and Conley and Molinari (2007) also extend such a framework in case of measurement error in the economic distance.

1.2 First order statistics

The problem of testing the null hypothesis

$$H_0 : \lambda = 0 \tag{1.2.1}$$

in (1.1.3), or in a related model where the spatial correlation potentially also affects the unobservable disturbances, i.e.

$$Y_n = \lambda W_{n,1} Y_n + X_n \beta + u_n \quad u_n = \rho W_{n,2} u_n + \epsilon_n, \tag{1.2.2}$$

$W_{n,1}$ and $W_{n,2}$ being suitable weight matrices and ρ a scalar parameter, is a long lasting issue in the spatial econometric literature.

When the focus of the investigation is both on estimation and testing of λ in models (1.1.3) or (1.2.2) various tests of (1.2.1) based on different estimates of λ have been proposed and widely used by practitioners. Ordinary Least Squares (OLS) estimation of λ and β in (1.1.3) was dismissed without thorough investigation in early work since WY in (1.1.3) is correlated with ϵ and hence OLS estimates are generally inconsistent. However, Lee (2002) shows that OLS estimates of λ and β in (1.1.3) can be consistent and asymptotically normal for some choices of W. Throughout, the OLS estimates of λ and β are denoted by $\hat{\lambda}$ and $\hat{\beta}$, respectively. Let $h = h_n$ be a sequence bounded away from zero for all n. For w_{ij} given in (1.1.1) and such that $\sum_{s=1}^{n} d_{is}$ is uniformly bounded away from zero at rate h, i.e. for all n

$$0 < c_1 < \frac{\sum_{s=1}^{n} d_{is}}{h},$$

where $\sqrt{n}/h = o(1)$ and c_1 being a generic arbitrarily small constant, Lee (2002) shows that as $n \to \infty$,

$$\sqrt{n}(\hat{\lambda} - \lambda, (\hat{\beta} - \beta)')' \xrightarrow{d} N(0, V_{OLS}), \tag{1.2.3}$$

$\xrightarrow{d}$ and prime indicating convergence in distribution and transposition, respectively, and

$$V_{OLS} = \sigma^2 \left(\lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} E(Y')W' \\ X' \end{pmatrix} (WE(Y), X) \right)^{-1}.$$

In case h diverges at rate $\sqrt{n}$, Lee (2002) shows

$$\sqrt{n}(\hat{\lambda} - \lambda, (\hat{\beta} - \beta)')' \xrightarrow{d} N(b, V_{OLS}), \tag{1.2.4}$$

where b is an asymptotic bias that vanishes only when $\lambda = 0$. Such results do not hold in case $h/\sqrt{n} = o(1)$. Although t-type of tests of (1.2.1) based on $\hat{\lambda}$ and $\hat{\beta}$ are computationally very simple, the aforementioned strong condition on W restricts their applicability.

Lee (2002) also shows that in case β in (1.1.3) is zero a priori, i.e. when the data follows a "pure" SAR

$$Y = \lambda WY + \epsilon, \tag{1.2.5}$$

$$\hat{\lambda} = \frac{Y'W'Y}{Y'W'WY} \tag{1.2.6}$$

is inconsistent and more generally, the estimate of λ is inconsistent when (1.1.3) only includes an intercept. However, in each of these cases, under H_0 in (1.2.1), the OLS estimate of λ actually does converge to zero in probability. Although the case $\lambda = 0$ is very limited when the interest is estimation, it is a leading one in testing and we will consider it in detail in Chapter 2. We will also show that when the data are driven by (1.2.5), under H_0 in (1.2.1), the rate of convergence of $\hat{\lambda}$ might be slower than the parametric $\sqrt{n}$, depending on assumptions on W.

Procedures based on Gaussian maximum likelihood estimates (MLE) for λ and β in (1.1.3) and (1.2.2) have been developed by Cliff and Ord (1975) and broadly considered. For an exhaustive survey about specification and implementation of tests of (1.2.1) based on the Gaussian MLE of parameters in (1.1.3) and (1.2.2), refer to Anselin (1988). Asymptotic properties of MLE and Pseudo-MLE (i.e. estimates obtained by maximization of a Gaussian log-likelihood function when normality of the error terms is not assumed) of λ, β and σ in (1.1.3), denoted $\tilde{\lambda}$, $\tilde{\beta}$, $\tilde{\sigma}$ henceforth, and relevant test statistics have been derived in Lee (2004). Henceforth PMLE indicates Pseudo-MLE.

The pseudo log-likehood function of (1.1.3) is defined as

$$l(\lambda,\beta,\sigma^2) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln\sigma^2 + ln(det(S(\lambda))) - \frac{1}{2\sigma^2}(S(\lambda)Y - X\beta)'(S(\lambda)Y - X\beta), \tag{1.2.7}$$

where, for every value of λ,

$$S(\lambda) = I - \lambda W \tag{1.2.8}$$

and $det(A)$ denotes the determinant of a generic square matrix A. It should be stressed that in (1.2.7), λ, β and σ denote any admissible value of the parameters in (1.1.3). Given λ, the PMLE of β and σ^2 are

$$\tilde{\beta}(\lambda) = (X'X)^{-1}X'S(\lambda)Y \tag{1.2.9}$$

and

$$\tilde{\sigma}^2(\lambda) = \frac{1}{n}(Y'S(\lambda)' - \tilde{\beta}(\lambda)'X')(S(\lambda)Y - X\tilde{\beta}(\lambda)), \tag{1.2.10}$$

respectively. Hence

$$\tilde{\lambda} = \arg\max_{\lambda\in\Lambda}\; l(\lambda, \tilde{\beta}(\lambda), \tilde{\sigma}^2(\lambda)), \tag{1.2.11}$$

where Λ is a compact set included in $(-1, 1)$. The PMLE of β and σ^2 are then defined as $\tilde{\beta} = \tilde{\beta}(\tilde{\lambda})$ and $\tilde{\sigma}^2 = \tilde{\sigma}^2(\tilde{\lambda})$.

In particular, assuming $w_{ij} = O(1/h)$, h being either divergent or bounded and such that $h/n \to 0$ as $n \to \infty$, Lee (2004) proves that under standard conditions,

$$\sqrt{n}(\tilde{\lambda} - \lambda, (\tilde{\beta} - \beta)')' \xrightarrow{d} N(0, V_{PMLE}), \tag{1.2.12}$$

where the explicit form of V_{PMLE} is given in Lee (2004) and is not reported here in order to avoid introducing further unnecessary notation. It should be stressed that when $\beta \neq 0$ and the elements of ϵ are normally distributed, $V_{PMLE} = V_{OLS}$. Although estimation of λ in (1.2.5) can be regarded as a constraint maximization of (1.2.11) when $\beta = 0$, Lee (2004) shows that in this case

$$\sqrt{\frac{n}{h}}\tilde{\lambda} \xrightarrow{d} N(0, V_{PMLE}^{pure}), \tag{1.2.13}$$

where V_{PMLE}^{pure} denotes the asymptotic variance of $\sqrt{n/h}\tilde{\lambda}$ in case the data follow (1.2.5). From (1.2.13) it is clear that the rate of convergence of $\tilde{\lambda}$ to λ can be slower than $\sqrt{n}$ when h is divergent. Test statistics based on $\tilde{\lambda}$ when the data are driven by (1.2.5) will be the focus of Chapter 3.

Although the MLE (or PMLE, more generally) has been extensively used for both estimation and testing, it is well known that it is computationally challenging when n is large (see e.g. Pace and Berry (1997)). In order to reduce the computational burden, tests of (1.2.1) based on alternative estimates of λ have been proposed. Instrumental Variable (IV) estimates of λ, and β have been introduced by Kelejian and Prucha

(1998) and subsequently improved by Lee (2003).

In particular, Kelejian and Prucha (1998) derive asymptotic properties of IV estimates of parameters in (1.2.2). Although Kelejian and Prucha (1998) show consistency and asymptotic normality of IV estimates of λ and β, denoted $\hat{\lambda}_{IV}$ and $\hat{\beta}_{IV}$ henceforth, in the more general model (1.2.2), we report here only their results pertaining to (1.1.3). Let Z be a $n \times (k+1)$ matrix of instruments. Under standard assumptions, Kelejian and Prucha (1998) prove

$$\sqrt{n}(\hat{\lambda}_{IV} - \lambda, (\hat{\beta}_{IV} - \beta)')' \xrightarrow{d} N(0, V_{IV}), \tag{1.2.14}$$

where

$$V_{IV} = \sigma^2 \left(\lim_{n\to\infty} \frac{1}{n} Z'(WE(Y), X) \right)^{-1} \left(\lim_{n\to\infty} \frac{1}{n} Z'Z \right) \left(\lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} E(Y')W' \\ X' \end{pmatrix} Z \right)^{-1}$$

The latter result holds under very weak conditions on W, and more details about those will be provided in Chapter 5. The "ideal" choice of instruments is $Z = (WE(Y), X)$ and with such choice $V_{IV} = V_{OLS}$. Also, with such choice for Z and under normality of the error terms, $V_{IV} = V_{MLE}$. Kelejian and Prucha (1998) propose to approximate the "ideal" instrument, which is clearly infeasible, with a subset of the linearly independent columns of $(X, WX, W^2X, ...)$. However, Kelejian and Prucha (1998) do not consider relative efficiency issues and, in general, their choice of Z is suboptimal.

In turn, Lee (2003), improves the asymptotic efficiency of the Kelejian and Prucha IV estimator. In a more recent paper, Kelejian et al. (2004) introduce a series-type IV estimator for model (1.2.2), which is proved to be asymptotically normal, efficient within the class of IV estimators and computationally simpler than one proposed in Lee (2003).

We should mention that, although the test of (1.2.1) is generally the main focus, we might be interested in testing restrictions on β in (1.1.3) and (1.2.2). However, IV estimates cannot be obtained in case $\beta = 0$ in (1.1.3) and (1.2.2) and hence tests for the joint significance of $\beta_1, ...\beta_k$ (β_i being the $i-$th element of β) are not possible in this setting.

In Chapter 5 we will focus on tests for λ and β in (1.1.3) based on $\hat{\lambda}_{IV}$ and $\hat{\beta}_{IV}$, starting from Kelejian and Prucha (1998) main framework.

On the other hand, when the interest of the practitioner is testing rather than estimation, a class of tests based on Langrange Multiplier (LM) statistics has received considerable attention starting from the early contribution by Moran (1950). Such tests are computationally very convenient as the estimation of λ in either model (1.1.3) or (1.2.2) is not required.

Moran (1950) presents a simple correlation test between neighbours in space based on a normalized quadratic form in the variables that are being tested, without specifying the alternative hypothesis. Moran's result has been applied by Cliff and Ord

(1972, 1981) to test H_0 in (1.2.1) in regression models with SAR disturbances, i.e.

$$Y = X\beta + u, \quad u = \lambda W u + \epsilon. \tag{1.2.15}$$

(1.2.15) is equivalent to (1.2.2) when λ in (1.2.2) is zero a priori, but we separate the two models to outline clearly whether the parameter that we wish to test is the SAR coefficient when the disturbances are also potentially correlated (i.e. λ in (1.2.2)) or that of the disturbance term of a linear regression (i.e. λ in (1.2.15)). In particular, assuming normality of the components of ϵ in (1.2.15), Cliff and Ord (1972) show that Moran's statistic under H_0 in (1.2.1) has a χ^2_1 asymptotic distribution . Cliff and Ord's (1972) normality assumption has been relaxed by Sen (1976), who derives the χ^2_1 asymptotic distribution of Moran's test for H_0 in (1.2.1) when the data follow (1.2.15) assuming that the components of ϵ are iid and under some specific moment conditions. Though Moran test statistic (and its aforementioned extensions) was not originally derived in a ML framework, Burridge (1980) shows it is indeed equivalent to a LM statistic for spatially uncorrelated disturbances. Details of LM statistics to test H_0 in (1.2.1) will be given below.

Although this is not explicitly considered in this thesis, Kelejian and Robinson (1992) derive an alternative test for spatial independence against correlation of unspecified form in the disturbance term of regression models (possibly nonlinear) based on regression residuals. Kelejian and Robinson (1992) do not refer explicitly to a weight matrix and the ordering of observations is based on first order contiguity. Similarly to Moran's, Kelejian and Robinson test has a χ^2 limiting distribution under (1.2.1) and its asymptotic properties have been derived without assuming normality of the error terms. However, the small sample performance of Kelejian and Robinson test is quite poor, as shown in a number of Monte Carlo studies (e.g. Anselin and Florax (1995) and Kelejian and Robinson (1998)).

Anselin (2001) provides an exhaustive survey of derivation and implementation issues of Moran/LM tests of (1.2.1) when the data follow either (1.1.3) or (1.2.2). As regarding asymptotic theory of Moran/LM test statistics, Kelejian and Prucha (2001) derive a central limit theorem for quadratic forms in random variables which allows to establish the asymptotic distribution of LM statistics for SAR models under H_0 in (1.2.1). This result is general enough to accommodate non linearity and the possibility of heteroskedastic error terms. Also, Pinske (1999, 2004) outlines a set of conditions for asymptotic normality (or asymptotic χ^2) of several Moran/LM-type of test statistics, which include LM statistics for testing (1.2.1) in (1.1.3), (1.2.2) and (1.2.15).

We briefly outline here some details on the construction of a version of the LM statistic to test H_0 in (1.1.3), its modification to either (1.2.2) or (1.2.15) is straightforward. Given (1.2.7), a version of the LM statistic to test (1.2.1) in model (1.1.3) is defined as

$$LM = \left(\frac{\partial l(\lambda, \beta, \sigma^2)}{\partial \lambda} |_{H_0} \right)^2 \left(-E \left(\frac{\partial^2 l(\lambda, \beta, \sigma^2)}{\partial \lambda^2} \right) |_{H_0} \right)^{-1}. \qquad (1.2.16)$$

From (1.2.7), by standard partial differentiation,

$$\frac{\partial l(\lambda, \beta, \sigma^2)}{\partial \lambda} |_{H_0} = n \frac{\hat{\epsilon}' W Y}{\hat{\epsilon}' \hat{\epsilon}},$$

where $\hat{\epsilon} = Y - X\hat{\beta}^r$ and $\hat{\beta}^r$ is the OLS estimate of β in (1.1.3) under H_0 ("restricted" model, hence the superscript "r") in (1.2.1) . Similarly,

$$-E \left(\frac{\partial^2 l(\lambda, \beta, \sigma^2)}{\partial \lambda^2} |_{H_0} \right) = tr(W^2) + tr(W'W) + n \frac{(WX\hat{\beta}^r)' P W X \hat{\beta}^r)}{\hat{\epsilon}' \hat{\epsilon}},$$

where

$$P = I - X(X'X)^{-1} X \qquad (1.2.17)$$

and tr indicates the trace operator. The derivation of (1.2.16) is based on a Gaussian likelihood but the same first order limit distribution obtains more generally. Indeed, as anticipated, under suitable conditions we have

$$LM \xrightarrow{d} \chi_1^2. \qquad (1.2.18)$$

Tests of (1.2.1) based on (1.2.16) (or its appropriate modification) for either (1.2.5) or (1.2.15) will be considered in detail in Chapter 4.

More generally, Robinson (2008b) derives the asymptotic distribution under the null hypothesis of lack of correlation of a class of residual-based test statistics, which include LM for either (1.1.3) or (1.2.2) as special cases. As expected, by considering the asymptotic distribution of such residual-based class of statistics under a local alternative, LM tests are motivated because they are locally optimal within this class. Finite sample improvements of test statistics under the null hypothesis of lack of correlation are also suggested. Robinson (2008b) results will be presented and discussed in detail in Chapter 4.

1.3 Edgeworth expansions

In Section 1.2 we mentioned which test statistics will be considered in Chapters 2-5. However, before illustrating more precisely the contribution of this thesis, in this section we give a brief account of existing literature on Edgeworth expansions, which are indeed the main methodological instruments for the derivation of our results. The literature on Edgeworth expansions and their applications in econometric and statistical theory is very broad and here we only aim to provide some of the main references together with a brief description of existing results that were useful to develop this project, although we acknowledge that this is not a complete survey.

The idea of (formally) expanding distribution functions was introduced by Edgeworth (1896, 1905) for sums of iid random variables. For useful and relatively simple surveys which deal with the derivation of Edgeworth expansions, refer to Rothenberg (1984) and Barndorff-Nielsen and Cox (1989, Chapter 4). Here, we illustrate briefly the derivation of the Edgeworth expansion for the cumulative distribution function (cdf) of the standardized sample mean of iid random variables and how such derivation can be extended to the case of quadratic statistics in normal random variables, which are the main focus of this thesis.

Specifically, let $U_1, U_2, ...U_n$ be a sample of iid random variables with mean $m = 0$ and variance $Var(U_i) = 1$. It is well known that the sample mean

$$\bar{U}_n = \bar{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$$

is a $\sqrt{n}-$consistent estimate of m. For notational simplicity, let $S_n = S = \sqrt{n}\bar{U}$. By the central limit theorem,

$$S \xrightarrow{d} N(0,1).$$

Equivalently, the latter result can be expressed in terms of the characteristic function of S, i.e.

$$E\left(e^{itS}\right) \to e^{-t^2/2}$$

as $n \to \infty$, where $e^{-t^2/2}$ is the characteristic function of $N(0,1)$ and $i = \sqrt{-1}$.

However, we might be interested in improving upon the approximation offered by the central limit theorem. Let κ_p be the $p-$th cumulant of U_i (throughout this thesis, κ_p will denote the $p-$th cumulant of various quantities and the reader will be reminded of this in each specific case in order to avoid notational confusion). It is known that the cumulant generating function of U_i, $\psi_{U_i}(t)$, can be written as an infinite series in κ_p, i.e.

$$\psi_{U_i}(t) = \sum_{p=1}^{\infty} \frac{(it)^p}{p!}\kappa_p.$$

Since $\kappa_1 = m = 0$ and $\kappa_2 = Var(U_i) = 1$, the latter expression becomes

$$\psi_{U_i}(t) = -\frac{1}{2}t^2 + \sum_{p=3}^{\infty} \frac{(it)^p}{p!}\kappa_p. \tag{1.3.1}$$

From (1.3.1), the cumulant generating function of S, $\psi_S(t)$, can be derived as

$$\psi_S(t) = n\psi_{U_i}(t/\sqrt{n}) = -\frac{1}{2}t^2 + \sum_{p=3}^{\infty} \frac{(it)^p}{n^{\frac{p}{2}-1}p!}\kappa_p. \tag{1.3.2}$$

From (1.3.2), it is clear that the normalized cumulants of S,

$$n^{-\frac{p}{2}+1}\kappa_p \quad \text{for } p \geq 3,$$

are decreasing in p. Hence, from (1.3.2),

$$\begin{aligned}
E\left(e^{itS}\right) = e^{\psi_S(t)} &= e^{-\frac{1}{2}t^2+\sum\limits_{p=3}^{\infty}\frac{(it)^p}{n^{\frac{p}{2}-1}p!}\kappa_p} \\
&= e^{-\frac{1}{2}t^2}\left(1+\sum_{p=3}^{\infty}\frac{(it)^p}{n^{\frac{p}{2}-1}p!}\kappa_p+\frac{1}{2}\left(\sum_{p=3}^{\infty}\frac{(it)^p}{n^{\frac{p}{2}-1}p!}\kappa_p\right)^2+\ldots\ldots\right) \\
&= e^{-\frac{1}{2}t^2}\left(1+\frac{1}{\sqrt{n}}\frac{(it)^3}{6}\kappa_3+\frac{1}{n}\left(\frac{(it)^4}{24}\kappa_4+\frac{(it)^6}{72}\kappa_3^2\right)+\ldots\ldots\right) \\
&= e^{-\frac{1}{2}t^2}\left(1+\frac{1}{\sqrt{n}}R_1(it)+\frac{1}{n}R_2(it)+\ldots\ldots\right), \qquad (1.3.3)
\end{aligned}$$

where

$$R_1(it) = \frac{(it)^3}{6}\kappa_3,$$

$$R_2(it) = \frac{(it)^4}{24}\kappa_4+\frac{(it)^6}{72}\kappa_3^2$$

and so on.

Since

$$e^{-\frac{1}{2}t^2} = \int_{\Re} e^{itx}d\Phi(x), \qquad (1.3.4)$$

$\Phi(x)$ being the standard normal cdf, (1.3.3) suggests the "inverse" expansion

$$Pr(S \leq x) = \Phi(x)+\frac{1}{\sqrt{n}}P_1(x)+\frac{1}{n}P_2(x)+\ldots. \qquad (1.3.5)$$

where $P_j(x)$ denotes a function whose Fourier-Stieltjes transform is $R_j(it)e^{-t^2/2}$, i.e.

$$\int_{\Re} e^{itx}dP_j(x) = R_j(it)e^{-t^2/2}.$$

By repeated integration by parts of (1.3.4),

$$P_j(x) = R_j(-d/dx)\Phi(x), \qquad (1.3.6)$$

where d/dx denotes the differential operator and $R_j(d/dx)$ should be interpreted as a polynomial in d/dx. For notational compactness, throughout we denote $g^{(i)}(x)$ the $i-$th derivative of a function g. From (1.3.6), (1.3.5) becomes

$$Pr(S \leq x) = \Phi(x)-\frac{1}{\sqrt{n}}\frac{1}{6}\kappa_3\Phi^{(3)}(x)+\frac{1}{n}\left(\frac{1}{24}\kappa_4\Phi^{(4)}(x)+\frac{1}{72}\kappa_3^2\Phi^{(6)}(x)\right)+\ldots. \qquad (1.3.7)$$

The latter is called Edgeworth expansion of the cdf of S.

Generally, (1.3.7) is truncated after a certain number of terms, e.g.

$$Pr(S \leq x) = \Phi(x) - \frac{1}{\sqrt{n}}\frac{1}{6}\kappa_3\Phi^{(3)}(x) + \frac{1}{n}\left(\frac{1}{24}\kappa_4\Phi^{(4)}(x) + \frac{1}{72}\kappa_3^2\Phi^{(6)}(x)\right) + O\left(\frac{1}{n^{3/2}}\right), \tag{1.3.8}$$

where the order of the remainder is conjectured from the rate and the parity of the coefficients. Such argument is purely formal. It is possible to prove validity of (1.3.8) by deriving the order of the remainder uniformly for all x. Starting from the work developed by Cramér (1946), Sargan (1976) and Bhattacharya and Ghosh (1978), among others, provided rigorous theory for validity of formal Edgeworth expansions. A seminal book-length account of Edgeworth expansions and rigorous results for validity issues is Bhattacharya and Rao (1976). In this thesis, we rely on formal Edgeworth expansions and validity proofs are left for future work.

As anticipated at the beginning of this section, this thesis will mainly deal with quadratic statistics in normal random variables and hence a short digression on how to extend the derivation of the Edgeworth expansion described above for the cdf of S to quadratic forms in normal random variables is worthwhile here. The characteristic function of a quadratic form $\epsilon' C\epsilon$, where the elements of ϵ are iid, normally distributed with mean zero and variance σ^2 and C is a $n \times n$ symmetric matrix, can be analytically evaluated by Gaussian integration as

$$\begin{aligned} E(e^{it\epsilon' C\epsilon}) &= \frac{1}{(2\pi\sigma^2)^{n/2}}\int_{\Re} e^{-\frac{\xi'\xi}{2\sigma^2}} e^{it\xi' C\xi} d\xi \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\int_{\Re} e^{-\frac{\xi'(I-2itC)\xi}{2\sigma^2}} d\xi = det(I-2itC)^{-1/2}. \end{aligned} \tag{1.3.9}$$

From (1.3.9), the cumulant generating function and hence the cumulants of $\epsilon' C\epsilon$ can be derived. Once such cumulants are known, after suitable algebraical manipulation, we can write an expansion for the characteristic function of a standardized version of $\epsilon' C\epsilon$, similar to one given in (1.3.3), and hence derive the corresponding Edgeworth expansion for the cdf. The derivation of Edgeworth expansions for the cdf of quadratic forms in normal random variables will be discussed thoroughly in the proofs of Theorems of Chapters 2-5.

It is clear that (1.3.8) (or a similar expansion for the statistic of interest) provides a more accurate approximation of the cdf of S (or of the cdf of the statistic of interest) than one offered by the central limit theorem. Also, (1.3.8) can be used to derive a better approximation for the quantiles of the cdf of S than one based on the quantiles of the standard normal. Alternatively, starting from (1.3.8), it is possible to derive a transformed statistic $g(S)$ so that its cdf is closer to the normal than that of S. Such results are very useful to derive improved testing procedures, since the former gives more accurate critical values than ones commonly used in first order theory, while the latter provides improved test statistics under the null hypothesis. Obviously, these results extend to the case of statistics other than S and the derivation of such Edgeworth-corrected critical values and Edgeworth-corrected test statistics will be shown and discussed in detail in this thesis in case of quadratic test statistics in normal random variables.

Several authors have applied Edgeworth expansions to derive refined test statistics in several context, starting from the work on the inverse of Edgeworth expansions by Cornish and Fisher (1937, 1960). Among these, Konishi et al (1988) derive the Edgeworth expansion and Edgeworth-corrected quantiles for the cdf of quadratic forms in normal random variables. Taniguchi (1986, 1988, 1991a, 1991b), derives higher order asymptotic properties of test statistics for time series data. Other relevant examples of derivation of refined test statistics in time series contexts include Magee (1989) and Kakizawa (1999). More specifically, Magee (1989) develops Edgeworth-corrected tests for linear restrictions when the data follow a linear regression with serially correlated disturbances. Also, Kakizawa (1999), starting from the earlier work by Ochi (1983), derives a valid Edgeworth expansion for the cdf of two different estimates of the correlation parameter in first order autoregressions and hence provides Edgeworth-corrected confidence intervals. In a different context, Phillips and Park (1988) derive Edgeworth-corrected Wald tests of nonlinar restrictions.

Before concluding this account on Edgeworth expansions, we should mention that Hall (1992) offers a very useful monograph which gives a view of the theory on Edgeworth expansions and Edgeworth-corrected tests in order to explain the performance of bootstrap methods. Indeed, it is well established, starting from the work by Singh (1981), that the bootstrap is a technique that can be used instead of the analytical derivation of Edgeworth expansions to improve upon the approximation offered by the central limit theorem. Indeed, Singh (1981) shows that the bootstrap automatically corrects for the first term after the normal cdf in an Edgeworth expansion (e.g. the second term at the RHS of (1.3.8)).

Although we do not aim to show theoretically the equivalence between the first Edgeworth correction and the bootstrap, in this thesis we compare by Monte Carlo the practical performance of Edgeworth corrections with bootstrap-based procedures and more specific references to relevant bootstrap literature will be given in Chapters 2-5.

1.4 Finite sample issues and contribution of this thesis

In Section 1.2 we provided an account of existing tests for (1.2.1) in SAR models, while in Section 1.3 we introduced the Edgeworth expansions and briefly discussed how they can be used to derive improved tests. Indeed, the main scope of this thesis is to derive refined tests for (1.2.1) in SAR models based on formal Edgeworth expansions. Specifically, in Chapters 2 and 3 we will derive Edgeworth-corrected tests of (1.2.1) in (1.2.5) based on $\hat{\lambda}$ and $\tilde{\lambda}$, respectively. The refined tests derived in Chapters 2 and 3 will also be applied in two small empirical examples. In Chapter 4 we will derive improved LM tests of (1.2.1) in (1.2.5) and (1.2.15). In Chapter 5 we will derive Edgeworth-corrected tests for (1.2.1) in (1.1.3) based on $\hat{\lambda}$ and $\hat{\lambda}_{IV}$. Although testing for spatial independence is the main focus of this work, in Chapter 5 improved tests for linear restrictions on β in (1.1.3) based on $\hat{\beta}_{IV}$ are also derived. Small sample performance of Edgeworth-corrected tests are assessed by Monte Carlo and compared with bootstrap-based procedures.

This thesis is motivated by the fact that although the literature on testing for spatial independence is very broad, analytical derivation of finite sample corrections for such tests has received little attention, other than the aforementioned contribution by Robinson (2008b). This issue is of particular concern in spatial econometrics since datasets are usually small/moderately-sized, as very often the practitioner is interested in estimating and testing spatial coefficients of SAR models when data are recorded across cities, regions of countries. For instance, the two empirical examples considered in this thesis deal with spatial correlation of variables recorded across 43 European countries and 103 Italian regions, respectively (hence $n = 43$ and $n = 103$, respectively). When n is small/moderately-sized, testing procedures based on the normal (or χ^2) approximation for the distribution of test statistics might be seriously unreliable.

Together with the likely limited sample size, another source of concern for the reliability of standard testing procedures in SAR models is given by the possibly slow rate of convergence of $\hat{\lambda}$ and $\tilde{\lambda}$ when the data follow (1.2.5), as outlined in Section 1.2. When this is the case, the cdf of statistics based on such estimates is poorly approximated by a normal and finite sample corrections are indeed crucial in order to obtain reliable tests. This issue provide even stronger motivation for the new tests presented in Chapters 2 and 3.

Although the spatial econometrics literature on analytical finite sample corrections is very limited, small sample performance of estimates of the parameters in (1.1.3), (1.2.2) and (1.2.15) and corresponding tests have been assessed quite extensively by Monte Carlo studies, see e.g. Anselin and Rey (1991), Anselin and Florax (1995), Das et al. (2003) and, more recently, Egger et al. (2009). More specifically, Anselin and Rey (1991) and Anselin and Florax (1995) report and discuss broad sets of Monte Carlo results to evaluate the practical performances of various existing tests for spatial independence. Das et al. (2003) perform a Monte Carlo study to assess the finite sample behaviour of IV-type of estimates of parameters in (1.2.2), while Egger et al. (2009) propose a similar analysis for Wald-type of tests of (1.2.1) in SAR models based on MLE and Generalized Method of Moments estimates.

Before concluding this section, we should acknowledge that an analytical procedure that attempts to derive Edgeworth-based corrections for the cdf of $\tilde{\lambda}$ in (1.2.5) has been derived by Bao and Ullah (2007). Using a stochastic expansion of the score function, they derive the second order bias and mean squared error of $\tilde{\lambda}$ in (1.2.5). However, Bao and Ullah (2007) do not outline the possibly slow rate of convergence of $\tilde{\lambda}$ in (1.2.5) and do not consider improved tests.

1.5 Some definitions and Assumptions

We first define some notation that will be used throughout and has not been introduced previously.

i) $\phi(z)$ denotes the the probability density function (pdf) of a standard normal random variable.

ii) $1(.)$ indicates the indicator function, i.e. $1(A \subset B) = 1$ if $A \subset B$ and $1(A \subset B) = 0$ otherwise.

iii) $H_j(x)$ denotes the $j-$th Hermite polynomial, i.e.

$$H_j(x) = (-1)^j e^{x^2/2} \frac{d^j}{dx^j} e^{-x^2/2}, \tag{1.5.1}$$

e.g.

$$H_1(x) = x, \quad H_2(x) = x^2 - 1 \quad \text{and} \quad H_3(x) = x^3 - 3x.$$

iv) $\sim$ denotes an exact rate, i.e. $a \sim b$ means that $|a/b|$ converges to a positive, finite limit.

v) $\eta_i(A)$, $i = 1, ...q$ denotes the eigenvalues of a generic $q \times q$ matrix A.

vi)

$$\bar{\eta}(A) = \max_{i=1,....q} \{|\eta_i(A)|\}. \tag{1.5.2}$$

vii)

$$\underline{\eta}(A) = \min_{i=1,....q} \{|\eta_i(A)|\}.$$

viii) $||.||$ indicates the spectral norm, i.e. for any $p \times q$ matrix B

$$||B||^2 = \bar{\eta}(B'B).$$

ix) $||A||_r$ denotes the maximum row sum matrix norm, i.e.

$$||A||_r = \max_i \sum_{j=1}^{q} |a_{ij}|,$$

a_{ij} being the (i, j)th element of A.

x) $||A||_c$ indicates the maximum column sum matrix norm, i.e.

$$||A||_c = \max_j \sum_{i=1}^{q} |a_{ij}| \, .$$

We introduce some assumptions, which are common to Chapters 2-5, while other relevant model-specific conditions are left to each chapter.

Assumption 1 *The elements of ϵ are independent and identically distributed normal random variables with mean zero and unknown variance σ^2.*

Assumption 2

(i) *For all n, $w_{ii} = 0$, $\Sigma_{j=1}^{n} w_{ij} = 1$, $i = 1, ..., n$, and $||W|| = 1$.*

(ii) *For all n, W is uniformly bounded in row and column sums in absolute value, i.e.*

$$||W||_r + ||W||_c \leq K,$$

where K is a finite generic constant.

(iii) *Uniformly in $i, j = 1, ..., n$, $w_{ij} = O(1/h)$, where $h = h_n$ is bounded away from zero for all n and $h/n \to 0$ as $n \to \infty$.*

As is common in much higher order literature, Gaussianity is assumed in this derivation. Assumption 1 can be relaxed at expense of considerable extra complications in the derivation of Edgeworth expansions. We should stress that Assumption 1 is not necessary for (1.2.3), (1.2.4), (1.2.12), (1.2.13), (1.2.14) and (1.2.18) which indeed hold more generally. However, in this thesis we are interested in deriving statistics with better finite-sample properties and we can only justify these under the precise distributional assumption.

The normalizations in Assumption 2(i) are not strictly necessary for the proofs of the results presented in Chapters 1-4, but they play a role in constructing the likelihood. Furthermore, Assumption 2(i) or some other normalization is required for identification when $\lambda \neq 0$. Assumption 2(i) requires that W is row normalized so that the elements in each row sum to one. It also imposes that the maximum eigenvalue of $W'W$ equals one. It should be stressed that actually $||W|| = 1$ would be sufficient for identification purposes and hence row normalization seems somehow redundant. However, the latter entails significant algebraical simplifications in the derivation of some of the results presented in this thesis (see e.g. Section 2.3) and is therefore retained. In general, $\bar{\eta}(W) \leq ||W||$ (see Horn and Johnson (1985), page 297) and, by Assumption 2(i), $||W|| = 1$. Since 1 is an eigenvalue of W when the latter is row normalized, we can conclude $\bar{\eta}(W) = 1$.

Assumption 2(ii) is standard. It has been introduced by Kelejian and Prucha (1998) to keep the spatial correlation to a manageable degree. Assumption 2(ii) plays an important role both in the proofs of the first order theory results, i.e. (1.2.3), (1.2.4), (1.2.12), (1.2.13), (1.2.14) and (1.2.18) and in the proofs of the results of this thesis. It is also worth mentioning that row normalization and non negative w_{ij}, for all $i, j = 1, ..., n$, implies $||W||_r = 1$.

Assumption 2(iii) formalises the definition and some general conditions on the sequence h already introduced in Section (1.2). Indeed, Assumption 2(iii) is the same as one in Lee (2004). As already outlined in Section 1.2, h can be bounded or divergent. A condition on w_{ij} is commonly required in asymptotic theory for statistics based on SAR models. Assumption 2(iii) will be suitably strengthened in Chapter 5, consistently with one in Lee (2002), as discussed in Section 1.2.

We should notice that W in (1.1.2) can satisfy Assumption 2. In this case $h = m - 1$. It is straightforward to verify that Assumptions 2(i)-(ii) are satisfied for this choice of W, whether h is bounded or divergent (that is, whether the number of households in each unit diverges or is bounded as n increases). Assumption 2(iii) holds provided that $r \to \infty$, m being either divergent or bounded.

2 Improved OLS Test Statistics for Pure SAR

Throughout most of this chapter we assume that for some scalar $\lambda \in (-1, 1)$ the data follow (1.2.5), which is model (1.1.3) when $\beta = 0$ a priori, i.e. none of the exogenous regressors is relevant, and we are interested in testing (1.2.1) when λ is estimated by OLS. An extension of the proposed procedures to the pure SAR with intercept term is also considered.

As discussed in Chapter 1, it is known (Lee (2002)) that $\hat{\lambda}$ in model (1.2.5) is inconsistent when $\lambda \neq 0$. However, it converges to zero in probability under (1.2.1) and although this case is limited when the interest is estimation, it is relevant in testing and should be investigated further. As anticipated in Chapter 1, in this chapter we show that, under H_0 in (1.2.1), the rate of convergence of $\hat{\lambda}$ might be slower than the parametric $\sqrt{n}$, depending on assumptions on W.

When the rate of convergence of $\hat{\lambda}$ is slower than $\sqrt{n}$, the cdf of the t-statistic based on $\hat{\lambda}$ under (1.2.1) is not accurately approximated by a normal. Our new tests are based on refined t-statistics, whose cdf are closer to the normal than those of the standard statistics and therefore entail better approximations. Alternatively, we show that inference based on standard statistics can be improved by considering more accurate approximations for critical values than ones of the normal cdf.

This chapter is organised as follows. In Sections 2.1 and 2.2 we present refined tests for (1.2.1) against one-sided and two-sided alternatives, respectively. In Section 2.3, we show that the results of Sections 2.1 and 2.2 can be easily extended when model (1.2.5) contains a location parameter. In Section 2.4 we present some results for the power of the test of (1.2.1) against a local alternative. In Section 2.5 we report and discuss the results of some Monte Carlo simulations of the tests presented in Sections 2.1-2.4. Relevant proofs are left to appendices.

2.1 Test against a one-sided alternative: Edgeworth-corrected critical values and corrected statistic

We suppose that model (1.2.5) holds and we are interested in testing (1.2.1) against a one-sided alternative

$$H_1 : \lambda > 0 \quad (< 0). \tag{2.1.1}$$

As previously mentioned, $\hat{\lambda}$ in (1.2.6) converges in probability to zero under H_0, as shown by a straightforward modification of Lemma 2.3 reported in the Appendix.

Let Assumptions 1-2 hold and in addition:

Assumption 3 *The limits*

$$\lim_{n\to\infty}\frac{h}{n}tr(W'W),\qquad \lim_{n\to\infty}\frac{h}{n}tr(WW'W),\qquad \lim_{n\to\infty}\frac{h}{n}tr((W'W)^2),$$
$$\lim_{n\to\infty}\frac{h}{n}tr(W^2),\qquad \lim_{n\to\infty}\frac{h}{n}tr(W^3) \tag{2.1.2}$$

are non-zero.

Under Assumption 2 the limits displayed in (2.1.2) exist and are finite by Lemma 2.1. Thus, the content of Assumption 3 is that such limits are also non-zero. Let ζ be any finite real number.

Theorem 2.1 *Let model (1.2.5) and Assumptions 1-3 hold. The cdf of $\hat{\lambda}$ under H_0 in (1.2.1) admits the third order formal Edgeworth expansion*

$$\begin{aligned}\Pr(a\hat{\lambda}\leq\zeta|H_0) &= \Phi(\zeta)+2b\zeta^2\phi(\zeta)-\frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta)-\left(\frac{tr((W'W)^2)}{(tr(W'W))^2}-6b^2\right)\zeta^3\phi(\zeta)\\ &+ 2b^2\zeta^4\Phi^{(2)}(\zeta)-\frac{\kappa_3^c}{3}b\zeta^2\Phi^{(4)}(\zeta)+\frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta)+O\left(\left(\frac{h}{n}\right)^{3/2}\right),\end{aligned} \tag{2.1.3}$$

where

$$a=\frac{tr(W'W)}{(tr(W'W+W^2))^{1/2}},\qquad b=\frac{tr(WW'W)}{(tr(W'W+W^2))^{1/2}tr(W'W)}, \tag{2.1.4}$$

$$\kappa_3^c\sim\frac{2tr(W^3)+6tr(W'W^2)}{(tr(W'W+W^2))^{3/2}} \tag{2.1.5}$$

and

$$\kappa_4^c\sim\frac{6tr(W^4)+24tr(W'W^3)+12tr((WW')^2)+6tr(W^2W'^2)}{(tr(W'W+W^2))^2}. \tag{2.1.6}$$

The proof of Theorem 2.1 is in the Appendix.

Under Assumption 3, as $n\to\infty$

$$b\sim\left(\frac{h}{n}\right)^{1/2},\quad \frac{tr((W'W)^2)}{(tr(W'W))^2}\sim\frac{h}{n},\quad \kappa_3^c\sim\left(\frac{h}{n}\right)^{1/2},\quad \kappa_4^c\sim\frac{h}{n}$$

and therefore

$$2b\zeta^2\phi(\zeta)-\frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta)\sim\left(\frac{h}{n}\right)^{1/2},$$

$$-\left(\frac{tr((W'W)^2)}{(tr(W'W))^2}-6b^2\right)\zeta^3\phi(\zeta)+2b^2\zeta^4\Phi^{(2)}(\zeta)-\frac{\kappa_3^c}{3}b\zeta^2\Phi^{(4)}(\zeta)+\frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta)\sim\frac{h}{n}.$$

Since $a \sim (n/h)^{1/2}$ from Assumption 3, when the sequence h is divergent the rate of convergence of $Pr(a\hat{\lambda} \leq \zeta|H_0)$ to the standard normal cdf is obviously slower than the usual $\sqrt{n}$. It must be stressed that the expansion in (2.1.3) is formal and hence the order of the remainder can only be conjectured by the rate of the coefficients.

As anticipated in Chapter 1, from the expansion (2.1.3) Edgeworth-corrected critical values can be obtained. We denote w_α and z_α the $\alpha-$quantiles of the null statistic $a\hat{\lambda}$ and the standard normal cdf, respectively. By inversion of (2.1.3) we can obtain an infinite series for w_α, i.e.

$$w_\alpha = z_\alpha + p_1(z_\alpha) + p_2(z_\alpha) + ..., \tag{2.1.7}$$

where $p_1(z_\alpha)$ and $p_2(z_\alpha)$ are polynomials of orders $(h/n)^{1/2}$ and h/n, respectively. Both $p_1(z_\alpha)$ and $p_2(z_\alpha)$ can be determined using the identity $\alpha = Pr(a\hat{\lambda} \leq w_\alpha|H_0)$ and the asymptotic expansion given in Theorem 2.1. Even though the procedure can be extended to higher orders, for algebraic simplicity we focus on the second order Edgeworth correction and therefore only $p_1(z_\alpha)$ has to be determined.

For convenience, we report the second order Edgeworth expansion

$$\Pr(a\hat{\lambda} \leq \zeta|H_0) = \Phi(\zeta) + 2b\zeta^2\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + O\left(\frac{h}{n}\right). \tag{2.1.8}$$

From (2.1.8) and the property (derived from (1.5.1))

$$(-d/dx)^j\Phi(x) = -H_{j-1}(x)\phi(x), \tag{2.1.9}$$

we have

$$\alpha = \Pr(a\hat{\lambda} \leq w_\alpha|H_0) = \Phi(w_\alpha) - (\frac{\kappa_3^c}{3!}H_2(w_\alpha) - 2bw_\alpha^2)\phi(w_\alpha) + O\left(\frac{h}{n}\right).$$

Moreover, expanding w_α according to (2.1.7) and dropping negligible terms,

$$\begin{aligned} \alpha &= \Pr(a\hat{\lambda} \leq w_\alpha|H_0) \\ &= \Phi(z_\alpha) + p_1(z_\alpha)\phi(z_\alpha) - (\frac{\kappa_3^c}{3!}H_2(z_\alpha) - 2bz_\alpha^2)\phi(z_\alpha) + O\left(\frac{h}{n}\right) \\ &= \alpha + p_1(z_\alpha)\phi(z_\alpha) - (\frac{\kappa_3^c}{3!}H_2(z_\alpha) - 2bz_\alpha^2)\phi(z_\alpha) + O\left(\frac{h}{n}\right), \end{aligned} \tag{2.1.10}$$

where the second equality follows by Taylor expansion of $\Phi(w_\alpha)$ around z_α. The last displayed identity holds up to order $O(h/n)$ when

$$p_1(z_\alpha) = \frac{\kappa_3^c}{3!}H_2(z_\alpha) - 2bz_\alpha^2.$$

Hence (2.1.7) becomes

$$w_\alpha = z_\alpha + \frac{\kappa_3^c}{3!} H_2(z_\alpha) - 2bz_\alpha^2 + O\left(\frac{h}{n}\right). \tag{2.1.11}$$

The size of the test of (1.2.1) obtained with the usual approximation of w_α by z_α, that is

$$Pr(a\hat{\lambda} > z_\alpha | H_0), \tag{2.1.12}$$

can be compared with the one obtained using the Edgeworth-corrected quantile as given in (2.1.11), i.e.

$$Pr(a\hat{\lambda} > z_\alpha + \frac{\kappa_3^c}{3!} H_2(z_\alpha) - 2bz_\alpha^2 | H_0). \tag{2.1.13}$$

When z_α is used to approximate w_α, the error has order $(h/n)^{1/2}$, while it is reduced to (h/n) when the Edgeworth-corrected critical value is used.

Rather than corrected critical values, an Edgeworth-corrected test statistic can be derived. By (2.1.9) and since $H_2(\zeta) = \zeta^2 - 1$, (2.1.8) can be written as

$$\Pr(a\hat{\lambda} \leq \zeta | H_0) = \Phi(\zeta + 2b\zeta^2 - \frac{\kappa_3^c}{3!}(\zeta^2 - 1)) + O\left(\frac{h}{n}\right).$$

When the transformation

$$v(\zeta) = \zeta + 2b\zeta^2 - \frac{\kappa_3^c}{3!}(\zeta^2 - 1) = \zeta + (2b - \frac{\kappa_3^c}{3!})\zeta^2 + \frac{\kappa_3^c}{3!}$$

is monotonic, we can write

$$\Pr(a\hat{\lambda} + (2b - \frac{\kappa_3^c}{3!})(a\hat{\lambda})^2 + \frac{\kappa_3^c}{3!} \leq \zeta) = \Phi(\zeta) + O\left(\frac{h}{n}\right)$$

and make inference on λ based on the corrected statistic $v(a\hat{\lambda})$. The function $v(\zeta)$ is strictly increasing when $\zeta > -1/(2(2b - \kappa_3^c/3!))$, however the latter does not hold in general and therefore a cubic transformation that does not affect the remainder but such that the resulting function is strictly increasing over the whole domain should be considered. A suitable transformation is in Hall (1992) or, in a more general case, Yanagihara et al (2005):

$$g(\zeta) = v(\zeta) + Q(\zeta), \qquad \text{with} \qquad Q(\zeta) = \frac{1}{3}\left(2b - \frac{\kappa_3^c}{3!}\right)^2 \zeta^3.$$

Indeed, it can be shown (Yanagihara et al (2005)) that for a statistic T that admits the general expansion

$$Pr(T \leq x) = \Phi(x) + p_1(x)\phi(x) + O\left(\frac{h}{n}\right),$$

where $p_1(x) \sim \sqrt{h/n}$, the transformation

$$g(x) = x + p_1(x) + \frac{1}{4}Q(x) \quad \text{with} \quad Q(x) = \int \left(\frac{d}{dx}p_1(x)\right)^2 dx \tag{2.1.14}$$

is strictly increasing and does not affect higher order terms, i.e.

$$Pr(g(T) \leq x) = \Phi(x) + O\left(\frac{h}{n}\right).$$

It is straightforward to verify that in the present case the function $g(\zeta)$ is strictly increasing for every ζ, its first derivative being $(1 + (2b - (\kappa_3^c/3!)\zeta))^2$.

The size of the test of (1.2.1) based on such corrected statistic,

$$Pr(g(a\hat{\lambda}) > z_\alpha | H_0), \tag{2.1.15}$$

can be compared with the standard (2.1.12). As previously mentioned, the error when the standard statistic is used has order $\sqrt{h/n}$, while it is reduced to h/n when considering the corrected variant.

2.2 Test against a two-sided alternative: Edgeworth-corrected critical values and corrected statistic

In Section 2.1 we focused on testing (1.2.1) against (2.1.1). However, in some circumstances the practitioner might not have a prior conjecture about the sign of λ in (1.2.5) and a test of (1.2.1) against a two-sided alternative may be more suitable. In this section we propose refined tests for (1.2.1) against a two-sided alternative

$$H_1 : \lambda \neq 0. \tag{2.2.1}$$

From Theorem 2.1, (2.1.9) and

$$\phi(-\zeta) = \phi(\zeta), \quad \Phi^{(2)}(-\zeta) = -\Phi^{(2)}(\zeta), \quad \Phi^{(3)}(-\zeta) = \Phi^{(3)}(\zeta), \quad \Phi^{(4)}(-\zeta) = -\Phi^{(4)}(\zeta),$$

we obtain

$$\begin{aligned} Pr(|a\hat{\lambda}| \leq \zeta | H_0) &= Pr(a\hat{\lambda} \leq \zeta) - Pr(a\hat{\lambda} \leq -\zeta) \\ &= \Phi(\zeta) - \Phi(-\zeta) - 2\left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right)\zeta^3\phi(\zeta) + 4b^2\zeta^4\Phi^{(2)}(\zeta) \\ &- 2\frac{\kappa_3^c}{3}b\zeta^2\Phi^{(4)}(\zeta) + 2\frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + O\left(\left(\frac{h}{n}\right)^2\right) \\ &= 2\Phi(\zeta) - 1 + \{-2\left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right)\zeta^3 - 4b^2\zeta^4 H_1(\zeta) \\ &+ 2\frac{\kappa_3^c}{3}b\zeta^2 H_3(\zeta) - \frac{\kappa_4^c}{12}H_3(\zeta)\}\phi(\zeta) + O\left(\left(\frac{h}{n}\right)^2\right). \end{aligned} \tag{2.2.2}$$

Under Assumption 3 the terms in braces of the last displayed expansion have order h/n, while, as previously mentioned, the order of the remainder is conjectured by the rate of the coefficients and the parity of the expansion.

As discussed in Section 2.1, Edgeworth-corrected critical values and corrected null statistics can be derived from (2.2.2). Let q_α be the $\alpha-$quantile of the null statistic $|a\hat{\lambda}|$. From (2.2.2),

$$\begin{aligned}
\alpha &= \Pr(|a\hat{\lambda}| \le q_\alpha | H_0) \\
&= 2\Phi(q_\alpha) - 1 - 2\left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right) q_\alpha^3 \phi(q_\alpha) - 4b^2 q_\alpha^4 H_1(q_\alpha)\phi(q_\alpha) \\
&+ 2\frac{\kappa_3^c}{3} b q_\alpha^2 H_3(q_\alpha)\phi(q_\alpha) - \frac{\kappa_4^c}{12} H_3(q_\alpha)\phi(q_\alpha) + O\left(\left(\frac{h}{n}\right)^2\right)
\end{aligned}$$

and therefore

$$\begin{aligned}
\frac{\alpha+1}{2} &= \Phi(q_\alpha) - \left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right) q_\alpha^3 \phi(q_\alpha) - 2b^2 q_\alpha^4 H_1(q_\alpha)\phi(q_\alpha) \\
&+ \frac{\kappa_3^c}{3} b q_\alpha^2 H_3(q_\alpha)\phi(q_\alpha) - \frac{\kappa_4^c}{4!} H_3(q_\alpha)\phi(q_\alpha) + O\left(\left(\frac{h}{n}\right)^2\right). \qquad (2.2.3)
\end{aligned}$$

Correspondingly, an infinite series for q_α in terms of $z_{(\alpha+1)/2}$ can be written as

$$q_\alpha = z_{\frac{\alpha+1}{2}} + p_1(z_{\frac{\alpha+1}{2}}) + O\left(\left(\frac{h}{n}\right)^2\right). \qquad (2.2.4)$$

Similarly to the case presented in Section 2.1, the size of the test of (1.2.1) against a two-sided alternative when q_α is approximated by $z_{(\alpha+1)/2}$ can be compared with that obtained when q_α is approximated by the Edgeworth-corrected quantity $z_{(\alpha+1)/2} + p_1(z_{(\alpha+1)/2})$. The error of the latter approximation is reduced to $O((h/n)^2)$. The polynomial $p_1(z_{(\alpha+1)/2})$ can be determined by substituting (2.2.4) into (2.2.3) and dropping negligible terms, i.e.

$$\begin{aligned}
\frac{\alpha+1}{2} &= \Phi(z_{(\alpha+1)/2} + p_1(z_{(\alpha+1)/2})) - \left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right) z_{(\alpha+1)/2}^3 \phi(z_{(\alpha+1)/2}) \\
&- 2b^2 z_{(\alpha+1)/2}^4 H_1(z_{(\alpha+1)/2})\phi(z_{(\alpha+1)/2}) + \frac{\kappa_3^c}{3} b z_{(\alpha+1)/2}^2 H_3(z_{(\alpha+1)/2})\phi(z_{(\alpha+1)/2}) \\
&- \frac{\kappa_4^c}{4!} H_3(z_{(\alpha+1)/2})\phi(z_{(\alpha+1)/2}) + O\left(\left(\frac{h}{n}\right)^2\right).
\end{aligned}$$

Hence, by Taylor expansion,

$$\frac{\alpha+1}{2} = \Phi(z_{(\alpha+1)/2}) + p_1\phi(z_{(\alpha+1)/2}) - \left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right) z_{(\alpha+1)/2}^3 \phi(z_{(\alpha+1)/2})$$

$$
\begin{aligned}
&- \quad 2b^2 z^4_{(\alpha+1)/2} H_1(z_{(\alpha+1)/2})\phi(z_{(\alpha+1)/2}) + \frac{\kappa^c_3}{3} b z^2_{(\alpha+1)/2} H_3(z_{(\alpha+1)/2})\phi(z_{(\alpha+1)/2}) \\
&- \quad \frac{\kappa^c_4}{4!} H_3(z_{(\alpha+1)/2})\phi(z_{(\alpha+1)/2}) + O\left(\left(\frac{h}{n}\right)^2\right).
\end{aligned}
$$

The last displayed identity holds up to order $O(h/n)^2$ if

$$
\begin{aligned}
p_1(z_{(\alpha+1)/2}) &= \left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right) z^3_{(\alpha+1)/2} + 2b^2 z^4_{(\alpha+1)/2} H_1(z_{(\alpha+1)/2}) \\
&- \frac{\kappa^c_3}{3} b z^2_{(\alpha+1)/2} H_3(z_{(\alpha+1)/2}) + \frac{\kappa^c_4}{4!} H_3(z_{(\alpha+1)/2}). \qquad (2.2.5)
\end{aligned}
$$

As discussed in Section 2.1, a corrected statistic under H_0 can also be derived from (2.2.2). Indeed, (2.2.2) can be written as

$$
\begin{aligned}
Pr(|a\hat{\lambda}| \leq \zeta) &= 2\Phi\left(\zeta - \left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right)\zeta^3 - 2b^2\zeta^4 H_1(\zeta) + \frac{\kappa^c_3}{3} b\zeta^2 H_3(\zeta) - \frac{\kappa^c_4}{4!} H_3(\zeta)\right) \\
&- 1 + O\left(\left(\frac{h}{n}\right)^2\right).
\end{aligned}
$$

By a straightforward modification of the procedure described in Section 2.1 (Yanagihara et al (2005)), we obtain

$$
Pr(v(|a\hat{\lambda}|) \leq \zeta | H_0) = 2\Phi(\zeta) - 1 + O\left(\left(\frac{h}{n}\right)^2\right),
$$

where

$$
v(x) = x - \left(\frac{tr((W'W)^2)}{(tr(W'W))^2} - 6b^2\right) x^3 - 2b^2 x^4 H_1(x) + \frac{\kappa^c_3}{3} b x^2 H_3(x) - \frac{\kappa^c_4}{4!} H_3(x). \quad (2.2.6)
$$

The error when the distribution of the corrected statistic under H_0 is approximated by the standard normal is reduced to $O((h/n)^2)$. As pointed out in Section 2.1, the latter result relies on the monotonicity (at least local) of $v(.)$. Because of the cumbersome functional form of the correction terms, in this case it is algebraically difficult to obtain the cubic transformation given in (2.1.14). Hence, we rely on some numerical work to assess whether $v(.)$ is indeed locally increasing and, eventually, implement numerically the cubic transformation in (2.1.14).

2.3 Corrected critical values and corrected statistic for pure SAR with a location parameter

In Sections 2.1 and 2.2 we considered model (1.2.5), which is a particular case of (1.1.3) where $\beta = 0$ a priori. In this section we extend the results derived in Section 2.1 to model

$$
Y = \mu l + \lambda W Y + \epsilon, \qquad (2.3.1)
$$

where μ denotes a scalar parameter. For simplicity we focus on one-sided test, but extensions of the results derived in Section 2.2 are also straightforward, at expense of extra algebraical burden.

Specifically, we obtain

Theorem 2.2 *Suppose that model (2.3.1) and Assumptions 1-3 hold. The cdf of $\hat{\lambda}$ under H_0 in (1.2.1) admits the second order formal Edgeworth expansion*

$$\begin{aligned} Pr(a\hat{\lambda} \le \zeta | H_0) &= \Phi(\zeta) + \left(\frac{1}{(tr(W^2 + W'W))^{1/2}} + 2b\zeta^2 \right) \phi(\zeta) \\ &- \frac{\kappa_3^c}{3!} \Phi^{(3)}(\zeta) + O\left(\frac{h}{n}\right), \end{aligned} \tag{2.3.2}$$

where a, b and κ_3^c have been defined in (2.1.4) and (2.1.5).

The proof of Theorem 2.2 is in the Appendix A.

From (2.3.2) corrected critical values and corrected statistics under H_0 can be obtained. The derivation is identical to one presented in Section 2.1 and is therefore omitted. Let w_α^l be the true $\alpha-$quantile of the cdf of $a\hat{\lambda}$ under H_0. From (2.3.2),

$$w_\alpha^l = z_\alpha - \left(\frac{1}{(tr(W^2 + W'W))^{1/2}} + 2bz_\alpha^2 \right) + \frac{\kappa_3^c}{3!} H_2(z_\alpha) + O\left(\frac{h}{n}\right).$$

Hence, as previously discussed, when z_α is used to approximate w_α^l, the error is $O((h/n)^{1/2})$, while it is reduced to $O(h/n)$ when the Edgeworth-corrected critical value is used.

Similarly, the corrected statistic under H_0 can be derived from (2.3.2). As discussed in Section 2.1, the transformation defined in (2.1.14) is strictly increasing and constructed so that the error obtained by approximating the cdf of the null transformed statistic with a normal is reduced to order h/n. In this case, from (2.3.2), (2.1.14) becomes

$$g(x) = x + \frac{1}{(tr(W^2 + W'W))^{1/2}} + 2bx^2 - \frac{\kappa_3^c}{3!}(x^2 - 1) + \frac{1}{3}\left(2b - \frac{\kappa_3^c}{3!}x^3\right).$$

2.4 Test against a local alternative

In this section we focus on testing (1.2.1) in model (1.2.5) against a local alternative hypothesis

$$H_1 : \lambda_n = c\left(\frac{h}{n}\right)^{1/2}, \quad c > 0 \ \ (< 0). \tag{2.4.1}$$

Although we previously specified that the subscript n would be omitted, we retain it in this case to stress the shrinking nature of the class of alternatives. For algebraic simplicity, the results in this section are derived assuming that W is symmetric. The extension to the case of non symmetric W is trivial, but algebraically more cumbersome. Without loss of generality, the following results are developed for $c > 0$ in (2.4.1). As already mentioned, when $\lambda \neq 0$ the OLS estimate of λ in (1.2.5) is inconsistent. However, under H_1 as defined in (2.4.1), $\hat{\lambda}_n$ converges in probability to zero, as shown in Lemma 2.3 reported in the Appendix. More specifically, by Lemma 2.3, $\hat{\lambda}_n = \lambda_n + O_p((h/n)^{1/2})$, i.e. the probability limit of $\hat{\lambda}_n - \lambda_n$ vanishes at least at fast as λ_n.

Under Assumption 2(ii), when W is symmetric, $||W||_r \leq K/2$ and $||W||_c \leq K/2$. The series representation

$$S^{-1}(\lambda_n) = \sum_{t=0}^{\infty} (\lambda_n W)^t \tag{2.4.2}$$

holds when $|||\lambda_n W||| < 1$, $|||.|||$ denoting any matrix norm (see e.g. Horn and Johnson (1985), page 301). Under H_1, using the spectral norm,

$$||\lambda_n W|| = c\left(\frac{h}{n}\right)^{1/2} ||W|| = c\left(\frac{h}{n}\right)^{1/2} < 1,$$

for $c < 1$ or n large enough.

Under H_1 in (2.4.1), $S^{-1}(\lambda_n)$ is also uniformly bounded in row sums in absolute value since

$$\begin{aligned} ||S^{-1}(\lambda_n)||_r &= ||\sum_{t=0}^{\infty}(\lambda_n W)^t||_r \leq \sum_{t=0}^{\infty}\lambda_n^t ||W^t||_r \leq \sum_{t=0}^{\infty}\lambda_n^t ||W||_r^t \\ &\leq \sum_{t=0}^{\infty}\left(\lambda_n \frac{K}{2}\right)^t = \frac{1}{1-\lambda_n \frac{K}{2}} < \infty \end{aligned}$$

for n large enough that $c(h/n)^{1/2}K/2 < 1$. Trivially, by symmetry of W, $S^{-1}(\lambda_n)$ is uniformly bounded in column sums in absolute value.

We obtain the following result

Theorem 2.3 *Suppose that model (1.2.5) and Assumptions 1-3 hold. The cdf of $\hat{\lambda}_n - \lambda_n$ under H_1 as defined in (2.4.1) admits the formal second order Edgeworth expansion*

$$Pr(a(\hat{\lambda}_n - \lambda_n) \leq \zeta | H_1) = \Phi(\zeta - \lambda_n a) - \omega(\zeta)\phi(\zeta - \lambda_n a) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta - \lambda_n a) + O\left(\frac{h}{n}\right), \tag{2.4.3}$$

where a and κ_3^c have been defined in (2.1.4) and (2.1.5), respectively, and

$$\omega(\zeta) = \frac{tr(W^3)}{2}(\lambda_n^2 a^{-1} - 2\lambda_n a^{-2}\zeta) + \frac{tr(W^3)}{4}(2\lambda_n a^{-2} - a^{-3}\zeta)(\zeta - a\lambda_n). \tag{2.4.4}$$

The proof of Theorem 2.3 is in the Appendix.

Since Theorem 2.3 has been proved for a symmetric W, a and κ_3^c can be simplified to $\sqrt{tr(W^2)}/\sqrt{2}$ and $8tr(W^3)/2\sqrt{2}tr(W^2)^{3/2}$, respectively. Under Assumption 3, $\omega(\zeta) \sim \sqrt{h/n}$ and $a\lambda_n$ has a positive limit. Hence, to a first approximation, under H_1, $a(\hat{\lambda}_n - \lambda_n)$ is normally distributed with mean $\lambda_n a$ and unit variance. It is straightforward to notice that when $\lambda_n = 0$ we recover the expansion given in (2.1.8). Intuitively, the term $a\lambda_n$ is a large sample bias that vanishes only when $\lambda_n = 0$ (or $\lambda_n = O((h/n)^\gamma)$, with $\gamma > 1/2$).

A very simple, straightforward result that can be derived using the expansion in Theorem 2.3 consists in the possibility of providing a better approximation of the (local) power of the test of (1.2.1) against (2.4.1) based on the statistic $a\hat{\lambda}_n$ than that given by the usual first order theory. Specifically, suppose H_0 is rejected when $a\hat{\lambda}_n > \tau$. The power of such test (as function of τ), denoted as $\Pi(\tau)$ henceforth, is defined as

$$\Pi(\tau) = Pr(a\hat{\lambda}_n > \tau | H_1) = 1 - Pr(a\hat{\lambda}_n \leq \tau | H_1) = 1 - Pr(a(\hat{\lambda}_n - \lambda) \leq \tau - a\lambda_n | H_1).$$

Obviously, $Pr(a(\hat{\lambda}_n - \lambda_n) \leq \tau - a\lambda_n | H_1)$ is unknown, but Theorem 2.3 can be used to obtain a more accurate approximation for $\Pi(\tau)$ than that based on the normal approximation. Indeed, standard first order theory offers the approximation

$$\Pi(\tau) = 1 - Pr(a(\hat{\lambda}_n - \lambda_n) \leq \tau - a\lambda_n | H_1) = 1 - \Phi(\tau - 2\lambda n a) + O\left(\left(\frac{h}{n}\right)^{1/2}\right), \quad (2.4.5)$$

while, by Theorem 2.3,

$$\begin{aligned} \Pi(\tau) &= 1 - Pr(a(\hat{\lambda}_n - \lambda_n) \leq \tau - a\lambda_n | H_1) = 1 - \Phi(\tau - 2\lambda_n a) \\ &+ \omega(\tau - a\lambda_n)\phi(\tau - 2\lambda_n a) + \frac{\kappa_3^c}{3!} H_2(\tau - 2\lambda_n a)\phi(\tau - 2\lambda_n a) + O\left(\frac{h}{n}\right). \end{aligned} \quad (2.4.6)$$

In Section 2.5 we will present some Monte Carlo results to confirm that the inclusion of the Edgeworth correction terms, as given in the RHS of (2.4.6), entails a closer approximation for $\Pi(\tau)$ than one based on the normal.

A more interesting result that can be derived starting from Theorem 2.3 is a "corrected" version of the test statistics under H_1. In Section 2.1 we have proposed a size-corrected statistic for testing (1.2.1) against a one-sided alternative. Now, from Theorem 2.3, we aim to derive a corrected statistic so that, under H_1 in (2.4.1), the error when its distribution is approximated by a normal is reduced. The corrected version under H_0 can be recovered when $\lambda_n = 0$ in the derivation that follows.

Similarly to the derivation in Section 2.1, (2.4.3) can be written in the equivalent form

$$Pr(a\hat{\lambda}_n \leq \zeta | H_1) = \Phi((\zeta - 2\lambda_n a) - \omega(\zeta - \lambda_n a) - \frac{\kappa_3^c}{3!} H_2(\zeta - 2\lambda_n a)) + O\left(\frac{h}{n}\right).$$

When the function

$$v(x) = x - \omega(x - \lambda_n a) - \frac{\kappa_3^c}{3!} H_2(x - 2\lambda_n a) \tag{2.4.7}$$

is monotonic,

$$Pr(v(a\hat{\lambda}_n) \leq \zeta | H_1) = \Phi(\zeta - 2\lambda_n a) + O\left(\frac{h}{n}\right). \tag{2.4.8}$$

The result in (2.4.8) can be derived by a modification of the argument in Yanagihara et al (2005). Specifically, when v is monotonic,

$$\begin{aligned} Pr(v(a\hat{\lambda}_n) \leq \zeta | H_1) &= Pr(a\hat{\lambda}_n \leq v^{-1}(\zeta) | H_1) \\ &= \Phi(v^{-1}(\zeta) - 2\lambda_n a) - (\omega(v^{-1}(\zeta) - \lambda_n a) \\ &- \frac{\kappa_3^c}{3!} H_2(v^{-1}(\zeta) - 2\lambda_n a))\phi(v^{-1}(\zeta) - 2\lambda_n a) + O\left(\frac{h}{n}\right). \end{aligned} \tag{2.4.9}$$

From (2.4.7),

$$\begin{aligned} \zeta &= v^{-1}\left(\zeta - \omega(\zeta - \lambda_n a) - \frac{\kappa_3^c}{3!} H_2(\zeta - 2\lambda_n a)\right) \\ &= v^{-1}(\zeta) - \left(\omega(\zeta - \lambda_n a) + \frac{\kappa_3^c}{3!} H_2(\zeta - 2\lambda_n a)\right) \frac{dv^{-1}(\zeta)}{d\zeta} + O\left(\frac{h}{n}\right), \end{aligned} \tag{2.4.10}$$

where the last equality follows by Taylor expansion. Let $y = v^{-1}(\zeta)$. Since v is monotonic,

$$\begin{aligned} \frac{dv^{-1}(\zeta)}{d\zeta} &= \left(\frac{dv(y)}{dy}\right)^{-1} = \left(1 - \frac{d\omega(y - \lambda_n a)}{dy} - \frac{\kappa_3^c}{3!} \frac{dH_2(y - 2\lambda_n a)}{dy}\right)^{-1} \\ &= \left(1 + tr(W^3)\left(\lambda_n a^{-2} + \frac{a^{-3}}{4}(y - \lambda_n a) - \frac{2\lambda_n a^{-2} - a^{-3} y}{4}\right) - \frac{\kappa_3^c}{3}(y - 2\lambda_n a)\right)^{-1} \\ &= 1 + O\left(\frac{h}{n}\right)^{1/2}. \end{aligned} \tag{2.4.11}$$

Therefore, by substituting (2.4.11) into (2.4.10),

$$v^{-1}(\zeta) = \zeta + \left(\omega(\zeta - \lambda_n a) + \frac{\kappa_3^c}{3!} H_2(\zeta - 2\lambda_n a)\right) + O\left(\frac{h}{n}\right).$$

Hence, by Taylor expansion,

$$\begin{aligned}\Phi(v^{-1}(\zeta)-2\lambda_n a) &= \Phi(\zeta-2\lambda_n a)\\ &+ \left(\omega(\zeta-\lambda_n a)+\frac{\kappa_3^c}{3!}H_2(\zeta-2\lambda_n a)\right)\phi(\zeta-2\lambda_n a)+O\left(\frac{h}{n}\right),\end{aligned} \tag{2.4.12}$$

$$\phi(v^{-1}(\zeta)-2\lambda_n a)=\phi(\zeta-2\lambda_n a)+O\left(\frac{h}{n}\right)^{1/2}, \tag{2.4.13}$$

$$\omega(v^{-1}(\zeta)-\lambda_n a)=\omega(\zeta-\lambda_n a)+O\left(\frac{h}{n}\right) \tag{2.4.14}$$

and

$$H_2(v^{-1}(\zeta)-2\lambda_n a)=H_2(\zeta-2\lambda_n a)+O\left(\frac{h}{n}\right)^{1/2}. \tag{2.4.15}$$

The result (2.4.8) follows by substituting (2.4.12), (2.4.13), (2.4.14) and (2.4.15) into (2.4.9).

A remark on the monotonicity of the function $v(.)$ is necessary at this stage. In Section 2.1, we explicitly derived the appropriate cubic transformation to guarantee the monotonicity of v over the whole domain without affecting the order of the remainder terms. However, this case is algebraically more complex and the inclusion of the cubic term in the corrected statistic would increase the computational burden (both theoretically and in terms of the simulation time) by a significant amount. Therefore, we rely on some numerical work to state that $v(a\hat{\lambda}_n)$ is indeed locally monotonic under H_1 in (2.4.1).

Hence, when inference is based on $v(a\hat{\lambda}_n)$ rather than $a\hat{\lambda}_n$,

$$\Pi(\tau)=Pr(v(a\hat{\lambda}_n)>\tau|H_1)=1-Pr(v(a\hat{\lambda}_n)\le\tau|H_1)=1-\Phi(\zeta-2\lambda_n a)+O\left(\frac{h}{n}\right).$$

By comparison with (2.4.5), it is straightforward to notice that the error of the approximation is reduced.

2.5 Bootstrap correction and simulation results

In this section we report and discuss some Monte Carlo simulations to investigate the finite sample performance of the tests derived in Sections 2.1, 2.2 and 2.4.

In this simulation work, we adopt the Case (1991) specification for W given in (1.1.2). With this choice, W is symmetric and hence a, b, κ_3^c and κ_4^c can be simplified accordingly. In each of 1000 replications the disturbance terms are $N(0,1)$, i.e. according to Assumption 1 with $\sigma^2=1$. We set $\alpha=95\%$.

A brief remark on W defined in (1.1.2) is necessary. As already mentioned in Chapter 1, it is straightforward to verify that Assumption 2 is satisfied for this choice of W, whether h is bounded or divergent. It is possible to verify that also Assumption

3 holds, i.e.

$$\lim_{n\to\infty} \frac{h}{n} tr(W^i) \neq 0 \quad \text{for} \quad i = 2, 3, 4,$$

by observing that

$$\frac{h}{n} tr((I_r \otimes B_m)^i) = \frac{h}{n} tr(I_r) tr(B_m^i) = \frac{h}{n} r tr(B_m^i).$$

By standard linear algebra, B_m has one eigenvalue equal to 1 and the other $m-1$ equal to $-1/(m-1)$. Therefore

$$tr(B_m^i) = 1 + (m-1)\left(\frac{-1}{m-1}\right)^i$$

and hence

$$\begin{aligned} \frac{h}{n} tr((I_r \otimes B_m)^i) &= \frac{h}{n} r tr(B_m^i) = \frac{m-1}{rm} r \left(1 + \frac{(-1)^i}{(m-1)^{i-1}}\right) \\ &= \frac{m-1}{m}\left(1 + \frac{(-1)^i}{(m-1)^{i-1}}\right), \end{aligned}$$

which is non-zero in the limit whether $m = h+1$ is bounded or not for $i = 2, 4$. When $i = 3$ and m is bounded, we require $m > 2$ (at least for large n) for the latter quantity to be non-zero.

In Tables 2.1-2.4 the empirical sizes of the test of (1.2.1) against a one-sided alternative (Tables 2.1-2.2) and two-sided alternative (Tables 2.3-2.4) based on the usual normal approximation are compared with the same quantities obtained with both the Edgeworth-corrected critical values and corrected test statistics. In addition, we consider the simulated sizes based on bootstrap critical values since it is well established that these may achieve the first Edgeworth correction and should then be comparable with the results obtained in Sections 2.1 and 2.2 (see e.g. Hall (1992) or DiCiccio and Efron (1996)).

Before discussing and comparing the simulation results, the procedure to obtain the bootstrap critical values should be outlined. It must be stressed that we focus on the implementation of the bootstrap procedure, without addressing validity issues. The bootstrap critical values are obtained by the following algorithm:

Step 1 Given model (1.2.5), under H_0, $Y = \epsilon$.

Step 2 Under Assumption 1, a parametric bootstrap can be used, i.e. we construct B $n-$dimensional vectors whose components are independently generated from $N(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2 = \epsilon'\epsilon/n = Y'Y/n$. We denote ϵ_j^*, for $j = 1, ...B$, each of these vectors. Hence, we generate B pseudo-samples as $Y_j^* = \epsilon_j^*$ for $j = 1, ...B$. (When the distribution of the disturbances is known, the parametric bootstrap proved to be more efficient than the usual procedure based on resampling the residuals with replacement, see e.g Hall (1992)).

Step 3 We obtain B bootstrap OLS null statistics as

$$Z_j = a\frac{Y_j^{*\prime}W'Y_j^*}{Y_j^{*\prime}W'WY_j^*}, \qquad j = 1, ...B.$$

Step 4 The $\alpha-$percentile is computed as the value w_α^* which solves

$$\frac{1}{B}\sum_{j=1}^{B} 1(Z_j \leq w_\alpha^*) \leq \alpha.$$

Step 5 The size of the test of (1.2.1) when the bootstrap critical value is used is then

$$Pr(a\hat{\lambda} > w_\alpha^* | H_0). \tag{2.5.1}$$

The extension of Steps 4-5 in the latter procedure to the test of (1.2.1) against a two-sided alternative is straightforward, i.e. the $\alpha-$percentile is computed as the value q_α^* which solves $\sum_{j=1}^{B} 1(|Z_j| \leq q_\alpha^*)/B = \alpha$ and the size based on such critical value is computed as

$$Pr(|a\hat{\lambda}| > q_\alpha^* | H_0). \tag{2.5.2}$$

In both cases, we set $B = 199$.

Regarding Step 1, a remark is needed. When we are interested in testing, the bootstrap procedure with H_0 imposed to obtain the residuals (and then to generate the pseudo-data) gives results at least as good as the same algorithm without imposing H_0 (see Paparoditis and Politis (2005)).

Tables 2.1 and 2.2 display the simulated values corresponding to (2.1.12), (2.1.13), (2.1.15) and (2.5.1) when m is increased monotonically and kept fixed, respectively. The former case is indeed consistent with divergent h, while the latter correspond to a bounded h. For such reason, henceforth we refer to "divergent" and "bounded" h. Moreover, Tables 2.3 and 2.4 display the simulated values corresponding to $Pr(|a\hat{\lambda}| > z_{(\alpha+1)/2}|H_0)$, $Pr(|a\hat{\lambda}| > z_{(\alpha+1)/2} + p_1(z_{(\alpha+1)/2})|H_0)$, where $p_1(.)$ is defined according to (2.2.5), $Pr(v(|a\hat{\lambda}|) > z_{(\alpha+1)/2}|H_0)$, with $v(.)$ given by (2.2.6), and (2.5.2) when h is either "divergent" or "bounded", respectively. All the values in Tables 2.1-2.4 have to be compared with the nominal 5%. For notational convenience, in the Tables we denote by "normal", "Edgeworth", "transformation" and "bootstrap" the simulated values corresponding to the size obtained with the standard approximation, Edgeworth-corrected critical values, Edgeworth-corrected null statistic and bootstrap critical values, respectively.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
normal	0	0	0.001	0.001
Edgeworth	0.125	0.117	0.110	0.099
transformation	0.056	0.055	0.052	0.048
bootstrap	0.039	0.061	0.053	0.054

Table 2.1: Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.1.1) when λ in model (1.2.5) is estimated by OLS and the sequence h is "divergent". The reported values have to be compared with the nominal 0.05.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
normal	0.001	0.001	0.001	0.011
Edgeworth	0.096	0.070	0.057	0.052
transformation	0.055	0.057	0.055	0.051
bootstrap	0.043	0.040	0.057	0.055

Table 2.2: Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.1.1) when λ in model (1.2.5)is estimated by OLS and the sequence h is "bounded". The reported values have to be compared with the nominal 0.05.

By observing the results in Tables 2.1 and 2.2, it is clear that the usual normal approximation does not work well in practice, since the simulated values for the size greatly underestimate the nominal 5% for all sample sizes. On the other hand, the Edgeworth-corrected results seem to perform reasonably well. However, the results obtained with the Edgeworth-corrected critical values exceed the target 0.05 for very small sample sizes, but the convergence to the nominal value appears to be fast. Indeed, such correction performs already quite well for moderate sample sizes such as $m = 18$, $r = 14$.

More specifically, when h is "divergent" and inference is based on Edgeworth-corrected critical values, the discrepancy between the simulated values and the nominal 5% appears to be 26% higher than such discrepancy obtained when inference is based on standard normal critical values, on average across sample sizes. However, we also notice that the difference between actual and nominal values only decreases by about 0.6%, on average, when sample size increases in case of the standard test, while it decreases by 13% when Edgeworth-corrected critical values are used. On the other hand, the simulated sizes based on the Edgeworth-corrected statistics are very satisfactory also for very small sample sizes. Indeed, on average across sample sizes, when h is "divergent" and inference is based on the Edgeworth-corrected statistic and bootstrap critical values, the simulated values are 92% and 85%, respectively, closer to 0.05 than values obtained with the standard t-statistic.

A similar pattern can be observed in Tables 2.2. When h is bounded the cdf of $a\hat{\lambda}$ under (1.2.1) converges faster to the normal. The figures displayed in Table 2.1

and 2.2 are consistent with this theoretical result. Indeed we notice from the first column of Table 2.2 that, on average, the difference between simulated values and the nominal 0.05 decreases by 6% as sample size increases when inference is based on the standard statistic (such value has to be compared with the aforementioned 0.6% decrease in case h is "divergent"). Also, we notice that in Table 2.2, the average improvements on average across sample sizes offered by Edgeworth-corrected critical values, Edgeworth-corrected statistic and bootstrap critical values over the standard OLS t-statistic are about 98%, 87% and 99%, respectively.

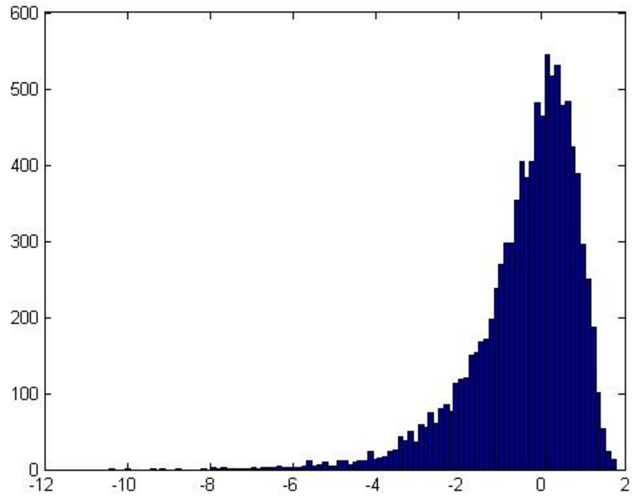

Figure 2.1: Simulated pdf of $a\hat{\lambda}$ under H_0 in (1.2.1)

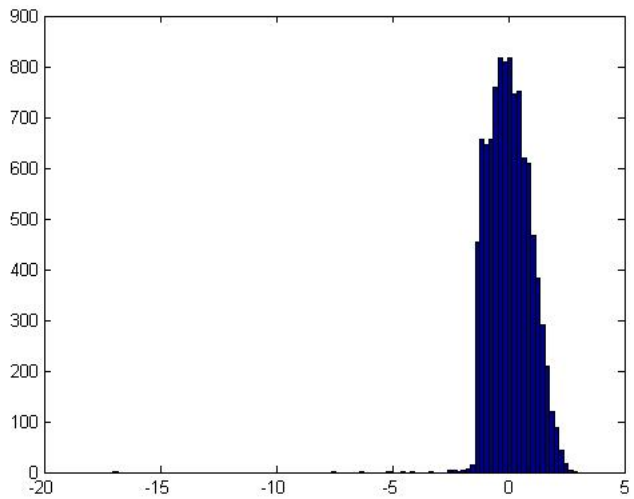

Figure 2.2: Simulated pdf of $g(a\hat{\lambda})$ under H_0 in (1.2.1)

In Figures 2.1 and 2.2 we plot the pdf obtained from the Monte Carlo simulation of the non-corrected OLS null statistic $a\hat{\lambda}$ and its corrected version $g(a\hat{\lambda})$. The pdf of the non-corrected statistics is very skewed to the left but most of this skewness is removed by the corrected version.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
normal	0.132	0.130	0.126	0.106
Edgeworth	0.062	0.060	0.056	0.055
transformation	0.130	0.128	0.105	0.098
bootstrap	0.048	0.044	0.045	0.047

Table 2.3: Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.2.1) when λ in model (1.2.5) is estimated by OLS and the sequence h is "divergent". The reported values have to be compared with the nominal 0.05.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
normal	0.096	0.078	0.068	0.061
Edgeworth	0.040	0.052	0.047	0.046
transformation	0.063	0.025	0.044	0.052
bootstrap	0.049	0.047	0.051	0.050

Table 2.4: Empirical sizes of the tests of H_0 in (1.2.1) against H_1 in (2.2.1) when λ in model (1.2.5)is estimated by OLS and the sequence h is "bounded". The reported values have to be compared with the nominal 0.05.

From Tables 2.3 and 2.4 it is clear again that the normal approximation does not produce satisfactory results, since the nominal 5% is greatly overestimated for all sample sizes, whether h is "divergent" or "bounded". In turn, results obtained with the Edgeworth-corrected critical values are very close to the nominal for all sample sizes. Indeed, when inference is based on Edgeworth-corrected critical values, the discrepancy between the simulated values and the nominal is reduced on average across sample sizes by 89% when h is "divergent" and by 79% when h is "bounded". On the other hand, simulated sizes based on the Edgeworth-corrected statistic seem still to greatly overestimate the target 5%, especially when h is "divergent" but appear to decrease to the nominal value quite fast. Specifically the improvement offered by the Edgeworth-corrected statistic over the standard one is 58% when h is "bounded", but only 13% when h is "divergent". Results based on bootstrap critical values are, as expected, comparable to the Edgeworth-corrected ones and are very close to 5% for all sample sizes. Again, the pattern of the results is similar for "divergent" and "bounded" h.

As mentioned in Section 2.2, a remark on the monotonicity of $v(.)$ in (2.2.6) is needed. Indeed, some numerical work shows that $v(.)$ cannot be considered locally strictly increasing unless n is very large. Hence, the corresponding results in Tables 2.3 and 2.4 have been derived by a numerical implementation of the cubic transformation in (2.1.14). Such numerical implementation can indeed be the reason of the less satisfactory performance of the Edgeworth-corrected statistic compared to the corrected critical values.

	$\bar{\lambda}$	$m = 8$, $r = 5$	$m = 12$, $r = 8$	$m = 18$, $r = 11$	$m = 28$, $r = 14$
normal	0.1	0	0	0.005	0.009
	0.5	0	0.335	0.673	0.854
	0.8	0.257	0.994	1	1
Edgeworth	0.1	0.561	0.610	0.663	0.693
	0.5	0.952	0.986	0.993	1
	0.8	1	1	1	1
bootstrap	0.1	0.111	0.119	0.155	0.164
	0.5	0.725	0.873	0.938	0.966
	0.8	0.996	1	1	1

Table 2.5: Empirical powers of the tests of H_0 in (1.2.1) against H_1 in (2.5.3), with $\bar{\lambda} = 0.1, 0.5, 0.8$, when λ in model (1.2.5) is estimated by OLS and the sequence h is "divergent". α is set to 0.95.

	$\bar{\lambda}$	$m = 5$, $r = 8$	$m = 5$, $r = 20$	$m = 5$, $r = 40$	$m = 5$, $r = 80$
normal	0.1	0.010	0.083	0.187	0.363
	0.5	0.551	0.988	1	1
	0.8	0.999	1	1	1
Edgeworth	0.1	0.640	0.739	0.852	0.693
	0.5	0.991	1	1	1
	0.8	1	1	1	1
bootstrap	0.1	0.139	0.203	0.296	0.451
	0.5	0.888	0.992	1	1
	0.8	1	1	1	1

Table 2.6: Empirical powers of the tests of H_0 in (1.2.1) against H_1 in (2.5.3), with $\bar{\lambda} = 0.1, 0.5, 0.8$, when λ in model (1.2.5) is estimated by OLS and the sequence h is "bounded". α is set to 0.95.

In Tables 2.5 and 2.6 we report some Monte Carlo results to assess the finite sample behaviour of the power of both standard and corrected tests of (1.2.1) against a fixed alternative hypothesis, i.e.

$$H_1 : \lambda = \bar{\lambda} > 0. \tag{2.5.3}$$

Obviously, the same argument can be carried on with very minor modifications in case $\bar{\lambda} < 0$. In Tables 2.5 and 2.6 we report the simulated quantities corresponding to $Pr(a\hat{\lambda} > z_\alpha | H_1)$, $Pr(a\hat{\lambda} > z_\alpha + p_1(z_\alpha)|H_1)$ and $Pr(a\hat{\lambda} > w_\alpha^* | H_1)$. We choose three different values of $\bar{\lambda}$, specifically $\bar{\lambda} = 0.1, 0.5, 0.8$. The values in Tables 2.5 and 2.6 are consistent with the empirical sizes reported in Tables 2.1 and 2.2. In particular, we observe that, when $\bar{\lambda} = 0.1$ for instance, the simulated power when inference is based on Edgeworth-corrected critical values is (on average across sample sizes) more than 300% higher than the corresponding result based on bootstrap critical values. Such a huge difference can be explained by the sign of the probability limit of $\hat{\lambda} - \lambda$ when W is chosen according to (1.1.2).

Indeed, as previously mentioned, $\hat{\lambda}$ is inconsistent when $\lambda \neq 0$. Therefore, in case $plim\hat{\lambda} < \lambda$ for $\lambda > 0$, it might be that under H_1, $plim\hat{\lambda} = 0$ (obviously, for $\lambda < 0$ the argument would be modified as: in case $plim\hat{\lambda} > \lambda$ as $n \to \infty$ it might be that under H_1, $plim\hat{\lambda} = 0$). In this case, the standard test of (1.2.1) against (2.5.3) would be inconsistent. Nevertheless, it is quite straightforward to evaluate the sign of the probability limit of $\hat{\lambda}$ for any particular choice of W. Specifically,

Theorem 2.4 *Suppose that model (1.2.5) holds. Under Assumption 1 and for W given by (1.1.2), $\underset{n\to\infty}{plim}(\hat{\lambda} - \lambda)$ is finite and has the same sign of λ.*

The proof of Theorem 2.4 is in the Appendix. It is worth mentioning that the sign of the probability limit in Theorem 2.4 can be computed similarly for any other choices of W, although it might not always be possible to obtain close form expressions. Obviously, Assumption 1 could be relaxed. However, Assumption 1 has been assumed throughout this project and is retained here for algebraic simplicity.

By Theorem 2.4, as $n \to \infty$, $plim\hat{\lambda} > \lambda$ when $\lambda > 0$ (or $plim\hat{\lambda} < \lambda$ when $\lambda < 0$) and hence it is straightforward to show that, as $n \to \infty$, $Pr(a\hat{\lambda} > z_\alpha|H_1) \to 1$, $Pr(a\hat{\lambda} > z_\alpha + p_1(z_\alpha)|H_1) \to 1$ and $Pr(g(a\hat{\lambda}) > z_\alpha|H_1) \to 1$, i.e. our new tests based on OLS estimates for λ are consistent when W chosen according to (1.1.2). As anticipated, the result of Theorem 2.4 also explains why the simulated values for the power of a test of (1.2.1) against (2.5.3) based on Edgeworth-corrected critical are so much higher than the same quantities obtained by bootstrap.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
Monte Carlo power	0	0.020	0.046	0.070
1^{st} order approximation	0.304	0.304	0.304	0.304
2^{nd} order approximation	0.054	0.089	0.111	0.127

Table 2.7: Numerical values corresponding to (2.5.5) (second row) and (2.5.6) (third row), compared with the simulated values for the power of a test of (1.2.1) against (2.4.1) when λ_n in model (1.2.5) is estimated by OLS and the sequence h is "divergent".

	$m=5$ $r=8$	$m=5$ $r=20$	$m=5$ $r=40$	$m=5$ $r=80$
Monte Carlo power	0.054	0.171	0.213	0.250
1^{st} order approximation	0.304	0.304	0.304	0.304
2^{nd} order approximation	0.138	0.199	0.229	0.252

Table 2.8: Numerical values corresponding to (2.5.5) (second row) and (2.5.6) (third row), compared with the simulated values for the power of a test of (1.2.1) against (2.4.1) when λ_n in model (1.2.5) is estimated by OLS and the sequence h is "bounded".

In the first row of Tables 2.7 and 2.8 we report the simulated values corresponding to

$$Pr(a\hat{\lambda}_n > z_\alpha | H_1), \tag{2.5.4}$$

where H_1 is given in (2.4.1). These should be compared with the values obtained by the normal approximation (reported in second row)

$$1 - 2\Phi(z_\alpha - 2a\lambda_n) \tag{2.5.5}$$

and with the values obtained by the Edgeworth-corrected approximation (reported in the third row), i.e.

$$1 - \Phi(\tau - 2\lambda_n a) + \omega(\tau - a\lambda_n)\phi(\tau - 2\lambda_n a) + \frac{\kappa_3^c}{3!} H_2(\tau - 2\lambda_n a)\phi(\tau - 2\lambda_n a), \tag{2.5.6}$$

where $\omega(.)$ is defined in (2.4.4). In Tables 2.7 and 2.8 the sample is increased consistently with a divergent and bounded h, respectively. We choose $c = 0.8$ in the expression for λ_n given in (2.4.1), although a different choice for c does not change the pattern of the results of the simulations.

As expected, the actual power obtained in the Monte Carlo simulations tends to the value corresponding to (2.5.5) when the sample size is large. However, the values obtained by (2.5.6) are 23% and 19% closer, on average across sample sizes, to the simulated ones when h is "divergent" and "bounded", respectively. The difference between the values obtained by (2.5.5) and those obtained by (2.5.6) becomes increasingly smaller as the sample size increases and, as expected, the convergence is faster in case of "bounded" h. Specifically, when h is "bounded", such difference decreases by 32% on average as sample size increases, but only 11% in case h is "divergent".

	$m=8$ $r=5$	$m=12$ $r=8$	$m=18$ $r=11$	$m=28$ $r=14$
Monte Carlo power	0	0.020	0.046	0.070
Monte Carlo power/corrected	0.180	0.231	0.248	0.253

Table 2.9: Simulated values of the power of a test of (1.2.1) against (2.4.1) based on the standard and corrected statistics when λ_n in model (1.2.5) is estimated by OLS and the sequence h is "divergent". The values should be compared with the target 0.304.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
Monte Carlo power	0.054	0.171	0.213	0.250
Monte Carlo power/corrected	0.223	0.275	0.278	0.290

Table 2.10: Simulated values of the power of a test of (1.2.1) against (2.4.1) based on the standard and corrected statistics when λ_n in model (1.2.5) is estimated by OLS and the sequence h is bounded. The values should be compared with the target 0.304.

Finally, in Tables 2.9 and 2.10 we compare the simulated values corresponding to (2.5.4), reported in the first row, with

$$Pr(v(a\hat{\lambda}_n) > z_\alpha | H_1), \tag{2.5.7}$$

where $v(.)$ is defined in (2.4.7). Tables 2.9 and 2.10 correspond to divergent and bounded h, respectively. The values in Tables 2.9 and 2.10 should be compared with (2.5.5), which is 0.304 for this particular setting. From both tables it is clear that, as expected, the results when inference is based on the corrected statistic are closer to the target value 0.304 for all sample sizes. In particular, the simulated values for the power based on the Edgeworth-corrected statistic are, on average across sample sizes, 72% and 19% closer to the nominal value than those based on the standard statistic, when h is "divergent" and "bounded", respectively. As expected, when h is "bounded", the discrepancy between actual and nominal values decreases faster as sample size increases. Specifically, when inference is based on the standard statistic, such discrepancy decreases by 8% on average as sample size increases when h is "divergent" and by 40% when h is "bounded".

A Appendix

Proof of Theorem 2.1

The OLS estimate of λ in (1.2.5) is defined as

$$\hat{\lambda} - \lambda = \frac{Y'W'\epsilon}{Y'W'WY}$$

and therefore, under H_0,

$$\hat{\lambda} = \frac{\epsilon'W'\epsilon}{\epsilon'W'W\epsilon}.$$

The cdf of $\hat{\lambda}$ under H_0 can be written in terms of a quadratic form in ϵ, i.e.

$$Pr(\hat{\lambda} \leq x) = Pr(f \leq 0),$$

where

$$\begin{aligned} f &= \frac{1}{2}\epsilon'(C + C')\epsilon, \\ C &= W' - xW'W \end{aligned} \tag{2.A.1}$$

and x is any real number.

Under Assumption 1, the characteristic function of f can be derived as

$$\begin{aligned} E(e^{it(\frac{1}{2}(\epsilon'(C+C')\epsilon)}) &= \frac{1}{(2\pi)^{n/2}\sigma^n}\int\limits_{\Re^n} e^{it(\frac{1}{2}(\xi'(C+C')\xi)}e^{-\frac{\xi'\xi}{2\sigma^2}}d\xi \\ &= \frac{1}{(2\pi)^{n/2}\sigma^n}\int\limits_{\Re^n} e^{-\frac{1}{2\sigma^2}\xi'(I - it\sigma^2(C+C'))\xi}d\xi \\ &= det(I - it\sigma^2(C + C'))^{-1/2} = \prod_{j=1}^{n}(1 - it\sigma^2\eta_j(C + C'))^{-1/2}, \end{aligned} \tag{2.A.2}$$

where $\eta_j(C+C')$ are the eigenvalues of $(C+C')$. From (2.A.2) the cumulant generating function of f is

$$\begin{aligned} \psi(t) &= -\frac{1}{2}\sum_{j=1}^{n} ln(1 - it\sigma^2\eta_j(C + C')) = \frac{1}{2}\sum_{j=1}^{n}\sum_{s=1}^{\infty}\frac{(it\sigma^2\eta_j(C + C'))^s}{s} \\ &= \frac{1}{2}\sum_{s=1}^{\infty}\frac{(it\sigma^2)^s}{s}\sum_{j=1}^{n}\eta_j(C + C')^s = \frac{1}{2}\sum_{s=1}^{\infty}\frac{(it\sigma^2)^s}{s}tr((C + C')^s). \end{aligned} \tag{2.A.3}$$

From (2.A.3) the s-th cumulant of f can be derived as

$$\kappa_1 = \sigma^2 tr(C), \tag{2.A.4}$$

$$\kappa_2 = \frac{\sigma^4}{2} tr((C + C')^2), \tag{2.A.5}$$

$$\kappa_s = \frac{\sigma^{2s} s!}{2} \frac{tr((C + C')^s)}{s}, s > 2. \tag{2.A.6}$$

Let

$$f^c = \frac{f - \kappa_1}{\kappa_2^{1/2}},$$

i.e. the centred and scaled version of f. The cumulant generating function of f^c is

$$\psi^c(t) = -\frac{1}{2}t^2 + \sum_{s=3}^{\infty} \frac{\kappa_s^c (it)^s}{s!},$$

where

$$\kappa_s^c = \frac{\kappa_s}{\kappa_2^{s/2}}, \tag{2.A.7}$$

so the characteristic function of f^c is

$$\begin{aligned}
E(e^{itf^c}) &= e^{-\frac{1}{2}t^2} \exp\{\sum_{s=3}^{\infty} \frac{\kappa_s^c (it)^s}{s!}\} = \\
&= e^{-\frac{1}{2}t^2}\{1 + \sum_{s=3}^{\infty} \frac{\kappa_s^c (it)^s}{s!} + \frac{1}{2!}(\sum_{s=3}^{\infty} \frac{\kappa_s^c (it)^s}{s!})^2 + \frac{1}{3!}(\sum_{s=3}^{\infty} \frac{\kappa_s^c (it)^s}{s!})^3 + ...\} \\
&= e^{-\frac{1}{2}t^2}\{1 + \frac{\kappa_3^c (it)^3}{3!} + \frac{\kappa_4^c (it)^4}{4!} + \frac{\kappa_5^c (it)^5}{5!} + \{\frac{\kappa_6^c}{6!} + \frac{(\kappa_3^c)^2}{(3!)^2}\}(it)^6 + ...\}.
\end{aligned}$$

Thus, by the Fourier inversion formula,

$$Pr(f^c \le z) = \int_{-\infty}^{z} \phi(z)dz + \frac{\kappa_3^c}{3!} \int_{-\infty}^{z} H_3(z)\phi(z)dz + \frac{\kappa_4^c}{4!} \int_{-\infty}^{z} H_4(z)\phi(z)dz + ...$$

Collecting the results derived above,

$$\begin{aligned}
Pr(\hat{\lambda} \le x) = Pr(f \le 0) = Pr(f^c \kappa_2^{1/2} + \kappa_1 \le 0) = Pr(f^c \le -\kappa_1^c) \\
= \Phi(-\kappa_1^c) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(-\kappa_1') + ...
\end{aligned} \tag{2.A.8}$$

From (2.A.4), (2.A.5) and (2.A.7),

$$\kappa_1^c = \frac{\sigma^2 tr(C)}{\sigma^2(\frac{1}{2}tr((C + C')^2))^{1/2}},$$

where C is defined according to (2.A.1). The numerator of κ_1^c is

$$\sigma^2 tr(W) - \sigma^2 x tr(W'W) = -\sigma^2 x tr(W'W),$$

while the denominator of κ_1^c is σ^2 times

$$(\frac{1}{2}tr(C+C')^2)^{1/2} \quad = \quad (tr(W^2)+tr(WW')-4xtr(WW'W)+2x^2tr((W'W)^2))^{1/2}.$$

Thus

$$\begin{aligned}\kappa_1^c &= \frac{-xtr(W'W)}{(tr(W^2)+tr(WW')-4xtr(WW'W)+2x^2tr((W'W)^2))^{1/2}}\\ &= \frac{-xtr(W'W)}{(tr(W^2+WW'))^{1/2}(1-\frac{4xtr(WW'W)+2x^2tr((W'W)^2)}{(tr(W^2+WW'))})^{1/2}}.\end{aligned}$$

We choose $x = a^{-1}\zeta$, where

$$a = \frac{tr(W'W)}{(tr(W'W+W^2))^{1/2}}.$$

Moreover,

$$b_1 = \frac{tr(WW'W)}{tr(W'W+W^2)}$$

and

$$b_2 = \frac{tr((W'W)^2)}{tr(W'W+W^2)}.$$

Now,

$$\begin{aligned}\kappa_1^c &= \frac{-xtr(W'W)}{(tr(W'W+W^2))^{1/2}(1-4xb_1+2x^2b_2)^{1/2}} = \frac{-\zeta}{(1-4xb_1+2x^2b_2)^{1/2}}\\ &= -\zeta\left(1+2a^{-1}\zeta b_1 - a^{-2}\zeta^2 b_2 + 6a^{-2}\zeta^2 b_1^2 + O\left(\left(\frac{h}{n}\right)^{3/2}\right)\right)\\ &= -\zeta - 2a^{-1}\zeta^2 b_1 + a^{-2}\zeta^3 b_2 - 6a^{-2}b_1^2\zeta^3 + O\left(\left(\frac{h}{n}\right)^{3/2}\right),\end{aligned}$$

where the third equality follows by performing a standard Taylor expansion of the term $(1-4xb_1+2x^2b_2)^{-1/2}$, i.e.

$$(1-4xb_1+2x^2b_2)^{-1/2} = 1+2xb_1-x^2b_2+6x^2b_1^2+O\left(\left(\frac{h}{n}\right)^{3/2}\right).$$

Under Assumption 3,

$$2a^{-1}\zeta^2 b_1 \sim \left(\frac{h}{n}\right)^{1/2}, \qquad a^{-2}\zeta^3 b_2 \sim \frac{h}{n}, \qquad 6a^{-2}b_1^2\zeta^3 \sim \frac{h}{n}.$$

Moreover, by Taylor expansion,

$$
\begin{aligned}
\Phi(-\kappa_1^c) &= \Phi\left(\zeta + 2a^{-1}\zeta^2 b_1 - a^{-2}\zeta^3 b_2 + 6a^{-2}\zeta^3 b_1^2 + O\left(\left(\frac{h}{n}\right)^{3/2}\right)\right) \\
&= \Phi(\zeta) + (2a^{-1}\zeta^2 b_1 - a^{-2}\zeta^3 b_2 + 6a^{-2}\zeta^3 b_1^2)\phi(\zeta) \\
&+ 2a^{-2}\zeta^4 b_1^2 \Phi^{(2)}(\zeta) + O\left(\left(\frac{h}{n}\right)^{3/2}\right) \qquad (2.A.9)
\end{aligned}
$$

and

$$
\Phi^{(3)}(-\kappa_1^c) = \Phi^{(3)}(\zeta) + 2a^{-1}\zeta^2 b_1 \Phi^{(4)}(\zeta) + O\left(\frac{h}{n}\right). \qquad (2.A.10)
$$

Collecting (2.A.8), (2.A.9) and (2.A.10), the third order Edgeworth expansion of the cdf of $a\hat{\lambda}$ under Assumptions 1-3, becomes

$$
\begin{aligned}
\Pr(a\hat{\lambda} \leq \zeta | H_0) &= \Phi(\zeta) + 2a^{-1}b_1\zeta^2\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) \\
&- (a^{-2}b_2 - 6a^{-2}b_1^2)\zeta^3\phi(\zeta) + 2a^{-2}b_1^2\zeta^4\Phi^{(2)}(\zeta) \\
&- \frac{\kappa_3^c}{3}a^{-1}b_1\zeta^2\Phi^{(4)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + O\left(\left(\frac{h}{n}\right)^{3/2}\right),
\end{aligned}
$$

where, from (2.A.5), (2.A.6) and (2.A.7),

$$
\kappa_3^c = \frac{\sigma^6 tr((C+C')^3)}{\sigma^6(\frac{1}{2}tr((C+C')^2))^{3/2}} \sim \frac{2tr(W^3) + 6tr(W'W^2)}{(tr(W'W + W^2))^{3/2}}
$$

and

$$
\kappa_4^c = \frac{3\sigma^8 tr((C+C')^4)}{\sigma^8(\frac{1}{2}tr((C+C')^2))^2} \sim \frac{6tr(W^4) + 24tr(W'W^3) + 12tr((WW')^2) + 6tr(W^2W'^2)}{(tr(W'W + W^2))^2}.
$$

Setting $b = b_1 a^{-1}$ and substituting the expression for a and b_2 into $a^{-2}b_2$, the expansion stated in Theorem 2.1 follows.

Proof of Theorem 2.2

The OLS estimate of λ in (2.3.1) is defined as

$$
\hat{\lambda} - \lambda = \frac{Y'W'P\epsilon}{Y'W'PWY},
$$

where $P = I - l(l'l)^{-1}l'$. Since W is row normalized, $Wl = l$. Hence, under H_0,

$$
\hat{\lambda} = \frac{\epsilon'W'P\epsilon}{\epsilon'W'PW\epsilon}.
$$

Similarly to the proof of Theorem 2.1, the cdf of $\hat{\lambda}$ under H_0 can be written in terms of a quadratic form in ϵ, i.e.

$$Pr(\hat{\lambda} \leq x) = Pr(f \leq 0),$$

where

$$f = \frac{1}{2}\epsilon'(C + C')\epsilon$$

and

$$C = W'P(I - xW). \tag{2.A.11}$$

The derivation of the cumulants is similar to one in the proof of Theorem 2.1 with C defined according to (2.A.11) and is therefore omitted. Given (2.A.11),

$$\kappa_1 = \sigma^2 tr(C) = -\sigma^2 \left(1 + xtr(W'W) - \frac{x}{n}(l'WW'l)\right),$$

since

$$\begin{aligned} tr(C) &= tr(W'P(I - xW)) = tr(W) - tr(W'l(l'l)^{-1}l') \\ &- xtr(W'W) + xtr(W'l(l'l)^{-1}l'W) \\ &= -\frac{1}{n}l'W'l - xtr(W'W) + \frac{x}{n}(l'WW'l) \\ &= -1 - xtr(W'W) + \frac{x}{n}(l'WW'l). \end{aligned}$$

Similarly, by straightforward algebra,

$$\begin{aligned} \frac{1}{2}tr((C + C')^2) &= \frac{1}{2}tr((W'P + PW - 2xW'PW)^2) \\ &= tr(W^2) + tr(W'W) + 1 - \frac{2}{n}l'Wl - \frac{1}{n}(l'WW'l) \\ &- 4xtr(W'PW'PW) + 2x^2tr((W'PW)^2) \end{aligned}$$

and hence

$$\begin{aligned} \kappa_2 &= \frac{\sigma^4}{2}tr((C + C')^2) \\ &= \sigma^4(tr(W^2) + tr(W'W) + 1 - \frac{2}{n}l'Wl - \frac{1}{n}(l'WW'l) \\ &- 4xtr(W'PW'PW) + 2x^2tr((W'PW)^2)). \end{aligned}$$

Proceeding as in the proof of Theorem 2.1, we obtain the first centred cumulant as

$$\kappa_1^c = \frac{-\left(xtr(W'W) + 1 - \frac{x}{n}l'WW'l\right)}{\gamma^{1/2}}$$

$$\times \quad \left(1+\frac{1-\frac{2}{n}l'Wl-\frac{1}{n}(l'WW'l)-4xtr(W'PW'PW)+2x^2tr((W'PW)^2)}{\gamma}\right)^{-1/2},$$

where $\gamma = tr(W^2+W'W)$. From Lemma 2.2, $l'Wl$ and $(l'WW'l)$ are $O(n)$, $tr(W'PW'PW)$ and $tr((W'PW)^2)$ are $O(n/h)$ by Lemma 2.1 (since P is uniformly bounded in row and column sums in absolute value and the product of matrices which are uniformly bounded in row and column sums retains the same property) and $\gamma \sim n/h$ under Assumption 3.

By setting $x = a^{-1}\zeta$ where a has been defined in (2.1.4) and by standard Taylor expansion,

$$\begin{aligned}
\kappa_1^c &= -\left(\zeta+\frac{1}{\gamma^{1/2}}+O\left(\frac{h}{n}\right)\right)\left(1+\frac{2a^{-1}tr(W'PW'PW)}{\gamma}\zeta+O\left(\frac{h}{n}\right)\right) \\
&= -\zeta-\frac{1}{\gamma^{1/2}}-\frac{2a^{-1}tr(W'PW'PW)}{\gamma}\zeta^2+O\left(\frac{h}{n}\right) \\
&= -\zeta-\frac{1}{\gamma^{1/2}}-\frac{2tr(W'PW'PW)}{\gamma^{1/2}tr(W^2)}\zeta^2+O\left(\frac{h}{n}\right) \\
&= -\zeta-\frac{1}{\gamma^{1/2}}-\frac{2tr(W'WW')}{\gamma^{1/2}tr(W^2)}\zeta^2+O\left(\frac{h}{n}\right),
\end{aligned}$$

where the last equality follows since $tr(W'PW'PW) = tr(W'WW') + O(1)$.

Proceeding as in the proof of Theorem 2.1,

$$Pr(a\hat{\lambda}\le\zeta|H_0) \;=\; \Phi(\zeta)+\left(\frac{1}{\gamma^{1/2}}+\frac{2tr(W'WW')}{\gamma^{1/2}tr(W^2)}\zeta^2\right)\phi(\zeta)-\frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta)+O\left(\frac{h}{n}\right),$$

where

$$\kappa_3^c \sim \frac{\sigma^6 tr(C+C')^3}{\sigma^6\left(\frac{1}{2}tr(C+C')^2\right)^{3/2}} \sim \frac{2tr((W'P)^3)+6tr((W'P)^2PW)}{(tr(W^2+W'W))^{3/2}} \sim \sqrt{\frac{h}{n}}.$$

The last displayed rate holds since the leading terms of $tr((W'P)^3)$ and $6tr((W'P)^2PW)$ are $tr(W^3)$ and $tr(W'WW')$, respectively, which have exactly order n/h under Assumption 3.

The expansion in Theorem 2.2 follows by observing that $tr(W'WW')/\gamma^{1/2}tr(W^2) = b$, where b is defined according to (2.1.4).

Proof of Theorem 2.3

The general structure of the proof is similar to ones of Theorems 2.1 and 2.2 and hence several details will be omitted.

Since $Y = S^{-1}(\lambda_n)\epsilon$,

$$\hat{\lambda}_n - \lambda_n = \frac{Y'W\epsilon}{Y'W^2Y} = \frac{\epsilon' S^{-1}(\lambda_n)W\epsilon}{\epsilon' S^{-1}(\lambda_n)W^2S^{-1}(\lambda_n)\epsilon}.$$

It is straightforward to see that, given (2.4.2), if W is symmetric so are $S^{-1}(\lambda_n)$ and $S^{-1}(\lambda_n)W$. By standard manipulation, the cdf of $\hat{\lambda}_n - \lambda_n$ under H_1 can be written in terms of a quadratic form in ϵ, i.e.

$$Pr(\hat{\lambda}_n - \lambda_n \leq x) = Pr(f \leq 0),$$

where $f = \epsilon' C\epsilon$,

$$C = S^{-1}(\lambda_n)W - xS^{-1}(\lambda_n)W^2S^{-1}(\lambda_n)$$

and x is any real number.

Similarly to the proofs of Theorems 2.1 and 2.2, the first cumulant of f is

$$\begin{aligned}
&\kappa_1 = \sigma^2 tr(C) = \sigma^2(tr(S^{-1}(\lambda_n)W) - xtr(S^{-1}(\lambda_n)W^2S^{-1}(\lambda_n)))\\
&= \sigma^2\left(tr\left(\sum_{t=0}^{\infty}(\lambda_n W)^t W\right) - xtr\left(\sum_{t=0}^{\infty}(\lambda_n W)^t W^2 \sum_{s=0}^{\infty}(\lambda_n W)^s\right)\right)\\
&= \sigma^2\left(\sum_{t=1}^{\infty}\lambda_n^t tr(W^{t+1}) - xtr(W^2) - 2x\lambda_n tr(W^3) - xtr(\sum_{t=1}^{\infty}(\lambda_n W)^t W^2 \sum_{s=1}^{\infty}(\lambda_n W)^s)\right)\\
&= \sigma^2(\lambda_n tr(W^2) + \lambda_n^2 tr(W^3) + \sum_{t=3}^{\infty}\lambda_n^t tr(W^{t+1}) - xtr(W^2) - 2x\lambda_n tr(W^3)\\
&- x\sum_{t,s=1}^{\infty}\lambda_n^t\lambda_n^s tr(W^{t+1}W^{s+1})),
\end{aligned} \tag{2.A.12}$$

where the third equality follows by (2.4.2). By Lemma 2.1,

$$tr(W^{t+1}) = O\left(\frac{n}{h}\right) \quad \text{and} \quad tr(W^{t+1}W^{s+1}) = O\left(\frac{n}{h}\right),$$

for every t and s. Hence, under H_1,

$$\sum_{t,s=1}^{\infty}\lambda_n^t\lambda_n^s tr(W^{t+1}W^{s+1}) = O\left(\frac{n}{h}\right)\left(\sum_{t=1}^{\infty}\lambda_n^t\right)^2 = O\left(\frac{n}{h}\right)\left(\frac{\lambda_n}{1-\lambda_n}\right)^2 = O(1)$$

and

$$\sum_{t=3}^{\infty}\lambda_n^t tr(W^{t+1}) = O\left(\frac{n}{h}\right)\sum_{t=3}^{\infty}\lambda_n^t = O\left(\frac{n}{h}\right)\left(\frac{\lambda_n^3}{1-\lambda_n}\right) = O\left(\sqrt{\frac{h}{n}}\right).$$

By a similar argument,

$$\begin{aligned}
&\kappa_2 = 2\sigma^4 tr(C^2) = 2\sigma^2(tr((S^{-1}(\lambda_n)W)^2) + x^2 tr((S^{-1}(\lambda_n)W^2S^{-1}(\lambda_n))^2)\\
&-2xtr(S^{-1}(\lambda_n)WS^{-1}(\lambda_n)W^2S^{-1}(\lambda_n)))
\end{aligned}$$

$$=2\sigma^4((tr(W^2)+2\lambda_n tr(W^3)+\sum_{t,s=1}^{\infty}\lambda_n^t\lambda_n^s tr(W^{t+1}W^{s+1}))+x^2tr((S^{-1}(\lambda_n)W^2S^{-1}(\lambda_n))^2)$$
$$-2xtr(W^3)-6x\lambda_n tr(W^4)-6x\lambda_n^2 tr(W^5)-2x\sum_{t,s,v=1}^{\infty}\lambda_n^t\lambda_n^s\lambda_n^v tr(W^{t+1}W^{s+1}W^{v+1})). \tag{2.A.13}$$

By choosing $x=a^{-1}\zeta$, where $a=tr(W^2)/\sqrt{2tr(W^2)}=\sqrt{2tr(W^2)}/2$ (which is (2.1.4) when W is symmetric), (2.A.12) and (2.A.13) become

$$\kappa_1=\sigma^2(-a^{-1}\zeta tr(W^2)+\lambda_n tr(W^2)+\lambda_n^2 tr(W^3)-2a^{-1}\zeta\lambda_n tr(W^3))+O\left(\sqrt{\frac{h}{n}}\right)$$

and

$$\begin{aligned}\kappa_2 &= 2\sigma^4(tr(W^2)+2\lambda_n tr(W^3)-2a^{-1}\zeta tr(W^3)+O(1))\\ &= 2\sigma^4 tr(W^2)\left(1+2\lambda_n\frac{tr(W^3)}{tr(W^2)}-2a^{-1}\zeta\frac{tr(W^3)}{tr(W^2)}+O\left(\frac{h}{n}\right)\right)\\ &= 2\sigma^4 tr(W^2)\left(1+2\lambda_n\frac{tr(W^3)}{tr(W^2)}-2\sqrt{2}\frac{tr(W^3)}{(tr(W^2))^{3/2}}\zeta+O\left(\frac{h}{n}\right)\right).\end{aligned}$$

Hence, by standard algebra,

$$\begin{aligned}\kappa_1^c &= \frac{-a^{-1}tr(W^2)\zeta+\lambda_n tr(W^2)+\lambda_n^2 tr(W^3)-2a^{-1}\zeta\lambda_n tr(W^3)+O\left(\sqrt{\frac{h}{n}}\right)}{\sqrt{2tr(W^2)}\left(1+2\lambda_n\frac{tr(W^3)}{tr(W^2)}-2\sqrt{2}\frac{tr(W^3)}{(tr(W^2))^{3/2}}\zeta+O\left(\frac{h}{n}\right)\right)^{1/2}}\\ &= \left(-\zeta+a\lambda_n+tr(W^3)\left(\frac{\lambda_n^2a^{-1}}{2}-\lambda_n a^{-2}\zeta\right)+O\left(\frac{h}{n}\right)\right)\\ &\times\left(1-\frac{tr(W^3)}{4}\left(2\lambda_n a^{-2}-a^{-3}\zeta\right)+O\left(\frac{h}{n}\right)\right)\\ &= -\zeta+a\lambda_n+\frac{tr(W^3)}{2}(\lambda_n^2a^{-1}-2\lambda_n a^{-2}\zeta)\\ &+\frac{tr(W^3)}{4}(2\lambda_n a^{-2}-a^{-3}\zeta)(\zeta-a\lambda_n)+O\left(\frac{h}{n}\right).\end{aligned} \tag{2.A.14}$$

For notational simplicity, let

$$\omega(\zeta) = \frac{tr(W^3)}{2}(\lambda_n^2a^{-1}-2\lambda_n a^{-2}\zeta)+\frac{tr(W^3)}{4}(2\lambda_n a^{-2}-a^{-3}\zeta)(\zeta-a\lambda_n).$$

Given (2.A.14) and proceeding as described in the proofs of Theorems 2.1 and 2.2, the expansion for the cdf of $\hat{\lambda}_n-\lambda_n$ under H_1 becomes

$$\begin{aligned}&Pr(a(\hat{\lambda}_n-\lambda_n)\le\zeta)=\Phi(-\kappa_1^c)-\frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c)+.....\\ =\ &\Phi(\zeta-\lambda_n a)-\omega(\zeta)\phi(\zeta-\lambda_n a)-\frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta-\lambda_n a)+O\left(\frac{h}{n}\right)\end{aligned}$$

$$= \quad \Phi(\zeta - \lambda_n a) - \omega(\zeta)\phi(\zeta - \lambda_n a) - \frac{\kappa_3^c}{3!} H_2(\zeta - \lambda_n a)\phi(\zeta - \lambda_n a) + O\left(\frac{h}{n}\right),$$

where

$$\kappa_3^c = \frac{8\sigma^6 tr(C^3)}{2^{3/2}\sigma^6 tr(C^2)} \sim \frac{2^{3/2} tr(W^3)}{(tr(W^2))^{3/2}} \sim \left(\frac{h}{n}\right)^{3/2}.$$

Proof of Theorem 2.4

The OLS estimate of λ is

$$\hat{\lambda} - \lambda = \frac{\frac{h}{n} Y' W \epsilon}{\frac{h}{n} Y' W^2 Y} = \frac{\frac{h}{n} \epsilon' S^{-1}(\lambda) W \epsilon}{\frac{h}{n} \epsilon' S^{-1}(\lambda) W^2 S^{-1}(\lambda)\epsilon}, \tag{2.A.15}$$

since W in (1.1.2) is symmetric and $Y = S^{-1}(\lambda)\epsilon$.

As regarding the numerator of the RHS of (2.A.15),

$$\begin{aligned} & \left(\frac{h}{n}\right)^2 E(\epsilon' S^{-1}(\lambda) W \epsilon - \sigma^2 tr(S^{-1}(\lambda) W))^2 \\ = & \left(\frac{h}{n}\right)^2 E(\epsilon' S^{-1}(\lambda) W \epsilon)^2 - \sigma^4 (tr(S^{-1}(\lambda) W))^2 \\ = & \ 2\sigma^4 \left(\frac{h}{n}\right)^2 tr((S^{-1}(\lambda) W)^2) \to 0 \end{aligned}$$

as $n \to \infty$, since $tr((S^{-1}(\lambda)W)^2) = O(n/h)$ by Lemma 2.1. Hence

$$\frac{h}{n}(\epsilon' S^{-1}(\lambda) W \epsilon - \sigma^2 tr(S^{-1}(\lambda) W)) \to 0$$

in second mean, implying

$$\underset{n\to\infty}{plim} \frac{h}{n} \epsilon' S^{-1}(\lambda) W \epsilon = \underset{n\to\infty}{lim} \sigma^2 \frac{h}{n} tr(S^{-1}(\lambda) W). \tag{2.A.16}$$

Similarly,

$$\underset{n\to\infty}{plim} \frac{h}{n} \epsilon' S^{-1}(\lambda) W^2 S^{-1}(\lambda) \epsilon = \underset{n\to\infty}{lim} \sigma^2 \frac{h}{n} tr((S^{-1}(\lambda) W)^2). \tag{2.A.17}$$

From (2.A.16) and (2.A.17),

$$\hat{\lambda} - \lambda \xrightarrow{p} \underset{n\to\infty}{lim} \frac{\frac{h}{n} tr(S^{-1}(\lambda) W)}{\frac{h}{n} tr((S^{-1}(\lambda) W)^2)}. \tag{2.A.18}$$

First, we show that the RHS of (2.A.18) is finite. Lemma 2.1 implies

$$\frac{h}{n} tr(S^{-1}(\lambda) W) = O(1).$$

The denominator in the RHS of (2.A.18) is non-negative and, by (2.4.2),

$$\frac{h}{n}tr((S^{-1}(\lambda)W)^2) \sim \frac{h}{n}tr(W^2),$$

which is non-zero for W given in (1.1.2), as shown in Section 2.5. Hence, the RHS of (2.A.18) is finite and its sign depends on its numerator.

From (1.1.2) and the series representation in (2.4.2),

$$tr(S^{-1}(\lambda)W) = tr(\sum_{i=0}^{\infty}\lambda^i tr(W^{i+1})) = r\sum_{i=0}^{\infty}\lambda^i tr(B_m^{i+1}).$$

Since B_m has one eigenvalue equal to 1 and the other $(m-1)$ equal to $-1/(m-1)$, we have

$$tr(B_m^{i+1}) = 1 + (m-1)\left(\frac{-1}{m-1}\right)^{i+1}$$

and hence, since $|\lambda| < 1$,

$$\begin{aligned} tr(S^{-1}(\lambda)W) &= r\sum_{i=0}^{\infty}\lambda^i\left(1-\left(\frac{-1}{m-1}\right)^i\right) \\ &= \frac{r}{1-\lambda} - \frac{r}{1+\frac{\lambda}{m-1}} = \frac{\lambda}{1-\lambda}\frac{rm}{m-1+\lambda}. \end{aligned} \tag{2.A.19}$$

By substituting $h = m-1$ and $n = mr$ into (2.A.19),

$$\frac{h}{n}tr(S^{-1}(\lambda)W) = \frac{m-1}{mr}\frac{\lambda}{1-\lambda}\frac{rm}{m-1+\lambda} = \frac{\lambda}{1-\lambda}\frac{m-1}{m-1+\lambda},$$

which has the same sign of λ, whether m is divergent or bounded, provided that $m > 1$.

Lemma 2.1 *If $w_{ij} = O(1/h)$, uniformly in i and j,*

$$tr(WA) = O\left(\frac{n}{h}\right),$$

where A is an $n \times n$ matrix so that $||A||_r + ||A||_c \leq K$.

Proof Let a_{ij} be the $(i-j)$th element of A. The $i-$th diagonal element of WA has absolute value given by

$$|(WA)_{ii}| \leq \max_j|w_{ij}|\sum_{j=1}^{n}|a_{ji}| = O\left(\frac{1}{h}\right),$$

uniformly in i. Therefore

$$|tr(WA)| \leq \sum_{i=1}^{n}|(WA)_{ii}| \leq n\max_i|(WA)_{ii}| = O\left(\frac{n}{h}\right).$$

Lemma 2.2 *Let R and S be $n \times 1$ vectors whose $i-th$ components are denoted by r_i and s_i, respectively. Let A be an $n \times n$ matrix. If, for all n,*

$$\underset{1\leq i\leq n}{max}|r_i| \leq K \qquad \underset{1\leq i\leq n}{max}|s_i| \leq K, \quad ||A||_r + ||A||_c \leq K,$$

then $|R'AS| = O(n)$.

Proof Let a_{ij} be the $(i-j)$th component of A.

$$\begin{aligned} |R'AS| \quad &= \quad |\sum_{i=1}^{n}\sum_{j=1}^{n} r_i a_{ij} s_j| \leq \sum_{i=1}^{n}\sum_{j=1}^{n}|r_i||a_{ij}||s_j| \leq \max_{1\leq i,j\leq n}|r_i||s_j|\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}| \\ &\leq K \quad \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}| \leq K \max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{ij}|\sum_{i=1}^{n}1 = O(n). \end{aligned}$$

Lemma 2.3 *Suppose model (1.2.5) and Assumptions 1-3 hold. For $\lambda_n = c(h/n)^\gamma$ with $0 < \gamma \leq 1/2$, $\hat{\lambda}_n = O_p(\lambda_n)$, while with $\gamma > 1/2$, $\hat{\lambda}_n = O_p((h/n)^{1/2})$.*

Proof We should stress that Assumption 1 could be relaxed. However, Gaussianity is assumed throughout this work and hence is retained. Here, Assumption 1 simplifies the derivation of the expectations of quadratic forms in ϵ in the following argument.

By the OLS formula,

$$\hat{\lambda}_n - \lambda_n = \frac{\frac{h}{n}Y'W'\epsilon}{\frac{h}{n}Y'W'W\epsilon} = \frac{\frac{h}{n}\epsilon'S^{-1}(\lambda_n)'W'\epsilon}{\frac{h}{n}\epsilon'S^{-1}(\lambda_n)'W'WS^{-1}(\lambda_n)\epsilon}.$$

Therefore, when $\gamma \leq 1/2$, we need to show that

$$\frac{\frac{h}{n}\epsilon'S^{-1}(\lambda_n)'W'\epsilon}{\frac{h}{n}\epsilon'S^{-1}(\lambda_n)'W'WS^{-1}(\lambda_n)\epsilon} = O_p\left(\left(\frac{h}{n}\right)^\gamma\right). \tag{2.A.20}$$

Similarly, when $\gamma > 1/2$, it suffices to show

$$\frac{\frac{h}{n}\epsilon'S^{-1}(\lambda_n)'W'\epsilon}{\frac{h}{n}\epsilon'S^{-1}(\lambda_n)'W'WS^{-1}(\lambda_n)\epsilon} = O_p\left(\left(\frac{h}{n}\right)^{1/2}\right) \tag{2.A.21}$$

and conclude the claim by observing that $O((h/n)^{1/2})$ dominates $\lambda_n = O((h/n)^\gamma)$ when $\gamma > 1/2$.

Under Assumptions 1-3, the denominator in (2.A.20) (and (2.A.21)) has a finite and positive probability limit. Indeed, let

$$V_n = \frac{h}{n}\epsilon'S^{-1}(\lambda_n)'W'WS^{-1}(\lambda_n)\epsilon - \frac{h}{n}E\left(\epsilon'S^{-1}(\lambda_n)'W'WS^{-1}(\lambda_n)\epsilon\right).$$

Then,

$$\begin{aligned}
E(V_n^2) &= \left(\frac{h}{n}\right)^2 E((\epsilon' S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)\epsilon)^2) \\
&- \left(\frac{h}{n}\right)^2 (E(\epsilon' S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)\epsilon))^2 \\
&= \left(\frac{h}{n}\right)^2 \sigma^4 (tr(S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)))^2 + 2\left(\frac{h}{n}\right)^2 \sigma^4 tr((S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n))^2) \\
&- \left(\frac{h}{n}\right)^2 \sigma^4 (tr(S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)))^2 \\
&= 2\left(\frac{h}{n}\right)^2 \sigma^4 tr((S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n))^2) = O\left(\frac{h}{n}\right),
\end{aligned}$$

where the last equality follows by Lemma 2.1, after observing that $||S^{-1}(\lambda_n)||_r + ||S^{-1}(\lambda_n)||_c \leq K$, as shown in Section 2.4. Hence, as $n \to \infty$, $E(V_n^2) \to 0$ and therefore $V_n \xrightarrow{p} 0$, i.e.

$$\underset{n\to\infty}{plim} \frac{h}{n} \epsilon' S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)\epsilon = \lim_{n\to\infty} \frac{h}{n} E(\epsilon' S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)\epsilon). \tag{2.A.22}$$

Let

$$Q = \lim_{n\to\infty} \frac{h}{n} tr(W'W).$$

Under Assumption 3 $Q > 0$. Moreover,

$$\frac{h}{n} E(\epsilon' S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)\epsilon) = \frac{h}{n}\sigma^2 tr(W'W) + O\left(\left(\frac{h}{n}\right)^\gamma\right).$$

The last displayed expression has been obtained by observing that

$$\begin{aligned}
\frac{h}{n} E(\epsilon' S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)\epsilon) &= \frac{h}{n}\sigma^2 tr(S^{-1}(\lambda_n)' W' W S^{-1}(\lambda_n)) \\
&= \sigma^2 \frac{h}{n}\left(tr(W'W) + 2\lambda_n tr((W')^2 W) + tr\left(\sum_{i=1}^{\infty}(\lambda_n W')^i W' W \sum_{j=1}^{\infty}(\lambda_n W)^j\right)\right) \\
&= \sigma^2 \frac{h}{n}\left(tr(W'W) + 2\lambda_n tr((W')^2 W) + tr\left(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\lambda_n^i \lambda_n^j W'^{i+1} W^{j+1}\right)\right) \\
&= \sigma^2 \frac{h}{n}\left(tr(W'W) + 2\lambda_n tr((W')^2 W) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\lambda_n^i \lambda_n^j tr(W'^{i+1} W^{j+1})\right) \\
&= \sigma^2 \frac{h}{n}\left(tr(W'W) + 2\lambda_n tr((W')^2 W) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\lambda_n^i \lambda_n^j O\left(\frac{n}{h}\right)\right) \\
&= \sigma^2 \frac{h}{n}\left(tr(W'W) + 2\lambda_n tr((W')^2 W) + \left(\frac{\lambda_n}{1-\lambda_n}\right)^2 O\left(\frac{n}{h}\right)\right),
\end{aligned}$$

where the fifth equality holds since $tr(W'^{i+1}W^{j+1}) = O(n/h)$ for every i, j by Lemma 2.1. The last equality is obtained by

$$\sum_{i=1}^{\infty}\lambda_n^i = \frac{\lambda_n}{1-\lambda_n},$$

since $|\lambda_n| < 1$ for n sufficiently large. Hence, from (2.A.22),

$$\underset{n\to\infty}{plim}\frac{h}{n}\epsilon' S^{-1}(\lambda_n)'W'WS^{-1}(\lambda_n)\epsilon = Q > 0. \tag{2.A.23}$$

On the other hand, the numerator in (2.A.20) is $O_p((h/n)^{\gamma})$, when $\gamma \leq 1/2$, and $O_p((h/n)^{1/2})$, when $\gamma > 1/2$, since

$$\left(\frac{h}{n}\right)^2 E((\epsilon' S^{-1}(\lambda_n)'W'\epsilon)^2) = \sigma^4\left(\frac{h}{n}\right)^2\left((tr(WS^{-1}(\lambda_n)))^2 + 2tr((WS^{-1}(\lambda_n))^2)\right)$$
$$= \sigma^4\left(\frac{h}{n}\right)^2\left((0 + tr(\sum_{i=1}^{\infty}\lambda_n^i W^{i+1}))^2 + O\left(\frac{n}{h}\right)\right), \tag{2.A.24}$$

while

$$tr(\sum_{i=1}^{\infty}\lambda_n^i W^{i+1}) = \sum_{i=1}^{\infty}\lambda_n^i tr(W^{i+1})$$
$$= \sum_{i=1}^{\infty}\lambda_n^i O\left(\frac{n}{h}\right) = \frac{\lambda_n}{1-\lambda_n}O\left(\frac{n}{h}\right) = O\left(\left(\frac{n}{h}\right)^{1-\gamma}\right). \tag{2.A.25}$$

Therefore, collecting (2.A.24) and (2.A.25),

$$\left(\frac{h}{n}\right)^2 E((\epsilon' S^{-1}(\lambda_n)'W'\epsilon)^2) = O\left(\left(\frac{h}{n}\right)^{2\gamma}\right),$$

when $\gamma \leq 1/2$, and

$$\left(\frac{h}{n}\right)^2 E((\epsilon' S^{-1}(\lambda_n)'W'\epsilon)^2) = O\left(\frac{h}{n}\right),$$

when $\gamma > 1/2$. By Markov's inequality,

$$\frac{h}{n}\epsilon' S^{-1}(\lambda_n)'W'\epsilon = O_p\left(\left(\frac{h}{n}\right)^{\gamma}\right), \tag{2.A.26}$$

when $\gamma \leq 1/2$, and

$$\frac{h}{n}\epsilon' S^{-1}(\lambda_n)'W'\epsilon = O_p\left(\left(\frac{h}{n}\right)^{1/2}\right), \tag{2.A.27}$$

when $\gamma > 1/2$.

Collecting (2.A.23), (2.A.26) and (2.A.27) the claims in (2.A.20) and (2.A.21) follow trivially.

3 Improved Test Statistics based on MLE for Pure SAR

In Chapter 2 we focused on the test of H_0 in (1.2.1) when λ in model (1.2.5) is estimated by OLS. As outlined, the OLS estimate of λ in (1.2.5) is inconsistent when $\lambda \neq 0$ and hence the results of Chapter 2 cannot be extended to test the more general null hypothesis

$$H_0 : \lambda = \lambda_0 \tag{3.0.1}$$

against the alternative

$$H_1 : \lambda > \lambda_0 \ \ (< \lambda_0) \tag{3.0.2}$$

for any fixed λ_0. More importantly, as discussed in Section 2.5, a test of H_0 in (1.2.1) might be inconsistent for some choices of W.

In this chapter we derive new tests of (1.2.1) based on $\tilde{\lambda}$ when again the data follow model (1.2.5), i.e. (1.1.3) when $\beta = 0$ a priori. As discussed in Chapter 1, $\tilde{\lambda}$ is consistent for every value of $\lambda \in (-1, 1)$ in model (1.2.5), as shown in Lee (2004). Hence, in principle we could extend the results presented in this section to test (3.0.1) against (3.0.2). Although the procedure would be identical, when $\lambda_0 \neq 0$ the algebraic burden would increase dramatically. In addition, in most of practical application, $\lambda = 0$ is probably the most interesting value one wishes to test, as discussed in the Chapter 1. Therefore, it seems reasonable to focus only on the test of H_0 as specified in (1.2.1).

Similarly to what discussed in Chapter 2 and as already outlined in Chapter 1, the rate of convergence of $\tilde{\lambda}$ can be slower than the parametric rate $\sqrt{n}$, depending on the choice of W. When this is the case, the normal cdf might not be an accurate approximation for the cdf of the t-statistic based on $\tilde{\lambda}$ under H_0. Thus, inference based on standard first order asymptotic theory can be unreliable and this provides motivation for employing instead refined statistics, based on formal Edgeworth expansions, which entail closer approximations.

In Section 3.1 we present refined tests based on both Edgeworth-corrected critical values and corrected t-statistics under H_0 in (1.2.1). In Section 3.2 we report the results of Monte Carlo simulations to assess the finite sample performance of the new tests. Finally, in Sections 3.3 and 3.4 the new tests based on both MLE and OLS estimates of λ in model (1.2.5) are applied in two empirical examples. It should be stressed that these examples are intended for illustrative purpose only and do not aim to be exhaustive analyses of the issues involved. Proofs are reported in the appendices.

3.1 Test against a one-sided alternative: Edgeworth-corrected critical values and corrected statistic

We suppose that model (1.2.5) holds and we are interested in testing (1.2.1) against (2.1.1). Extensions of the following results to testing (1.2.1) against a two-sided alternative are straightforward in principle, but algebraically very cumbersome, since the derivation or a third order Edgeworth expansion, rather than the second order one, would be necessary (similarly to what was discussed in Section 2.2). The Gaussian log-likelihood function for model (1.2.5) is given by (1.2.7) when $\beta = 0$, i.e.

$$l(\lambda, \sigma^2) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln\sigma^2 + ln(det(S(\lambda))) - \frac{1}{2\sigma^2} Y'S(\lambda)'S(\lambda)Y. \tag{3.1.1}$$

Given λ, the MLE of σ^2 is

$$\tilde{\sigma}^2(\lambda) = \frac{1}{n} Y'S(\lambda)'S(\lambda)Y \tag{3.1.2}$$

and hence

$$\tilde{\lambda} = \arg\max_{\lambda \in \Lambda} \; l(\lambda, \tilde{\sigma}^2(\lambda)),$$

where $\Lambda \in (-1, 1)$ and here λ denotes any admissible value.

When $\Lambda \in (-1, 1)$, $det(S(\lambda))$ in (3.1.1) is positive for every $\lambda \in \Lambda$. Indeed, $det(S(\lambda)) = det(I - \lambda W)$ is positive when $|\lambda| < 1/\bar{\eta}(W)$, where $\bar{\eta}(W)$ is defined in (1.5.2). On the other hand under Assumption 2(i) $\bar{\eta}(W) = 1$, as discussed in Chapter 1. Furthermore, under Assumption 2(i), existence of $S^{-1}(\lambda)$ is guaranteed from (2.4.2) provided that $|\lambda| < 1$.

We have the following result

Theorem 3.1 *Let model (1.2.5) and Assumptions 1-3 hold. The cdf of $\tilde{\lambda}$ under H_0 in (1.2.1) admits the second order formal Edgeworth expansion*

$$Pr(\tilde{a}\tilde{\lambda} \leq \zeta | H_0) = \Phi(\zeta) + \left(2\frac{tr(WW'W)}{\tilde{a}^3} + \frac{tr(W^3)}{\tilde{a}^3}\right)\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(\zeta) + o\left(\sqrt{\frac{h}{n}}\right),$$

or equivalently

$$Pr(\tilde{a}\tilde{\lambda} \leq \zeta | H_0) = \Phi(\zeta) + \left(2\frac{tr(WW'W)}{\tilde{a}^3} + \frac{tr(W^3)}{\tilde{a}^3}\right)\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}H_2(\zeta)\phi(\zeta) + o\left(\sqrt{\frac{h}{n}}\right), \tag{3.1.3}$$

where

$$\tilde{a} = \sqrt{tr(W^2 + W'W)} \tag{3.1.4}$$

and

$$\tilde{\kappa}_3^c \sim -\frac{4tr(W^3)+6tr(WW'W)}{\tilde{a}^3} \sim \sqrt{\frac{h}{n}}.$$

The proof of Theorem 3.1 is in Appendix A.1.

Under Assumption 3,

$$\left(2\frac{tr(WW'W)}{\tilde{a}^3}+\frac{tr(W^3)}{\tilde{a}^3}\right)\phi(\zeta)-\frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(\zeta) \sim \sqrt{\frac{h}{n}}.$$

It should again be stressed that the expansion in (3.1.3) is formal and hence the order of the remainder can only be conjectured by the rate of the coefficients. Without considering validity issues, the error order $o(\sqrt{h/n})$ is the sharpest one can conjecture since several approximations are used to obtain (3.1.3), as explained in detail in Appendices A.1 and A.2

Under Assumption 3 $\tilde{a}$ is finite and strictly positive for large n and hence the rate of convergence of $Pr(\tilde{a}\tilde{\lambda} \leq \zeta|H_0)$ to the standard normal cdf is slower than the usual $\sqrt{n}$ when the sequence h is divergent.

From expansion (3.1.3), Edgeworth-corrected critical values and the corrected null statistic can be obtained. The derivation is very similar to that reported in Chapter 2, Section 2.1, for the cdf of $a\hat{\lambda}$ and is omitted here. The size of the test of (1.2.1) obtained with the usual standard normal approximation

$$Pr(\tilde{a}\tilde{\lambda} > z_\alpha|H_0) \tag{3.1.5}$$

can be compared with that for the Edgeworth-corrected critical value, that is

$$Pr(\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed}|H_0), \tag{3.1.6}$$

where

$$\tilde{t}^{Ed} = z_\alpha - \left(2\frac{tr(WW'W)}{\tilde{a}^3}+\frac{tr(W^3)}{\tilde{a}^3}\right)+\frac{\tilde{\kappa}_3^c}{3!}H_2(z_\alpha).$$

As discussed in Chapter 2, when z_α is used to approximate the true quantile, we have an error of order $\sqrt{h/n}$, while the error is decreased to $o(\sqrt{h/n})$ when the Edgeworth-corrected critical value is used.

Finally, (3.1.5) can be compared with the size based on the corrected statistic, i.e.

$$Pr(\tilde{g}(\tilde{a}\tilde{\lambda}) > z_\alpha|H_0), \tag{3.1.7}$$

where

$$\tilde{g}(x) = x + 2\frac{tr(WW'W)}{\tilde{a}^3} + \frac{tr(W^3)}{\tilde{a}^3} - \frac{\tilde{\kappa}_3^c}{3!}H_2(x) + \tilde{Q}(x),$$

and

$$\tilde{Q}(x) = \left(\frac{\tilde{\kappa}_3^c}{3!}\right)^2 \frac{x^3}{3}.$$

As discussed in detail in Section 2.1, $\tilde{Q}(x)$ can be derived from (2.1.14) and is cubic so that $\tilde{g}(x)$ is strictly increasing over the whole domain, but does not affect the order of the remainder.

3.2 Bootstrap correction and Monte Carlo results

In this section we report and discuss some Monte Carlo simulations to investigate the finite sample performance of the tests derived in Section 3.1. As in Chapter 2, we adopt the Case (1991) specification for W, specified in (1.1.2). The setting of the Monte Carlo study is identical to that described in Section 2.5.

The empirical sizes of the test of (1.2.1) based on the usual normal approximation are compared with the same quantities obtained with both the Edgeworth-corrected critical values and corrected test statistics. In addition, we consider the simulated sizes based on bootstrap critical values. The bootstrap algorithm to obtain the critical values is similar to that outlined in Section 2.5. Once B pseudo-samples Y_j^*, $j = 1, ...B$, are obtained from $N(0, Y'Y/n)$, we obtain B bootstrap MLE null statistics

$$\tilde{Z}_j = \tilde{a}\tilde{\lambda}_j^*, \quad j = 1, ...B,$$

where

$$\tilde{\lambda}_j^* = \arg\max_{\lambda \in \Lambda} l_j^*(\lambda)$$

and

$$l_j^*(\lambda) = -\frac{n}{2}(ln(2\pi) + 1) - \frac{n}{2}ln(\frac{1}{n}Y_j^{*\prime}S(\lambda)'S(\lambda)Y_j^*) + ln(det(S(\lambda))).$$

The $\alpha-$percentile is computed as the value $\tilde{w}_\alpha^*$ which solves

$$\frac{1}{B}\sum_{j=1}^{B} 1(\tilde{Z}_j \leq \tilde{w}_\alpha^*) = \alpha.$$

The size of the test of (1.2.1) when the bootstrap critical value is used is then

$$Pr(\tilde{a}\tilde{\lambda} > \tilde{w}_\alpha^* | H_0). \tag{3.2.1}$$

Similarly to Section 2.5, in the Tables we denote by “normal”, “Edgeworth”, "transformation" and "bootstrap" the simulated values corresponding to the size obtained with the standard approximation, Edgeworth-corrected critical values, Edgeworth-corrected null statistic and bootstrap critical values, respectively. Also, similarly to Section 1.5, we denote by "divergent" and "bounded" h the cases where m is monotonically increased and kept fixed, respectively.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
normal	0.005	0.006	0.004	0.013
Edgeworth	0.118	0.091	0.074	0.060
transformation	0.056	0.052	0.052	0.045
bootstrap	0.058	0.052	0.054	0.046

Table 3.1: Empirical sizes of the tests of H_0 in (1.2.1) when λ in (1.2.5) is estimated by MLE and the sequence h is "divergent". The reported values have to be compared with the nominal 0.05.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
normal	0.012	0.025	0.032	0.038
Edgeworth	0.090	0.075	0.068	0.049
transformation	0.057	0.055	0.049	0.051
bootstrap	0.062	0.056	0.058	0.052

Table 3.2: Empirical sizes of the tests of H_0 in (1.2.1) when λ in (1.2.5) is estimated by MLE and the sequence h is "bounded". The reported values have to be compared with the nominal 0.05.

Tables 3.1 and 3.2 display the simulated values corresponding to (3.1.5), (3.1.6), (3.1.7) and (3.2.1) when h is "divergent" and "bounded", respectively. All the values in Tables 3.1 and 3.2 have to be compared with the nominal 5%.

For both "divergent" and "bounded" h, it is clear that the usual normal approximation does not work well in practice, since the simulated values for the size greatly underestimate the nominal 5% for all sample sizes. On the other hand, the Edgeworth-corrected results seem to perform reasonably well. Similarly to what discussed in Section 2.5, the results obtained with the Edgeworth-corrected critical values exceed the target 0.05 for very small sample sizes, but the convergence to the nominal value appears to be fast. Specifically, on average across sample sizes, the difference between the values obtained with Edgeworth-corrected critical values and the nominal 0.05 is only 19% and 21% smaller than the same quantity obtained with the standard t-statistic, h being "divergent" and "bounded", respectively. However, as sample size increases, such difference decreases at a faster rate when inference is based on corrected critical values. Indeed, the difference between actual and nominal values decreases by 46% and 53% when inference is based on Edgeworth-corrected critical values, and only by 6% and 32% when we rely on the standard statistic, h being "divergent" or "bounded", respectively. The simulated sizes based on the corrected statistics, instead, are very satisfactory also for very small sample sizes, whether h is "divergent" or "bounded". Finally, for all sample sizes, the bootstrap results appear to be very similar to ones based on the Edgeworth-corrected statistic, whether h is "divergent" or "bounded". Specifically, when h is "divergent", the values obtained by Edgeworth-corrected statistic and by bootstrap critical values are 91% and 89% closer to 0.05 than values obtained by the standard statistic. Such improvements become 87% and 71% when h is "bounded".

As expected, by comparing Tables 3.1 and 3.2 with Tables 2.1 and 2.2 (reported in Section 2.5), we notice that the results are similar whether λ is estimated by OLS or MLE. However, for "divergent" h and when considering the Edgeworth-corrected critical values, the results obtained when λ is estimated by MLE slightly outperform those based on the OLS estimate. Other than this case, the values appear to be comparable.

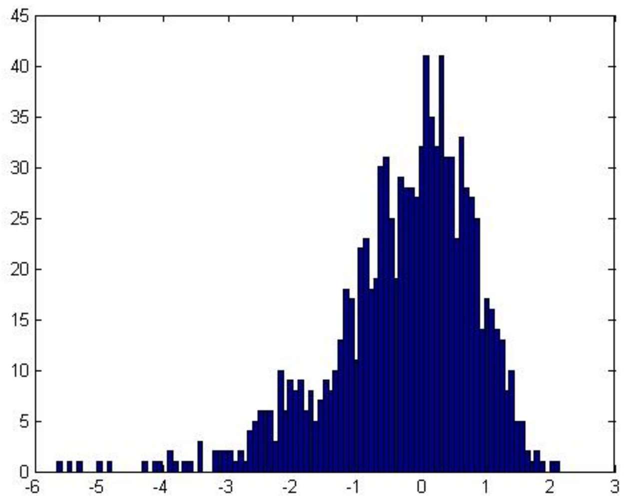

Figure 3.1: Simulated pdf of $\tilde{a}\tilde{\lambda}$ under H_0 in (1.2.1)

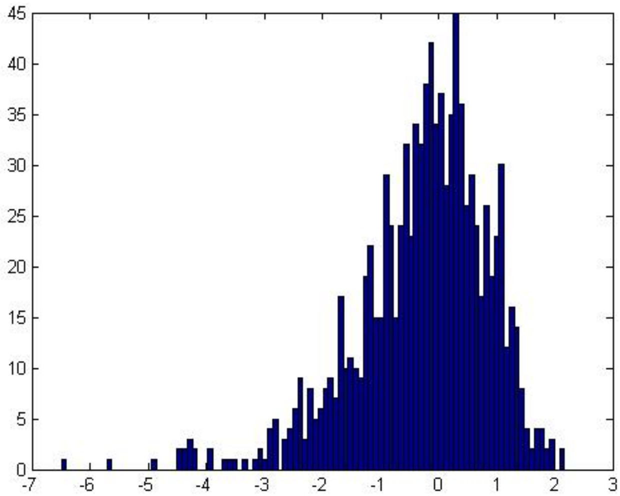

Figure 3.2: Simulated pdf of $\tilde{g}(\tilde{a}\tilde{\lambda})$ under H_0 in (1.2.1)

In Figures 3.1 and 3.2 we plot the pdf obtained from the Monte Carlo simulation of the non-corrected MLE null statistic $\tilde{a}\tilde{\lambda}$ and its corrected version $\tilde{g}(\tilde{a}\tilde{\lambda})$, respectively. We notice that the non-corrected statistic is skewed to the left but most of this skewness is removed when we consider the corrected version.

	$m = 8$ $r = 5$		$m = 12$ $r = 8$		$m = 18$ $r = 11$		$m = 28$ $r = 14$	
normal	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.0100	0.1	0.0370	0.1	0.0380	0.1	0.0560
	0.5	0.4740	0.5	0.7270	0.5	0.8640	0.5	0.8930
	0.8	0.9850	0.8	0.9990	0.8	1	0.8	1
Edgeworth	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.1270	0.1	0.1300	0.1	0.1410	0.1	0.1740
	0.5	0.7600	0.5	0.8710	0.5	0.9270	0.5	0.9750
	0.8	0.9900	0.8	1	0.8	1	0.8	1
bootstrap	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.0940	0.1	0.1220	0.1	0.1300	0.1	0.1450
	0.5	0.7480	0.5	0.8560	0.5	0.9180	0.5	0.9990
	0.8	0.9980	0.8	1	0.8	1	0.8	1

Table 3.3: Empirical powers of the tests of H_0 in (1.2.1) against H_1 in (2.5.3) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ in (1.2.5) is estimated by MLE and the sequence h is "divergent". α is set to 0.95.

	$m=5$, $r=8$		$m=5$, $r=20$		$m=5$, $r=40$		$m=5$, $r=80$	
normal	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.0510	0.1	0.1260	0.1	0.2070	0.1	0.4000
	0.5	0.7910	0.5	0.9890	0.5	0.9980	0.5	1
	0.8	1	0.8	1	0.8	1	0.8	1
Edgeworth	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.1260	0.1	0.1920	0.1	0.2720	0.1	0.4530
	0.5	0.8820	0.5	0.9950	0.5	1	0.5	1
	0.8	1	0.8	1	0.8	1	0.8	1
bootstrap	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.1140	0.1	0.1940	0.1	0.3020	0.1	0.5220
	0.5	0.8920	0.5	1	0.5	1	0.5	1
	0.8	.	0.8	1	0.8	1	0.8	1

Table 3.4: Empirical powers of the tests of H_0 in (1.2.1) against (2.5.3) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ in (1.2.5) is estimated by MLE and the sequence h is "bounded". α is set to 0.95.

In Tables 3.3 and 3.4 we report some Monte Carlo results to assess the finite sample value of the power of both standard and corrected tests of (1.2.1) against a fixed alternative hypothesis, as specified in (2.5.3). In Tables 3.3 and 3.4 we report the simulated quantities corresponding to $Pr(\tilde{a}\tilde{\lambda} > z_\alpha|H_1)$, $Pr(\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed}|H_1)$ and $Pr(\tilde{a}\tilde{\lambda} > \tilde{w}^*_\alpha|H_1)$. As in Section 2.5, we choose three different values of $\bar{\lambda}$, specifically $\bar{\lambda} = 0.1, 0.5, 0.8$. The values in Tables 3.3 and 3.4 are consistent with the empirical sizes reported in Tables 3.1-3.4. By comparison of the results in Tables 3.3 and 3.4 with Tables 2.5 and 2.6, we notice that the simulated values for the power obtained with Edgeworth-corrected critical values when λ is estimated by MLE are significantly smaller than the corresponding ones when λ is estimated by OLS. This is due to the sign of the probability limit of $(\hat{\lambda} - \lambda)$ (Theorem 2.4) when W is chosen as in (1.1.2) and does not necessarily extend to other choices of W.

3.3 Empirical evidence: the geography of happiness

In this section the corrected tests presented in Sections 2.1 and 3.1 are applied to a small empirical example based on Stanca (2009). We first shortly describe the methodology and main results in Stanca (2009) and then outline the purpose and results of our analysis. The main goal of the empirical work in Stanca (2009) is to investigate the spatial distribution of the effects of both income and unemployment on well-being for a sample of $n = 81$ countries. For the purpose of our analysis we only focus on income effects. Several specifications are considered in Stanca (2009), the three main ones being

$$P = \lambda W P + X\gamma + \epsilon, \tag{3.3.1}$$

$$P = \lambda W P + \epsilon, \tag{3.3.2}$$

$$P = X\gamma + \epsilon, \tag{3.3.3}$$

where $\epsilon \sim N(0, \sigma^2 I)$. P is the $n-$ dimensional vector of sensitivities of well-being to income in each country, X is a $n \times k$ matrix, $k = 10$, containing exogenous macroeconomic conditions, which include GDP per capita, unemployment rate, government size and trade openness. W is the usual matrix of spatial weights and more details about the choice of W will be given below.

The components of P are clearly unobservable. Stanca (2009) provides a proxy for each component of P by estimating n country-specific, micro-level linear models where well-being (denoted W^b, henceforth) is regressed on income (denoted In, henceforth) as well as unemployment status, demographic factors, social conditions, personality traits and environmental characteristics. For notational convenience we denote Z_j, for $j = 1, ...81$, the $n_1 \times k_1$ matrix of all the regressors other than income, $k_1 = 19$. The sample sizes n_1 of each country-specific analysis varies country by country, on average $n_1 = 2300$. Specifically, for each country $j = 1, ...81$,

$$W_j^b = \beta_{1,j} In_j + \beta_{2,j} Z_j + u_j, \tag{3.3.4}$$

where u_j is a normally distributed error term. Stanca (2009) chooses the OLS estimate of $\beta_{1,j}$, denoted $\hat{\beta}_{1,j}$, for $j = 1, ...81$, as proxy for each component of P.

For each individual in the sample, W^b (intended as life satisfaction) is a self reported number from 1 to 10 while income is measured by self reported deciles in the national distribution of income. The data source for the analyses in (3.3.1), (3.3.2) and (3.3.3) is the database "World Development Indicators" (World Bank (2005)). The data source for the country-specific regressions in (3.3.4) is the "World Values Survey".

The results in Stanca (2009) indicate that by estimating λ by MLE in model (3.3.2) the presence of spatial correlation is detected. However, when the macroeconomic conditions are included among the regressors, such as in specification (3.3.1), the estimate of λ becomes insignificant, suggesting that the geographical correlation is mainly explained by similar underlying macroeconomic conditions in neighbouring countries. Therefore, either specification (3.3.2) or (3.3.3) can be appropriate, as the estimate of λ in model (3.3.2) should reflect the macroeconomic similarities among countries.

By a closer inspection of the results in Stanca (2009), we notice that the estimates of the relevant components of γ in (3.3.3) are strongly significant (1% or 0.5% level), while the estimate of λ in specification (3.3.2) is barely significant at 5%. Given that specifications (3.3.2) and (3.3.3) should both be appropriate, in principle we would expect the estimates of the relevant coefficients of the two specifications to be equally significant (at least roughly). Therefore, it can be useful to investigate whether an Edgeworth-corrected test gives a different result.

We only consider a sub-sample of the 43 European countries, rather than the 81 worldwide ones considered in Stanca (2009). Since P in specification (3.3.2) is a vector of estimates and not actual data, some heterogeneity issues might be eliminated by considering only European countries. Indeed, we expect that the micro-level analysis to obtain $\hat{\beta}_{1,j}$, $j = 1, ...43$, does not exhibit strong structural differences across a sample of 43 European countries. On the other hand, when considering a broader sample, some systematic differences in the relationship among the country-specific variables might occur. In practice, (3.3.4) might not be the correct specification for all countries, when such countries are very heterogeneous. In turn, when such differences occur, the reliability of $\hat{\beta}_{1,j}$ as proxies for the components of P is not clear. This problem might be reduced by considering only a sub-sample of less heterogeneous countries.

Since the dependent variable in (3.3.2) is a vector of proxies and not actual data, we acknowledge that the corrections derived in Sections 2.1 and 3.1 do not fully hold. In principle, we might be neglecting some relevant term arising from the approximation of the components of P by $\hat{\beta}_{1,j}$, $j = 1, ...43$, in the Edgeworth expansions of the cdf of the OLS and MLE statistics under H_0 in (1.2.1). However, at least for illustrative purpose, we think that a preliminary investigation of the effects of the inclusion of the small sample corrections derived in Sections 2.1 and 3.1 is worthwhile.

The choice of W is not described in Stanca (2009). We construct W based on a contiguity criterion, i.e. $w_{ij} = 1$ if country i and country j share a border and $w_{ij} = 0$ otherwise.

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$
$a\hat{\lambda} > z_\alpha$	reject H_0 $(1.713 > 1.645)$	fail to reject H_0 $(1.713 < 2.326)$
$a\hat{\lambda} > t^{Ed}$	reject H_0 $(1.713 > 1.287)$	reject H_0 $(1.713 > 1.666)$

Table 3.5: Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.3.2) is estimated by OLS

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$
$\tilde{a}\tilde{\lambda} > z_\alpha$	reject H_0 $(2.869 > 1.645)$	reject H_0 $(2.869 > 2.326)$
$\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed}$	reject H_0 $(2.869 > 1.429)$	reject H_0 $(2.869 > 1.922)$

Table 3.6: Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.3.2) is estimated by MLE

In Tables 3.5 and 3.6, we report the outcome of the tests of (1.2.1) when λ is estimated by OLS and MLE, respectively. The actual values of statistics and critical values are reported in brackets. When λ is estimated by OLS, $\hat{\lambda}$ is only (barely) significant at 5%, while it becomes significant at 1% when corrected critical values are used. We notice that in case λ is estimated by MLE, the outcome of the test does not change when corrected critical values are used. This is a result that could be expected, to some extent. From the simulation work, the non-corrected results for the MLE appear to be slightly better than OLS in very small samples.

3.4 Empirical evidence: the distribution of crimes in Italian provinces

The second example we consider to assess the practical performance of the new tests derived in Sections 2.1 and 3.1 is based on a paper by Buonanno et al. (2009) and deals with crime rates in Italian provinces. In particular, Buonanno et al. (2009) aim to investigate whether social capital, intended as civic norms and associational networks, affects the property crime rate at a provincial level. The 103 Italian provinces are especially suitable for this purpose since Italy displays significant provincial disparities despite being politically, ethically and religiously quite homogeneous. The literature about the influence of social capital on crime rate is broad and a survey is beyond the scope of this example. Similarly, for a discussion about the peculiar contribution of Buonanno et al. (2009), we refer to the paper.

For the purpose of our investigation, we consider the following three models

$$Y = \lambda WY + \epsilon, \tag{3.4.1}$$

$$Y = \lambda WY + \beta_1 SC + \beta_2 X + \delta D + \epsilon, \tag{3.4.2}$$

and

$$Y = \beta_1 SC + \beta_2 X + \delta D + \epsilon, \tag{3.4.3}$$

where $\epsilon \sim N(0, \sigma^2 I)$. Y is the $n-$dimensional vector of crime rates in each province, where $n = 103$. Each component of Y is obtained by dividing the reported crime rate at provincial level by the corresponding overall report rate at regional level. The dataset, originally constructed by Buonanno et al. and mainly based on ISTAT ("Istituto Nazionale di Statistica") records, contains three sets of observations, regarding car thefts, robberies and general thefts rates. SC is the vector of social capital observations. Buonanno et al. (2009) proposes four different measures of social capital, which are used separately, namely the number of recreational associations, voluntary associations, referenda turnout and blood donation. X is a $n \times k$ matrix of exogenous regressors, with $k = 8$, containing deterrence (such as the average length of judicial process and the crime specific clear up rate), demographic and socio-economic variables. In addition, X contains a measure of criminal association at provincial level. Finally, D is a matrix of geographical dummies to control for heterogeneity among the north, centre and south of the country. Our analysis is conducted with and without the inclusion of the geographical dummies and the results do not appear to vary significantly. The data pertain to 2002 or, when an average is considered, to the period 2000-2002.

In Buonanno et al. (2009) the parameters in model (3.4.2) are estimated for each crime type, with different variants of W and measures of social capital. Details of the estimation methods used are not provided in the paper. The results in Buonanno et al. (2009) indicate that the estimate of λ in model (3.4.2) is insignificant in each of the regressions considered (or only barely significant at 10%, in few cases). However, we observe that when we estimate λ in model (3.4.1) we detect spatial correlation, suggesting that the effect of geographical contiguity is mostly taken into account by the regressors included in model (3.4.2). Hence, both models (3.4.1) and (3.4.3) seem to be appropriate and we expect the estimate of λ in model (3.4.1) to reflect the overall similarities across neighbouring provinces.

For the purpose of out analysis, in order to investigate more specifically which are the main determinants of Y, we perform an OLS estimation of the parameters in model (3.4.3) and observe that Y is strongly affected by the measure of criminal association (denoted CA, henceforth). Indeed, the estimate of the coefficient of CA is significant at 0.5% level. In turn, we expect that CA displays significant correlation across provinces and to confirm our conjecture we estimate the spatial parameter μ of the additional model

$$CA = \mu WCA + \epsilon. \tag{3.4.4}$$

As expected, the estimate of μ is strongly significant (0.5% level) when inference is based on the normal approximation.

When regressors are not included, such as in (3.4.1), we would expect to detect a similarly strong spatial correlation in the dependent variable. However, the estimate of λ in (3.4.1) is only significant at 5% level, when inference is based on the normal approximation.

As discussed for the previous example, we investigate whether we obtain a different outcome of the test of (1.2.1) by including the small sample corrections derived in Sections 2.1 and 3.1. We report the results obtained for the robberies rates, W defined by a contiguity criterion as described in Section 3.3 (the same choice of W is adopted in Buonanno et al. (2009)), and blood donation as a measure of social capital, although similar results can be derived for the other crime rates and alternative measures of social capital.

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
$a\hat{\lambda} > z_\alpha$	reject H_0 $(1.9998 > 1.645)$	fail to reject H_0 $(1.9998 < 2.326)$	fail to reject H_0 $(1.9998 < 2.5776)$
$a\hat{\lambda} > t^{Ed}$	reject H_0 $(1.9998 > 1.4042)$	reject H_0 $(1.9998 > 1.8821)$	fail to reject H_0 $(1.9998 < 2.0410)$

Table 3.7: Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.4.1) is estimated by OLS

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
$\tilde{a}\tilde{\lambda} > z_\alpha$	reject H_0 $(2.2934 > 1.645)$	fail to reject H_0 $(2.2934 < 2.326)$	fail to reject H_0 $(2.2934 < 2.5776)$
$\tilde{a}\tilde{\lambda} > \tilde{t}^{ed}$	reject H_0 $(2.2934 > 1.5227)$	reject H_0 $(2.2934 > 2.0767)$	reject H_0 $(2.2934 > 2.2704)$

Table 3.8: Outcomes of the tests of H_0 in (1.2.1) when λ in model (3.4.1) is estimated by MLE

The outcomes of the tests of H_0 in (1.2.1) when λ is estimated by OLS and MLE are reported in Tables 3.7 and 3.8, respectively. We notice that when the usual normal approximation is adopted, we are able to reject H_0 only at 5% level, λ being estimated by either OLS or MLE. Instead, when the Edgeworth correction is included, we are able to reject H_0 at 1% level when λ is estimated by OLS and at 0.5% level when λ is estimated by MLE. As is the case in the previous example, these results confirm those of the simulation work, i.e. for small/moderate sample sizes, the results obtained when λ is estimated by MLE slightly outperform those obtained by OLS estimation.

A Appendix

A.1 Proof of Theorem 3.1

We first introduce some notation that will be used throughout the proof. We write

$$l(\lambda) = l(\lambda, \tilde{\sigma}^2(\lambda)),$$

where $l(\lambda, \sigma^2)$ and $\tilde{\sigma}^2(\lambda)$ are defined in (3.1.1) and (3.1.2), respectively. Define also

$$Z^{(1)}(\lambda) = \sqrt{\frac{h}{n}}\frac{\partial l(\lambda)}{\partial \lambda}, \qquad Z^{(2)}(\lambda) = \sqrt{\frac{h}{n}}\left(\frac{\partial^2 l(\lambda)}{\partial \lambda^2} - E\left(\frac{\partial^2 l(\lambda)}{\partial \lambda^2}\right)\right),$$

$$J(\lambda) = \frac{h}{n}\frac{\partial^3 l(\lambda)}{\partial \lambda^3}, \qquad K(\lambda) = -\frac{h}{n}E\left(\frac{\partial^2 l(\lambda)}{\partial \lambda^2}\right), \qquad \frac{\partial l(0)}{\partial \lambda} = \frac{\partial l(\lambda)}{\partial \lambda}|_{\lambda=0}.$$

Finally, $O_e(.)$ indicates an exact rate (in probability). In order to establish whether the orders of the coefficients appearing in Theorem 3.1 hold as exact rates, it is relevant here to distinguish $O_e(.)$ from $O_p(.)$.

By (3.1.1),

$$\frac{\partial l(\lambda)}{\partial \lambda} = n\frac{(Y'WY - \lambda Y'W'WY)}{Y'S(\lambda)'S(\lambda)Y} - tr(S^{-1}(\lambda)W) \tag{3.A.1}$$

and

$$\begin{aligned}\frac{\partial^2 l(\lambda)}{\partial \lambda^2} = &-n\frac{Y'W'WY}{Y'S(\lambda)'S(\lambda)Y} + 2n\frac{(\lambda Y'W'WY - Y'WY)^2}{(Y'S(\lambda)'S(\lambda)Y)^2} \\ &- tr(S^{-1}(\lambda)WS^{-1}(\lambda)W).\end{aligned} \tag{3.A.2}$$

Therefore, under H_0,

$$Z^{(1)}(0) = \sqrt{hn}\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \tag{3.A.3}$$

and

$$Z^{(2)}(0) = \sqrt{\frac{h}{n}}\{-n\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon} + 2n\left(\frac{\epsilon'W\epsilon}{\epsilon'\epsilon}\right)^2 - tr(W^2) + nE\left(\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon}\right)$$

$$
\begin{aligned}
& - 2nE\left(\frac{\epsilon'W\epsilon}{\epsilon'\epsilon}\right)^2 + tr(W^2)\} \\
&= \sqrt{\frac{h}{n}}\{-n\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon} + 2n\left(\frac{\epsilon'W\epsilon}{\epsilon'\epsilon}\right)^2 - tr(W^2) + n\frac{E(\epsilon'W'W\epsilon)}{E(\epsilon'\epsilon)} \\
& - 2n\frac{E\left(\frac{1}{2}\epsilon'(W+W')\epsilon\right)^2}{E(\epsilon'\epsilon)^2} + tr(W^2)\} \\
&= \sqrt{\frac{h}{n}}\{-n\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon} + 2n\frac{(\epsilon'W\epsilon)^2}{(\epsilon'\epsilon)^2} + tr(W'W) \\
& - \frac{1}{n}tr((W+W')^2)(1+\frac{2}{n})^{-1}\}, \qquad (3.A.4)
\end{aligned}
$$

since

$$E(\epsilon'W'W\epsilon) = \sigma^2 tr(W'W), \qquad (3.A.5)$$

$$E((\epsilon'(W+W')\epsilon)^2) = 2\sigma^4 tr((W+W')^2) \qquad (3.A.6)$$

and

$$E((\epsilon'\epsilon)^2) = \sigma^4(n^2+2n). \qquad (3.A.7)$$

The second equality in (3.A.4) follows because both the ratios

$$\frac{\epsilon'W\epsilon}{\epsilon'\epsilon} = \frac{\frac{1}{2}\epsilon'(W+W')\epsilon}{\epsilon'\epsilon} \quad \text{and} \quad \frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon}$$

are independent of their own denominator and therefore the expectation of the ratio is equal to the ratio of the expectations (Pitman (1937)). Similarly,

$$J(0) = \frac{h}{n}\left(\frac{-6n\epsilon'W\epsilon\epsilon'W'W\epsilon}{(\epsilon'\epsilon)^2} + \frac{8n(\epsilon'W\epsilon)^3}{(\epsilon'\epsilon)^3} - 2tr(W^3)\right) \qquad (3.A.8)$$

and, using (3.A.5), (3.A.6), (3.A.7),

$$
\begin{aligned}
K(0) &= -\frac{h}{n}\left(-n\frac{E(\epsilon'W'W\epsilon)}{E(\epsilon'\epsilon)} + 2n\frac{E\left(\epsilon'\frac{1}{2}(W+W')\epsilon\right)^2}{E(\epsilon'\epsilon)^2}\right) + \frac{h}{n}tr(W^2) \\
&= \frac{h}{n}tr(W^2) + \frac{h}{n}tr(W'W) - \frac{h}{n^2}tr((W+W')^2)\left(1+\frac{2}{n}\right)^{-1}. \qquad (3.A.9)
\end{aligned}
$$

By Lemmas 3.1, 3.2 and 3.3 (reported in Appendix A.2) $Z^{(1)}(0) = O_e(1)$, $Z^{(2)}(0) = O_p(1)$ and $J(0) = O_p(1)$, respectively. In addition, under Assumption 3, $K(0)$ is finite and positive for large n.

By the Mean Value Theorem,

$$0 = \frac{h}{n}\frac{\partial l(\tilde{\lambda})}{\partial\lambda} = \frac{h}{n}\frac{\partial l(0)}{\partial\lambda} + \frac{h}{n}\frac{\partial^2 l(0)}{\partial\lambda^2}\tilde{\lambda} + \frac{1}{2}\frac{h}{n}\frac{\partial^3 l(0)}{\partial\lambda^3}\tilde{\lambda}^2 + \frac{h}{6n}\frac{\partial^4 l(\bar{\lambda})}{\partial\lambda^2}\tilde{\lambda}^3,$$

where $\bar{\lambda}$ is an intermediate point between $\tilde{\lambda}$ and 0. Therefore,

$$0 = \sqrt{\frac{h}{n}}Z^{(1)}(0) + \sqrt{\frac{h}{n}}Z^{(2)}(0)\tilde{\lambda} - K(0)\tilde{\lambda} + \frac{1}{2}J(0)\tilde{\lambda}^2 + \frac{h}{6n}\frac{\partial^4 l(\bar{\lambda})}{\partial\lambda^4}\tilde{\lambda}^3$$

and rearranging,

$$\sqrt{\frac{n}{h}}\tilde{\lambda} = \frac{Z^{(1)}(0)}{K(0)} + \frac{Z^{(2)}(0)}{K(0)}\tilde{\lambda} + \frac{1}{2}\sqrt{\frac{n}{h}}\frac{J(0)}{K(0)}\tilde{\lambda}^2 + \frac{1}{6}\sqrt{\frac{h}{n}}\frac{\partial^4 l(\bar{\lambda})}{\partial\lambda^4}\tilde{\lambda}^3. \quad (3.A.10)$$

The first term of the RHS of (3.A.10) is $O_e(1)$, the second and the third are $O_p(\sqrt{h/n})$, since it is known that $\tilde{\lambda} = O_e(\sqrt{h/n})$ (see Lee (2004)) while $Z^{(2)}(0)$ and $J(0)$ are $O_p(1)$, by Lemma 3.2 and Lemma 3.3, respectively. The last term is $o_p(\sqrt{h/n})$ since $\bar{\lambda} \xrightarrow{p} 0$ and $\partial^4 l(0)/\partial\lambda^4 \sim tr(W^4) \sim (n/h)$. Hence,

$$\sqrt{\frac{n}{h}}\tilde{\lambda} = \frac{Z^1(0)}{K(0)} + \sqrt{\frac{h}{n}}\frac{Z^{(2)}(0)Z^{(1)}(0)}{K(0)^2} + \frac{1}{2}\sqrt{\frac{h}{n}}\frac{J(0)(Z^{(1)}(0))^2}{K(0)^3} + o_p\left(\sqrt{\frac{h}{n}}\right),$$

where the last displayed expression has been obtained by substituting into (3.A.10) the approximation for $\tilde{\lambda}$ implicit in (3.A.10), i.e.

$$\tilde{\lambda} \sim \sqrt{\frac{h}{n}}\frac{Z^{(1)}(0)}{K(0)}.$$

Let x be any finite real number. We have

$$\begin{aligned} & Pr(\sqrt{\frac{n}{h}}\tilde{\lambda} \le x) \\ = & Pr(\frac{Z^1(0)}{K(0)} + \sqrt{\frac{h}{n}}\frac{Z^{(2)}(0)Z^{(1)}(0)}{K(0)^2} + \frac{1}{2}\sqrt{\frac{h}{n}}\frac{J(0)(Z^{(1)}(0))^2}{K(0)^3} + o_p\left(\sqrt{\frac{h}{n}}\right) \le x) \\ = & Pr(\frac{1}{K(0)}\sqrt{\frac{h}{n}}\frac{\epsilon' W\epsilon}{\frac{1}{n}\epsilon'\epsilon} + \sqrt{\frac{h}{n}}\frac{Z^{(2)}(0)Z^{(1)}(0)}{K(0)^2} \\ + & \frac{1}{2}\sqrt{\frac{h}{n}}\frac{J(0)(Z^{(1)}(0))^2}{K(0)^3} + o_p\left(\sqrt{\frac{h}{n}}\right) \le x), \end{aligned}$$

where the last equality is obtained by substituting (3.A.3) and multiplying both numerator and denominator of the first term by $1/n$. We write

$$\begin{aligned} \tilde{f} = & \sqrt{\frac{h}{n}}\epsilon' W\epsilon - x\frac{K(0)}{n}\epsilon'\epsilon + \sqrt{\frac{h}{n}}\frac{Z^{(2)}(0)Z^{(1)}(0)}{K(0)}\frac{1}{n}\epsilon'\epsilon \\ & + \frac{1}{2}\sqrt{\frac{h}{n}}\frac{J(0)(Z^{(1)}(0))^2}{K(0)^2}\frac{1}{n}\epsilon'\epsilon + o_p\left(\sqrt{\frac{h}{n}}\right) \end{aligned}$$

$$= \frac{1}{2}\epsilon'(\tilde{C}+\tilde{C}')\epsilon + \sqrt{\frac{h}{n}}\frac{Z^{(2)}(0)Z^{(1)}(0)}{K(0)}\frac{1}{n}\epsilon'\epsilon + \frac{1}{2}\sqrt{\frac{h}{n}}\frac{J(0)(Z^{(1)}(0))^2}{K(0)^2}\frac{1}{n}\epsilon'\epsilon + o_p\left(\sqrt{\frac{h}{n}}\right), \tag{3.A.11}$$

where

$$\tilde{C} = \sqrt{\frac{h}{n}}W - x\frac{K(0)}{n}I. \tag{3.A.12}$$

Therefore, by standard algebraic manipulation,

$$Pr(\sqrt{\frac{n}{h}}\tilde{\lambda} \le x) = Pr(\tilde{f} \le 0).$$

Under Assumption 3 and by a slight modification of the argument in Lemma 3.1 the first term of the RHS of (3.A.11) is $O_e(1)$. The second and the third terms are both $O_p(\sqrt{h/n})$ by Lemmas 3.1, 3.2 and 3.3, and since $K(0)$ is finite and positive in the limit.

Under Assumption 1 the characteristic function of $\tilde{f}$ can be written as

$$\begin{aligned}
E(e^{it\tilde{f}}) =& \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\Re^n} e^{it\tilde{f}}e^{-\frac{u'u}{2\sigma^2}}du \\
=& \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\Re^n} e^{\frac{1}{2}itu'(\tilde{C}+\tilde{C}')u}\{1 + it\sqrt{\frac{h}{n}}\frac{Z^{(2)}(0)Z^{(1)}(0)}{K(0)}\frac{1}{n}u'u \\
&+ \frac{1}{2}it\sqrt{\frac{h}{n}}\frac{J(0)(Z^{(1)}(0))^2}{K(0)^2}\frac{1}{n}u'u + o_p\left(\sqrt{\frac{h}{n}}\right)\} \times e^{-\frac{u'u}{2\sigma^2}}du,
\end{aligned}$$

where, from (3.A.3), (3.A.4) and (3.A.8), it is clear than in $Z^{(1)}(0)$, $Z^{(2)}(0)$ and $J(0)$ appearing in the integrand function of the last displayed expression, are functions of u. Next,

$$\begin{aligned}
E(e^{it\tilde{f}}) =& \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\Re^n} e^{-\frac{1}{2\sigma^2}u'(I-it\sigma^2(\tilde{C}+\tilde{C}'))u}du \\
&+ it\sqrt{\frac{h}{n}}\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{K(0)}\int_{\Re^n} e^{-\frac{1}{2\sigma^2}u'(I-it\sigma^2(\tilde{C}+\tilde{C}'))u}\frac{Z^{(1)}(0)Z^{(2)}(0)u'u}{n}du \\
&+ \frac{1}{2}it\sqrt{\frac{h}{n}}\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{K(0)^2}\int_{\Re^n} e^{-\frac{1}{2\sigma^2}u'(I-it\sigma^2(\tilde{C}+\tilde{C}'))u}\frac{(Z^{(1)}(0))^2J(0)u'u}{n}du + o\left(\sqrt{\frac{h}{n}}\right).
\end{aligned}$$

Let

$$\Sigma = (I - it\sigma^2(\tilde{C}+\tilde{C}')). \tag{3.A.13}$$

By the change of variable

$$u \to v = \Sigma^{1/2}u, \tag{3.A.14}$$

$$
\begin{aligned}
E(e^{it\tilde{f}}) &= det(I - it\sigma^2(\tilde{C}+\tilde{C}'))^{-1/2}\frac{1}{\sqrt{2\pi\sigma^2}} \\
&\times\{\int_{\Re^n} e^{-\frac{v'v}{2\sigma^2}}(1+it\sqrt{\frac{h}{n}}\frac{1}{K(0)}\frac{Z^{(1)}(0)Z^{(2)}(0)v'\Sigma^{-1}v}{n}dv \\
&+\frac{1}{2}it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}\frac{(Z^{(1)}(0))^2J(0)v'\Sigma^{-1}v}{n})dv\} + o\left(\sqrt{\frac{h}{n}}\right) \\
&=\prod_{j=1}^{n}(1-it\sigma^2\eta_j(\tilde{C}+\tilde{C}'))^{-1/2}\{1+it\sqrt{\frac{h}{n}}\frac{1}{K(0)}E\left(\frac{Z^{(1)}(0)Z^{(2)}(0)\epsilon'\Sigma^{-1}\epsilon}{n}\right) \\
&+\frac{1}{2}it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}E\left(\frac{(Z^{(1)}(0))^2J(0)\epsilon'\Sigma^{-1}\epsilon}{n}\right)\} + o\left(\sqrt{\frac{h}{n}}\right), \qquad (3.A.15)
\end{aligned}
$$

where $det(I-it\sigma^2(\tilde{C}+\tilde{C}'))^{-1/2}$ is the Jacobian of the transformation in (3.A.14) and $\eta_j(\tilde{C}+\tilde{C}'), j = 1....n$, are the eigenvalues of $(\tilde{C}+\tilde{C}')$. It should be stressed that, after the transformation in (3.A.14), $Z^{(1)}(0)$, $Z^{(2)}(0)$ and $J(0)$ in the expectations displayed in (3.A.15) are functions of $V = \Sigma^{-1/2}\epsilon$ instead of ϵ only.

For notational simplicity, let

$$Q = Q_1 + Q_2 + o\left(\sqrt{\frac{h}{n}}\right),$$

where

$$Q_1 = it\sqrt{\frac{h}{n}}\frac{1}{K(0)}E\left(\frac{Z^{(1)}(0)Z^{(2)}(0)\epsilon'\Sigma^{-1}\epsilon}{n}\right)$$

and

$$Q_2 = \frac{1}{2}it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}E\left(\frac{(Z^{(1)}(0))^2J(0)\epsilon'\Sigma^{-1}\epsilon}{n}\right).$$

From (3.A.15) the cumulant generating function for $\tilde{f}$, denoted $\tilde{\psi}(t)$, can be written as

$$
\begin{aligned}
\tilde{\psi}(t) &= -\frac{1}{2}\sum_{j=1}^{n} ln(1-it\sigma^2\eta_j(\tilde{C}+\tilde{C}')) + ln(1+Q) \\
&= \frac{1}{2}\sum_{s=1}^{\infty}\frac{(it\sigma^2)^s}{s}tr((\tilde{C}+\tilde{C}')^s) + \sum_{s=1}^{\infty}\frac{(-1)^{s+1}}{s}Q^s. \qquad (3.A.16)
\end{aligned}
$$

Let $\tilde{\kappa}_s$ be the sth cumulant of f. The contributions of the first term of the RHS of (3.A.16) to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ are given by

$$\sigma^2 tr(\tilde{C}) = -\sigma^2 xK(0), \qquad (3.A.17)$$

$$\frac{\sigma^4}{2}tr((\tilde{C}+\tilde{C}')^2)) = \frac{h}{n}\sigma^4(tr(W^2+W'W)) + O\left(\frac{1}{n}\right) \qquad (3.A.18)$$

and

$$\sigma^6 tr(\tilde{C}+\tilde{C}')^3 = \sigma^6 \left(\frac{h}{n}\right)^{3/2} \left(2tr(W^3) + 6tr(W^2W')\right) + o\left(\sqrt{\frac{h}{n}}\right), \tag{3.A.19}$$

respectively. The contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ of the second term of the RHS of (3.A.16) are evaluated in Appendix A.2. Collecting (3.A.17), (3.A.18), (3.A.19) and the results in Appendix A.2,

$$\tilde{\kappa}_1 = -\sigma^2 xK(0) - 2\sigma^2 \left(\frac{h}{n}\right)^{3/2} \frac{tr(WW'W)}{K(0)} - \sigma^2 \left(\frac{h}{n}\right)^{3/2} \frac{tr(W^3)}{K(0)} + o\left(\sqrt{\frac{h}{n}}\right), \tag{3.A.20}$$

$$\tilde{\kappa}_2 = \sigma^4 \frac{h}{n}(tr(W^2 + W'W)) + o\left(\sqrt{\frac{h}{n}}\right) \tag{3.A.21}$$

and

$$\tilde{\kappa}_3 = -4\sigma^6 \left(\frac{h}{n}\right)^{3/2} tr(W^3) - 6\sigma^6 \left(\frac{h}{n}\right)^{3/2} tr(WW'W) + o\left(\sqrt{\frac{h}{n}}\right). \tag{3.A.22}$$

By centring and scaling the statistic $\tilde{f}$,

$$\tilde{f}^c = \frac{\tilde{f} - \tilde{\kappa}_1}{\tilde{\kappa}_2^{1/2}},$$

the cumulant generating function of $\tilde{f}^c$ can be written as

$$\tilde{\psi}^c(t) = -\frac{1}{2}t^2 + \sum_{s=3}^{\infty} \frac{\tilde{\kappa}_s^c (it)^s}{s!}, \tag{3.A.23}$$

where $\tilde{\kappa}_s^c = \tilde{\kappa}_s/\tilde{\kappa}_2^{s/2}$. From (3.A.23), the characteristic function of $\tilde{f}^c$ becomes

$$\begin{aligned}
E(e^{it\tilde{f}^c}) &= e^{-\frac{1}{2}t^2} \exp\{\sum_{s=3}^{\infty} \frac{\tilde{\kappa}_s^c (it)^s}{s!}\} = \\
&= e^{-\frac{1}{2}t^2}\{1 + \sum_{s=3}^{\infty} \frac{\tilde{\kappa}_s^c (it)^s}{s!} + \frac{1}{2!}(\sum_{s=3}^{\infty} \frac{\tilde{\kappa}_s^c (it)^s}{s!})^2 + \frac{1}{3!}(\sum_{s=3}^{\infty} \frac{\tilde{\kappa}_s^c (it)^s}{s!})^3 + ...\} \\
&= e^{-\frac{1}{2}t^2}\{1 + \frac{\tilde{\kappa}_3^c (it)^3}{3!} + \frac{\tilde{\kappa}_4^c (it)^4}{4!} + \frac{\tilde{\kappa}_5^c (it)^5}{5!} + \{\frac{\tilde{\kappa}_6^c}{6!} + \frac{(\tilde{\kappa}_3^c)^2}{(3!)^2}\}(it)^6 + ...\}.
\end{aligned}$$

Thus, by the Fourier inversion formula,

$$Pr(\tilde{f}^c \le z) = \int_{-\infty}^{z} \phi(z)dz + \frac{\tilde{\kappa}_3^c}{3!}\int_{-\infty}^{z} H_3(z)\phi(z)dz + \frac{\tilde{\kappa}_4^c}{4!}\int_{-\infty}^{z} H_4(z)\phi(z)dz + ...$$

Collecting the results derived above,

$$Pr(\sqrt{\frac{n}{h}}\tilde{\lambda} \leq x) = Pr(\tilde{f} \leq 0) = Pr(\tilde{f}^c\tilde{\kappa}_2^{1/2} + \tilde{\kappa}_1 \leq 0) = Pr(\tilde{f}^c \leq -\tilde{\kappa}_1^c)$$
$$= \Phi(-\tilde{\kappa}_1^c) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(-\tilde{\kappa}_1^c) + \frac{\tilde{\kappa}_4^c}{4!}\Phi^{(4)}(-\tilde{\kappa}_1^c) + ... \quad (3.A.24)$$

Now, from (3.A.20) and (3.A.21),

$$\begin{aligned}
\tilde{\kappa}_1^c &= \frac{-\sigma^2 xK(0) - 2\sigma^2\left(\frac{h}{n}\right)^{3/2}\frac{tr(WW'W)}{K(0)} - \sigma^2\left(\frac{h}{n}\right)^{3/2}\frac{tr(W^3)}{K(0)}}{\sigma^2(\frac{h}{n}(tr(W^2+W'W)))^{1/2}} + o\left(\sqrt{\frac{h}{n}}\right) \\
&= \frac{-x\frac{h}{n}(tr(W^2+W'W)) - 2\sqrt{\frac{h}{n}}\frac{tr(WW'W)}{(tr(W^2+W'W))} - \sqrt{\frac{h}{n}}\frac{tr(W^3)}{(tr(W^2+W'W))}}{(\frac{h}{n}(tr(W^2+W'W)))^{1/2}} \\
&+ o\left(\sqrt{\frac{h}{n}}\right),
\end{aligned}$$

where the second equality has been obtained by substituting

$$K(0) = \frac{h}{n}(tr(W^2+W'W)) + O\left(\frac{1}{n}\right),$$

according to (3.A.9). We set $x = \sqrt{n/h}\tilde{a}^{-1}\zeta$, where

$$\tilde{a} = \sqrt{tr(W^2+W'W)}.$$

Therefore

$$\tilde{\kappa}_1^c = -\zeta - 2\frac{tr(WW'W)}{\tilde{a}^3} - \frac{tr(W^3)}{\tilde{a}^3} + o\left(\sqrt{\frac{h}{n}}\right)$$

and, from (3.A.21) and (3.A.22),

$$\tilde{\kappa}_3^c \sim \frac{-4tr(W^3) - 6tr(WW'W)}{(tr(W^2+W'W))^{3/2}} = -\frac{4tr(W^3)+6tr(WW'W)}{\tilde{a}^3} \sim \sqrt{\frac{h}{n}}.$$

By Taylor expansion of the function $\Phi(-\tilde{\kappa}_1^c)$ in (3.A.24),

$$Pr(\tilde{a}\tilde{\lambda} \leq \zeta) = \Phi(\zeta) + \left(2\frac{tr(WW'W)}{\tilde{a}^3} + \frac{tr(W^3)}{\tilde{a}^3}\right)\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(\zeta) + o\left(\sqrt{\frac{h}{n}}\right).$$

A.2 Auxiliary results

In this appendix we will present and prove some of the auxiliary results used in the proof of Theorem 3.1. As already stressed, the expansion in Theorem 3.1 is formal, so we do not deal with convergence issues in some of the results that follow. Moreover, it must be mentioned that for notational simplicity, we prove Lemmas 3.1, 3.2 and

3.3 for a symmetric W. When W is not symmetric the same results hold simply by substituting $(W+W')/2$ instead of W where necessary.

Lemma 3.1 *Under Assumptions 1-3*

$$Z^{(1)}(0)=\sqrt{hn}\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}=O_e(1).$$

Proof We have

$$E\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 = \frac{E(\epsilon' W\epsilon)^2}{E(\epsilon'\epsilon)^2}=\frac{2tr(W^2)}{n^2+2n}\sim\frac{1}{nh},$$

under Assumptions 1-3. Hence, by Markov's inequality,

$$\sqrt{hn}\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}=O_e(1).$$

Lemma 3.2 *Under Assumptions 1-3*

$$Z^{(2)}(0)=O_p(1),$$

where $Z^{(2)}(.)$ is defined according to (3.A.4).

Proof By rearranging terms in the first two lines of (3.A.4),

$$Z^{(2)}(0)=-\sqrt{\frac{h}{n}}\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}-nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right)+\sqrt{\frac{h}{n}}\left(2n\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2-2nE\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2\right).$$

By the C_r inequality,

$$\begin{aligned}&E(Z^{(2)}(0))^2\\ &\leq 2\frac{h}{n}E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}-nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right)^2+2\frac{h}{n}E\left(2n\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2-2nE\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2\right)^2.\end{aligned} \tag{3.A.25}$$

Now,

$$\begin{aligned}E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}-nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right)^2&=E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}-tr(W'W)\right)^2\\ =n^2\frac{E(\epsilon' W' W\epsilon)^2}{E(\epsilon'\epsilon)^2}&+(tr(W'W))^2-2ntr(W'W)\frac{E(\epsilon' W' W\epsilon)}{E(\epsilon'\epsilon)}\end{aligned}$$

$$
\begin{aligned}
&=((tr(W'W))^2 + 2tr((W'W)^2))\left(1+\frac{2}{n}\right)^{-1} + (tr(W'W))^2 - 2(tr(W'W))^2 \\
&=((tr(W'W))^2 + 2tr((W'W)^2))\left(1-\frac{2}{n}+O\left(\frac{1}{n^2}\right)\right) - (tr(W'W))^2 \\
&=2tr((W'W)^2)\left(1-\frac{2}{n}+O\left(\frac{1}{n^2}\right)\right) - (tr(W'W))^2\left(\frac{2}{n}+O\left(\frac{1}{n^2}\right)\right) \qquad (3.A.26)
\end{aligned}
$$

and hence,

$$
E\left(n\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon} - nE\left(\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon}\right)\right)^2 \sim 2tr((W'W)^2) \sim \frac{n}{h}, \qquad (3.A.27)
$$

under Assumption 3. In case the sequence h is bounded, the latter result would be modified as

$$
E\left(n\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon} - nE\left(\frac{\epsilon'W'W\epsilon}{\epsilon'\epsilon}\right)\right)^2 \sim 2tr((W'W)^2) - \frac{2}{n}(tr(W'W))^2 \sim n.
$$

It is worth stressing that, despite we are not attempting to provide an exact rate, we could not use the inequality

$$
E\left(X - E(X)\right)^2 \leq E(X^2)
$$

instead of (3.A.26), as it would neglect relevant terms. Moreover,

$$
\begin{aligned}
&4n^2 E\left(\left(\frac{\epsilon'W\epsilon}{\epsilon'\epsilon}\right)^2 - E\left(\frac{\epsilon'W\epsilon}{\epsilon'\epsilon}\right)^2\right)^2 \leq 4n^2 E\left(\frac{\epsilon'W\epsilon}{\epsilon'\epsilon}\right)^4 \\
=&4n^2\frac{E(\epsilon'W\epsilon)^4}{E(\epsilon'\epsilon)^4} \sim 4n^2\frac{12(tr(W^2))^2 + 48tr(W^4)}{n^4} \sim \frac{1}{h^2}. \qquad (3.A.28)
\end{aligned}
$$

Collecting (3.A.25), (3.A.27), (3.A.28) and by Markov's inequality, we conclude $Z^{(2)}(0) = O_p(1)$.

Lemma 3.3 *Under Assumptions 1-3*

$$
J(0) = O_p(1),
$$

where $J(0)$ is defined according to (3.A.8).

Proof By the C_r inequality (applied twice),

$$
\begin{aligned}
E(J(0))^2 \leq & 2\frac{h^2}{n^2}\left(E\left(\frac{6n\epsilon'W\epsilon\epsilon'W'W\epsilon}{(\epsilon'\epsilon)^2}\right)^2 + E\left(\frac{8n(\epsilon'W\epsilon)^3}{(\epsilon'\epsilon)^3} - 2tr(W^3)\right)^2\right) \\
\leq & 2\frac{h^2}{n^2}E\left(\frac{6n\epsilon'W\epsilon\epsilon'W'W\epsilon}{(\epsilon'\epsilon)^2}\right)^2 + 4\frac{h^2}{n^2}E\left(\frac{8n(\epsilon'W\epsilon)^3}{(\epsilon'\epsilon)^3}\right)^2
\end{aligned}
$$

$$+4\frac{h^2}{n^2}(2tr(W^3))^2. \tag{3.A.29}$$

In order to evaluate the rate of the first term in (3.A.29), we use $E(A/B) \sim E(A)/E(B)$, without deriving the exact order of the remainder. As previously mentioned, $E(A/B) = E(A)/E(B)$ when A/B is independent of B. When the latter fails, we are able to justify $E(A/B) \sim E(A)/E(B)$ as an approximation using an argument similar to Lieberman (1994). Using standard results on the expectations of quadratic forms,

$$E\left(\frac{6n\epsilon' W\epsilon\epsilon' W' W\epsilon}{(\epsilon'\epsilon)^2}\right)^2 \sim 36n^2\frac{E(\epsilon' W\epsilon\epsilon' W' W\epsilon)^2}{E(\epsilon'\epsilon)^4} \sim 36n^2\frac{2tr(W^2)(tr(W'W))^2}{n^4} \sim \frac{n}{h^3}. \tag{3.A.30}$$

Moreover, by a recursive formula (Ghazal (1996)),

$$E(\epsilon' W\epsilon)^n = \sum_{i=0}^{n-1} g_i E(\epsilon' W\epsilon)^{n-1-i}, \tag{3.A.31}$$

where

$$g_i = \begin{pmatrix} n-1 \\ i \end{pmatrix} i!2^i\sigma^{2i+2}tr((W)^{i+1}),$$

we have

$$E\left(\frac{8n(\epsilon' W\epsilon)^3}{(\epsilon'\epsilon)^3}\right)^2 = \frac{64n^2E(\epsilon' W\epsilon)^6}{E(\epsilon'\epsilon)^6} \sim 64n^2\frac{120(tr(W^2))^3}{n^6} \sim \frac{1}{nh^3}. \tag{3.A.32}$$

Hence, the term

$$4(tr(W^3)^2) \sim \frac{n^2}{h^2}$$

in (3.A.29) dominates both (3.A.30) and (3.A.32), whether h is divergent or bounded. Therefore,

$$E(J(0))^2 = O(\frac{h^2}{n^2}\frac{n^2}{h^2}) = O(1),$$

implying $J(0) = O_p(1)$.

Evaluation of cumulants

Here we evaluate the contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ of the term

$$Q_1 = it\sqrt{\frac{h}{n}}\frac{1}{K(0)}\frac{1}{n}E(Z^{(1)}(0)Z^{(2)}(0)\epsilon'\Sigma^{-1}\epsilon)$$

appearing in (3.A.15). Since the expansion in Theorem 3.1 is formal, $E(A/B) \sim E(A)/E(B)$ is used without proving the exact order of the remainder terms. Substituting (3.A.3) and (3.A.4),

$$Q_1 = Q_{11} + Q_{12} + Q_{13},$$

where

$$Q_{11} = -it\sqrt{\frac{h}{n}}\frac{1}{K(0)}hE\left(\frac{\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon\epsilon'\Sigma^{-1/2}W'W\Sigma^{-1/2}\epsilon}{\epsilon'\Sigma^{-1}\epsilon}\right),$$

$$Q_{12} = 2it\sqrt{\frac{h}{n}}\frac{1}{K(0)}hE\left(\frac{(\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon)^3}{(\epsilon'\Sigma^{-1}\epsilon)^2}\right)$$

and

$$Q_{13} = it\sqrt{\frac{h}{n}}\frac{1}{K(0)}\frac{h}{n}\left(tr(W'W) - \frac{1}{n}tr((W+W')^2)\left(1+\frac{2}{n}\right)^{-1}\right)E(\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon).$$

Contribution from term Q_{11}

By standard results on the expectations of quadratic forms in normal random variables, we have

$$\begin{aligned} Q_{11} &\sim -it\sqrt{\frac{h}{n}}\frac{1}{K(0)}h\frac{E\left(\frac{1}{2}\epsilon'\Sigma^{-1/2}(W+W')\Sigma^{-1/2}\epsilon\epsilon'\Sigma^{-1/2}W'W\Sigma^{-1/2}\epsilon\right)}{E(\epsilon'\Sigma^{-1}\epsilon)} \\ &= -it\sqrt{\frac{h}{n}}\frac{1}{K(0)}h\frac{\sigma^4\left(\frac{1}{2}tr((W+W')\Sigma^{-1})tr(W'W\Sigma^{-1}) + tr(\Sigma^{-1}(W+W')\Sigma^{-1}W'W)\right)}{\sigma^2 tr(\Sigma^{-1})}. \end{aligned}$$

Since

$$\Sigma^{-1} = (I - it\sigma^2(\tilde{C}+\tilde{C}'))^{-1} = \sum_{s=0}^{\infty}(it\sigma^2(\tilde{C}+\tilde{C}'))^s$$

by (2.4.2), it is straightforward to show that $tr(\Sigma^{-1}) \sim n$.

The contribution from Q_{11} to $\tilde{\kappa}_1$ is then

$$-2\sqrt{\frac{h}{n}}\frac{1}{K(0)}\frac{h}{n}\sigma^2 tr(WW'W) + o\left(\sqrt{\frac{h}{n}}\right) = -2\sigma^2\sqrt{\frac{h}{n}}\frac{tr(WW'W)}{tr(W^2)+tr(W'W)} + o\left(\sqrt{\frac{h}{n}}\right), \tag{3.A.33}$$

since

$$K(0) = \frac{h}{n}(trW^2 + W'W)) + O(\frac{1}{n}),$$

according to (3.A.9).

The contribution to $\tilde{\kappa}_2$ comes from the term

$$-(it)^2\sigma^4(\frac{h}{n})^{3/2}\frac{1}{K(0)}\{\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}'))tr(W'W)$$

$$+ \quad tr((\tilde{C}+\tilde{C}')(W+W')W'W)+tr((W+W')(\tilde{C}+\tilde{C}')W'W)\},$$

with $\tilde{C}$ given by (3.A.12). The contribution to $\tilde{\kappa}_2$ is given by

$$-\sigma^4(\frac{h}{n})^2\frac{1}{K(0)}(tr((W+W')^2)tr(W'W)+o_p\left(\sqrt{\frac{h}{n}}\right), \tag{3.A.34}$$

since

$$tr((W+W')(\tilde{C}+\tilde{C}'))=(tr(W+W')^2)\sqrt{\frac{h}{n}}\sim\left(\sqrt{\frac{n}{h}}\right), \tag{3.A.35}$$

$$tr((\tilde{C}+\tilde{C}')(W+W')W'W)\sim\left(\sqrt{\frac{n}{h}}\right) \tag{3.A.36}$$

and

$$tr((W+W')(\tilde{C}+\tilde{C}')W'W)\sim\left(\sqrt{\frac{n}{h}}\right). \tag{3.A.37}$$

Similarly, the contribution to $\tilde{\kappa}_3$ comes from the term

$$\begin{aligned}
&-(it)^3(\frac{h}{n})^{3/2}\frac{1}{K(0)}\frac{1}{\sigma^2}\sigma^8(\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}'))tr(W'W(\tilde{C}+\tilde{C}'))\\
+\quad &\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}')^2)tr(W'W)+tr((\tilde{C}+\tilde{C}')^2(W+W')W'W)\\
+\quad &tr((W+W')(\tilde{C}+\tilde{C}')^2W'W)+tr((\tilde{C}+\tilde{C}')(W+W')(\tilde{C}+\tilde{C}')W'W)).
\end{aligned}$$

Now,

$$tr(W'W(\tilde{C}+\tilde{C}'))\sim\sqrt{\frac{h}{n}}2tr(WW'W), \tag{3.A.38}$$

$$tr(W(\tilde{C}+\tilde{C}')^2)\sim\frac{h}{n}tr((W+W')^3), \tag{3.A.39}$$

$$tr((\tilde{C}+\tilde{C}')^2(W+W')W'W)=o\left(\sqrt{\frac{n}{h}}\right), \tag{3.A.40}$$

$$tr((W+W')(\tilde{C}+\tilde{C}')^2W'W)=o\left(\sqrt{\frac{n}{h}}\right) \tag{3.A.41}$$

and

$$tr((\tilde{C}+\tilde{C}')(W+W')(\tilde{C}+\tilde{C}')W'W)=o\left(\sqrt{\frac{n}{h}}\right). \tag{3.A.42}$$

Using (3.A.35), (3.A.36), (3.A.38)-(3.A.42), and after some tedious but straightforward algebra we conclude that the contribution to $\tilde{\kappa}_3$ is

$$\begin{aligned}
&-6\left(\frac{h}{n}\right)^{5/2}\frac{1}{K(0)}\sigma^6(2tr(W^2)tr(WW'W)+5tr(W'W)tr(WW'W)\\
&+tr(W'W)tr(W^3))+o\left(\sqrt{\frac{h}{n}}\right).
\end{aligned} \tag{3.A.43}$$

When W is symmetric (e.g. W given in (1.1.2)), the latter expression simplifies to

$$-24\left(\frac{h}{n}\right)^{3/2}\sigma^6 tr(W^3) + o\left(\sqrt{\frac{h}{n}}\right),$$

as

$$K(0) = 2\frac{h}{n}trW^2 + O\left(\frac{1}{n}\right),$$

according to (3.A.9).

Contribution from term Q_{12}

When Σ^{-1} is positive definite, $1/2(\epsilon'\Sigma^{-1/2}(W+W')\Sigma^{-1/2}\epsilon)/\epsilon'\Sigma^{-1}\epsilon$ and $\epsilon'\Sigma^{-1}\epsilon$ are independent (see e.g. Heijmans (1999), who provides sufficient conditions for Pitman (1937) general result to hold). Without considering validity issues for (2.4.2), Σ^{-1} in (3.A.13) is indeed positive definite, since $\Sigma^{-1} \sim \sigma^2 I$. Hence,

$$\begin{aligned}
Q_{12} &= 2it\sqrt{\frac{h}{n}}\frac{1}{K(0)}hE\left(\left(\frac{\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon}{\epsilon'\Sigma^{-1}\epsilon}\right)^3 \epsilon'\Sigma^{-1}\epsilon\right) \\
&= 2it\sqrt{\frac{h}{n}}\frac{1}{K(0)}hE\left(\frac{\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon}{\epsilon'\Sigma^{-1}\epsilon}\right)^3 E(\epsilon'\Sigma^{-1}\epsilon) \\
&= 2it\sqrt{\frac{h}{n}}\frac{1}{K(0)}h\frac{E\left(\frac{1}{2}\epsilon'\Sigma^{-1/2}(W+W')\Sigma^{-1/2}\epsilon\right)^3}{E(\epsilon'\Sigma^{-1}\epsilon)^3}E(\epsilon'\Sigma^{-1}\epsilon) \\
&= 2it\sqrt{\frac{h}{n}}\frac{1}{K(0)}h\sigma^6\{(\frac{1}{2}tr((W+W')\Sigma^{-1}))^3 \\
&+ 6tr\left(\frac{1}{2}(W+W')\Sigma^{-1}\right)tr\left(\left(\frac{1}{2}(W+W')\Sigma^{-1}\right)^2\right) \\
&+ 8tr\left(\left(\frac{1}{2}\Sigma^{-1}(W+W')\right)^3\right)\}\frac{\sigma^2 tr(\Sigma^{-1})}{\sigma^6\left((tr\Sigma^{-1})^3 + 6tr(\Sigma^{-1})tr(\Sigma^{-2}) + 8tr(\Sigma^{-3})\right)}.
\end{aligned}$$

We have

$$(tr\Sigma^{-1})^3 + 6tr(\Sigma^{-1})tr(\Sigma^{-2}) + 8tr(\Sigma^{-3}) \sim n^3$$

and $tr(\Sigma^{-1}) \sim n$.

The contribution from Q_{12} to $\tilde{\kappa}_1$ is then

$$2\sigma^2\sqrt{\frac{h}{n}}\frac{1}{K(0)}\frac{h}{n^2}tr((W+W')^3) = o\left(\sqrt{\frac{h}{n}}\right), \tag{3.A.44}$$

A similar argument holds also for the contribution from Q_{12} to both $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$.

Contribution from term Q_{13}

We have

$$Q_{13} = it\left(\frac{h}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W) - \frac{1}{n}tr((W+W')^2)(1+\frac{2}{n})^{-1})\sigma^2 tr(\frac{1}{2}(W+W')\Sigma^{-1}).$$

It is straightforward to see that there are no contributions to $\tilde{\kappa}_1$, since

$$\sigma^2\left(\frac{h}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W) - \frac{1}{n}tr(W^2)(1+\frac{2}{n})^{-1})tr(W) = 0. \tag{3.A.45}$$

The contribution to $\tilde{\kappa}_2$ comes from

$$(it)^2\sigma^4\left(\frac{h}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W) - \frac{4}{n}tr(W^2)(1+\frac{2}{n})^{-1})\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}'))$$

and by (3.A.35) we conclude that Q_{13} contributes to κ_2 with

$$\sigma^4\left(\frac{h}{n}\right)^2\frac{1}{K(0)}tr((W+W')^2)tr(W'W) + o_p\left(\sqrt{\frac{h}{n}}\right). \tag{3.A.46}$$

The contribution to $\tilde{\kappa}_3$ comes from

$$(it)^3\sigma^6\left(\frac{h}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W) - \frac{1}{n}tr(W^2)(1+\frac{2}{n})^{-1})\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}')^2)$$

and hence, from (3.A.39), we conclude that Q_{13} contributes to κ_3 with

$$6\sigma^6\left(\frac{h}{n}\right)^{5/2}\frac{1}{K(0)}tr(W'W)(tr(W^3)+3tr(W(W')^2)) + o\left(\sqrt{\frac{h}{n}}\right). \tag{3.A.47}$$

When W is symmetric, the latter simplifies to

$$12\sigma^6(\frac{h}{n})^{3/2}tr(W^3) + o\left(\sqrt{\frac{h}{n}}\right).$$

From (3.A.33), (3.A.44) and (3.A.45) we conclude that Q_1 contributes to $\tilde{\kappa}_1$ with the term

$$-2\sigma^2\left(\frac{h}{n}\right)^{3/2}\frac{1}{K(0)}tr(WW'W) + o\left(\sqrt{\frac{h}{n}}\right). \tag{3.A.48}$$

From (3.A.34) and (3.A.46) we conclude that any contribution to $\tilde{\kappa}_2$ from Q_1 is neglegible, while collecting (3.A.43) and (3.A.47) we have that the contribution to $\tilde{\kappa}_3$

from Q_1 is

$$\begin{aligned}
&-12\sigma^6\left(\frac{h}{n}\right)^{5/2}\frac{1}{K(0)}tr(WW'W)(tr(W^2)+tr(W'W))+o\left(\sqrt{\frac{h}{n}}\right)\\
=&-12\sigma^6\left(\frac{h}{n}\right)^{3/2}tr(WW'W)+o\left(\sqrt{\frac{h}{n}}\right). \qquad (3.A.49)
\end{aligned}$$

Finally, we report the main steps for the evaluation of the contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ from

$$Q_2=\frac{1}{2}it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}\frac{1}{n}E((Z^{(1)}(0))^2J(0)\epsilon'\Sigma^{-1}\epsilon).$$

Substituting (3.A.3) and (3.A.8), we write

$$Q_2=Q_{21}+Q_{22}+Q_{23},$$

where, by independence between $1/2(\epsilon'\Sigma^{-1/2}(W+W')\Sigma^{-1/2}\epsilon)/\epsilon'\Sigma^{-1}\epsilon$ and $\epsilon'\Sigma^{-1}\epsilon$,

$$Q_{21}\sim -3it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}h^2\frac{E((\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon)^3\epsilon'\Sigma^{-1/2}W'W\Sigma^{-1/2}\epsilon)}{E(\epsilon'\Sigma^{-1}\epsilon)^3},$$

$$\begin{aligned}
Q_{22} &= 4it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}h^2E\left(\left(\frac{\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon}{\epsilon'\Sigma^{-1}\epsilon}\right)^5\epsilon'\Sigma^{-1}\epsilon\right)\\
&= 4it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}h^2\frac{E(\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon)^5}{E(\epsilon'\Sigma^{-1}\epsilon)^5}E(\epsilon'\Sigma^{-1}\epsilon)
\end{aligned}$$

and

$$\begin{aligned}
Q_{23} &= -it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}\frac{h^2}{n}tr(W^3)E\left(\left(\frac{\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon}{\epsilon'\Sigma^{-1}\epsilon}\right)^2\epsilon'\Sigma^{-1}\epsilon\right)\\
&= -it\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}\frac{h^2}{n}tr(W^3)\frac{E(\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon)^2}{E(\epsilon'\Sigma^{-1}\epsilon)^2}E(\epsilon'\Sigma^{-1}\epsilon).
\end{aligned}$$

Contribution from term Q_{21}

Using some standard results on the expectations of quadratic forms in normal random variables,

$$\begin{aligned}
&E((\frac{1}{2}\epsilon'\Sigma^{-1/2}(W+W')\Sigma^{-1/2}\epsilon)^3\epsilon'\Sigma^{-1/2}W'W\Sigma^{-1/2}\epsilon)\\
&=\sigma^8((tr(\Sigma^{-1}\frac{1}{2}(W+W')))^3tr(\Sigma^{-1}W'W)
\end{aligned}$$

$$
\begin{aligned}
&+6(tr(\Sigma^{-1}\frac{1}{2}(W+W')))^2 tr(\Sigma^{-1}\frac{1}{2}(W+W')\Sigma^{-1}W'W) \\
&+6tr((\Sigma^{-1}\frac{1}{2}(W+W'))^2)tr(\Sigma^{-1}\frac{1}{2}(W+W'))tr(\Sigma^{-1}W'W) \\
&+8tr(\Sigma^{-1}\frac{1}{2}(W+W'))tr((\Sigma^{-1}\frac{1}{2}(W+W'))^2\Sigma^{-1}W'W) \\
&+12tr(\Sigma^{-1}\frac{1}{2}(W+W')\Sigma^{-1}W'W)tr((\Sigma^{-1}\frac{1}{2}(W+W'))^2) \\
&+48tr((\Sigma^{-1}\frac{1}{2}(W+W'))^3\Sigma^{-1}W'W))
\end{aligned}
$$

and $E(\epsilon'\Sigma^{-1}\epsilon)^3 \sim \sigma^6 n^3$. Therefore, the contribution to $\tilde{\kappa}_1$ is

$$
\begin{aligned}
& -3\sigma^2\sqrt{\frac{h}{n}}\frac{1}{(K(0))^2}\frac{h^2}{n^3}(\frac{3}{2}tr((W+W')^2)tr((W+W')WW') \\
+ \quad & 6tr((W+W')^3W'W)) = o\left(\sqrt{\frac{h}{n}}\right).
\end{aligned}
$$

A similar argument holds for the contribution to both $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$.

Contribution from term Q_{22}

We have

$$
\frac{E(\epsilon'\Sigma^{-1}\epsilon)}{E(\epsilon'\Sigma^{-1}\epsilon)^5} \sim \frac{\sigma^2 n}{\sigma^{10}n^5} = \frac{1}{\sigma^8 n^4}.
$$

Also, we can evaluate the fifth moment of $\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon$ by the recursive formula given in (3.A.31). By tedious, but straightforward algebra, it is possible to show that the contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ are $o(\sqrt{h/n})$. Intuitively, this is because no term in $E((\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon)^5)$ is large enough to offset the factor h^2/n^4.

Contribution from term Q_{23}

We have

$$
\begin{aligned}
E((\epsilon'\Sigma^{-1/2}W\Sigma^{-1/2}\epsilon)^2) &= \frac{1}{4}E(\epsilon'\Sigma^{-1/2}(W+W')\Sigma^{-1/2}\epsilon)^2 \\
&= \sigma^4\left(\frac{1}{4}(tr(\Sigma^{-1}(W+W')))^2 + \frac{1}{2}tr((\Sigma^{-1}(W+W'))^2)\right)
\end{aligned}
$$

and

$$
\frac{E(\epsilon'\Sigma^{-1}\epsilon)}{E(\epsilon'\Sigma^{-1}\epsilon)^2} \sim \frac{n\sigma^2}{n^2\sigma^4} = \frac{1}{n\sigma^2}.
$$

Therefore, the contribution to $\tilde{\kappa}_1$ is

$$
-\sigma^2\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}\frac{h^2}{n^2}(tr(W^2)+tr(W'W))tr(W^3)
$$

$$= -\sigma^2\sqrt{\frac{h}{n}}\frac{tr(W^3)}{(tr(W^2+W'W))} + o\left(\sqrt{\frac{h}{n}}\right). \tag{3.A.50}$$

Similarly, the contribution to $\tilde{\kappa}_2$ comes from the term

$$-(it)^2\sigma^4\sqrt{\frac{h}{n}}\frac{1}{K(0)^2}\frac{h^2}{n^2}tr(W^3)tr((W+W')^2(\tilde{C}+\tilde{C}'))$$

and, by (3.A.36), is $o(\sqrt{h/n})$.

Finally, the contribution to $\tilde{\kappa}_3$ comes from the term

$$-\frac{1}{4}(it)^3\sigma^6\left(\frac{h}{n}\right)^{7/2}\frac{1}{K(0)^2}tr(W^3)(tr((W+W')^2))^2$$

and hence the actual contribution to $\tilde{\kappa}_3$ is

$$\begin{aligned}&-6\sigma^6\left(\frac{h}{n}\right)^{7/2}\frac{1}{K(0)^2}tr(W^3)(tr(W^2)+tr(W'W))^2 + o\left(\sqrt{\frac{h}{n}}\right)\\ &= -6\sigma^6\left(\frac{h}{n}\right)^{3/2}tr(W^3) + o\left(\sqrt{\frac{h}{n}}\right).\end{aligned} \tag{3.A.51}$$

Collecting (3.A.48) and (3.A.50), we conclude that the contribution to $\tilde{\kappa}_1$ from Q_1+Q_2 is

$$-2\sigma^2\left(\frac{h}{n}\right)^{3/2}\frac{tr(WW'W)}{K(0)} - \sigma^2\left(\frac{h}{n}\right)^{3/2}\frac{tr(W^3)}{K(0)} + o(\sqrt{\frac{h}{n}}).$$

The overall contribution to $\tilde{\kappa}_2$ from Q_1+Q_2 is neglegible, while that to $\tilde{\kappa}_3$ is

$$-12\sigma^6\left(\frac{h}{n}\right)^{3/2}tr(WW'W) - 6\sigma^6\left(\frac{h}{n}\right)^{3/2}tr(W^3) + o\left(\sqrt{\frac{h}{n}}\right),$$

by collecting (3.A.49) and (3.A.51).

4 Finite Sample Corrections for the LM Test in SAR Models

As already outlined in Chapter 1, LM testing is especially computationally convenient because it depends on the null model, and thus does not require estimating the spatial coefficient. An LM test can be expected to be efficient against local SAR alternatives, and to have an asymptotic χ^2 distribution under the null. However, the χ^2 approximation may not be accurate in modest samples, so a test based on it may be badly sized. Thus we develop tests with improved finite-sample properties.

The main contribution of this chapter is to develop tests based on the Edgeworth expansion of the cdf of the LM statistic. Specifically, in Section 4.1 we derive a refined test for H_0 in (1.2.1) against a (2.2.1) when the data follow model (1.2.5). We focus here on a two-sided test because in some circumstances the practitioner might not have an ex ante evidence regarding the sign of λ. We then provide corresponding tests of (1.2.1) in linear regression models with SAR disturbances. In both cases the proofs of the theorems are left to the Appendix. In Section 4.3 we describe the finite sample corrections of Robinson (2008b), so that the finite sample performance of the latter can be compared with that of the Edgeworth-corrected tests. In Section 4.4 we compare the corrected tests presented in Sections 4.1-4.3 with bootstrap-based ones in a Monte Carlo study of finite sample. Section 4.5 compares the Edgeworth approximation with the the exact distribution of the LM statistic.

4.1 Edgeworth-corrected LM tests for independence in pure SAR

We suppose that model (1.2.5) holds and we focus on testing (1.2.1) against (2.2.1). For any admissible values of λ and σ^2, the Gaussian log-likelihood for Y in model (1.2.5) is given by (3.1.1). As discussed in Section 3.1, any $\lambda \in \Lambda$ where Λ is any closed subset of $(-1, 1)$ is admissible.

By standard linear algebra,

$$\frac{\partial l(\lambda,\sigma^2)}{\partial \lambda} = -tr(S^{-1}(\lambda)W) + \frac{1}{\sigma^2}Y'S(\lambda)W'Y,$$

$$\frac{\partial^2 l(\lambda,\sigma^2)}{\partial \lambda^2} = -tr((S^{-1}(\lambda)W)^2) - \frac{1}{\sigma^2}Y'W'WY.$$

Hence, given the MLE for σ^2 displayed in (3.1.2)

$$\frac{\partial l(\lambda,\sigma^2)}{\partial \lambda}|_{H_0} = \frac{Y'WY}{\frac{1}{n}Y'Y},$$

and

$$-E\left(\frac{\partial^2 l(\lambda,\sigma^2)}{\partial\lambda^2}\right)|_{H_0} = tr(W'W + W^2).$$

Therefore, a version of the LM statistic is

$$LM = \frac{n^2}{tr(W^2 + W'W)}\left(\frac{Y'WY}{Y'Y}\right)^2 \tag{4.1.1}$$

so, under H_0,

$$LM = \frac{n^2}{tr(W^2 + W'W)}\left(\frac{\epsilon'W\epsilon}{\epsilon'\epsilon}\right)^2. \tag{4.1.2}$$

The latter corresponds to (1.2.16) when $\beta = 0$ a priori in (1.1.3).

As noted by Burridge (1980), (4.1.2) is also the LM statistic for testing (1.2.1) against the spatial moving average model

$$Y = \epsilon + \lambda W\epsilon$$

(a corresponding equivalence to that found with time series models).

As already discussed in Chapter 1, under suitable conditions we have

$$P(LM \leq \eta) = F(\eta) + o(1)$$

for any $\eta > 0$, where F denotes the cdf of a χ^2_1 random variable. Thus (1.2.1) is rejected in favour of (2.2.1) if LM exceeds the appropriate percentile of the χ^2_1 distribution. We can likewise test (1.2.1) against (2.1.1) by comparing $\sqrt{LM}$ with the appropriate upper or lower percentiles of the standard normal distribution. However, except in Section 4.5, we focus throughout on a two-sided tests.

Throughout this chapter f denotes the χ^2 pdf.

Theorem 4.1 *Suppose that model (1.2.5) and Assumptions 1-3 hold. Under H_0 in (1.2.1), the cdf of LM admits the formal Edgeworth expansion*

$$Pr(LM \leq \eta|H_0) = F(\eta) + \frac{\kappa}{4}\eta f(\eta) - \frac{\kappa}{12}\eta^2 f(\eta) + o\left(\frac{h}{n}\right) \tag{4.1.3}$$

in case h is divergent, and

$$Pr(LM \leq \eta) = F(\eta) + \frac{\kappa}{4}\eta f(\eta) - \frac{\kappa}{12}\eta^2 f(\eta) - \frac{2}{n}\eta^2 f(\eta) + o\left(\frac{1}{n}\right) \tag{4.1.4}$$

when h is bounded, where

$$\kappa = \frac{3tr(W' + W)^4}{\tilde{a}^4} \sim \frac{h}{n} \tag{4.1.5}$$

and $\tilde{a}$ defined according to (3.1.4)

The proof of Theorem 4.1 is in the Appendix. Again, it must be stressed that both expansions in Theorem 4.1 are formal.

Clearly, (4.1.3) and (4.1.4) entail better approximations than (1.2.18). The leading terms in (4.1.3) and (4.1.4) depend on known quantities, so they can be used directly for approximating the cdf. The two outcomes in Theorem 4.1 create a dilemma for the practitioner because it cannot be determined from a finite data set whether to treat h as divergent or bounded. However, (4.1.4) is justified also when h is divergent because the extra term in the expansion, $-2\eta^2 f(\eta)/n$, is $o(h/n)$.

Theorem 4.1 can be used to derive Edgeworth-corrected critical values. Let w_α^{LM} be the α-quantile of LM. By inverting either expansion, we can expand w_α^{LM} as an infinite series

$$w_\alpha^{LM} = z^2_{(1+\alpha)/2} + p_1(z^2_{(1+\alpha)/2}) + ..., \qquad (4.1.6)$$

where $p_1(z^2_{(1+\alpha)/2})$ is a polynomial whose coefficients have order h/n, and that can be determined using the identity $\alpha = Pr(LM \leq w_\alpha^{LM})$ and the expansions given in Theorem 4.1. It is worth recalling that, consistently with the notation used in Chapters 2 and 3, z_α denoted the $\alpha-$quantile of the standard normal variate. Specifically, when h is divergent, we have

$$\alpha = Pr(LM \leq w_\alpha^{LM}) = F(w_\alpha^{LM}) + \left(\frac{\kappa}{4}w_\alpha^{LM} - \frac{\kappa}{12}(w_\alpha^{LM})^2\right) f((w_\alpha^{LM})^2) + o\left(\frac{h}{n}\right).$$

By substituting (4.1.6), the leading terms of the LHS are

$$\begin{aligned} & F(z^2_{(1+\alpha)/2}) + p_1(z^2_{(1+\alpha)/2}) f(z^2_{(1+\alpha)/2}) \\ + & \left(\frac{\kappa}{4}z^2_{(1+\alpha)/2} - \frac{\kappa}{12}z^4_{(1+\alpha)/2}\right) f(z^2_{(1+\alpha)/2}) + o\left(\frac{h}{n}\right) \\ = & \ \alpha + p_1(z^2_{(1+\alpha)/2}) f(z^2_{(1+\alpha)/2}) \\ + & \left(\frac{\kappa}{4}z^2_{(1+\alpha)/2} - \frac{\kappa}{12}z^4_{(1+\alpha)/2}\right) f(z^2_{(1+\alpha)/2}) + o\left(\frac{h}{n}\right). \end{aligned}$$

The latter is $\alpha + o(h/n)$ (rather than $\alpha + O(h/n)$), when we take

$$p_1(x) = -\left(\frac{\kappa}{4}x - \frac{\kappa}{12}x^2\right) \sim \frac{h}{n}. \qquad (4.1.7)$$

Similarly, when h is bounded, we take

$$p_1(x) = -\left(\frac{\kappa}{4}x - \frac{\kappa}{12}x^2 - \frac{2}{n}x^2\right) \sim \frac{1}{n}. \qquad (4.1.8)$$

If w_α^{LM} were known, the size of a test of H_0 in (1.2.1) would obviously be $Pr(LM > w_\alpha^{LM}|H_0) = 1 - \alpha$. We can compare the size of the test of (1.2.1) against (2.2.1) based on the usual first order approximation, i.e.

$$Pr(LM > z^2_{(\alpha+1)/2}|H_0) \tag{4.1.9}$$

with

$$Pr(LM > z^2_{(\alpha+1)/2} + p_1(z^2_{(\alpha+1)/2})|H_0), \tag{4.1.10}$$

where $p_1(.)$ is defined according to (4.1.7) if h is divergent and (4.1.8) if h is bounded.

Thus, the error of the approximation of (4.1.9) is $O(h/n)$, while that of (4.1.10) is $o(h/n)$ when the sequence h is divergent, or $o(1/n)$ when it is bounded.

As an alternative to using corrected critical values, we can also apply Theorem 4.1 to construct a transformation of LM whose distribution better approximates χ^2 than LM itself. Starting from the expansion in (4.1.3), we consider the cubic transformation

$$g(x) = x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 + \frac{1}{4}Q(x), \quad Q(x) = \left(\frac{\kappa}{4}\right)^2 \left(\frac{4}{27}x^3 - \frac{2}{3}x^2 + x\right), \tag{4.1.11}$$

such that

$$Pr(g(LM) \leq \eta) = F(\eta) + o\left(\frac{h}{n}\right).$$

Similarly, from (4.1.4), we can write

$$\begin{aligned} g(x) &= x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 - \frac{2}{n}x^2 + \frac{1}{4}Q(x), \\ Q(x) &= \left(\frac{\kappa}{4}\right)^2 x + \frac{1}{3}\left(\frac{\kappa}{6} + \frac{4}{n}\right)^2 x^3 - \frac{\kappa}{4}\left(\frac{\kappa}{6} + \frac{4}{n}\right)x^2, \end{aligned} \tag{4.1.12}$$

such that

$$Pr(g(LM) \leq \eta) = F(\eta) + o\left(\frac{1}{n}\right).$$

As already outlined in Section 2.1, the transformations (4.1.11) and (4.1.12) were proposed in case of a standard normal limiting distribution by Hall (1992), or, in a slightly more general setting, Yanagihara et al. (2005). In Lemma 4.2 (reported in the Appendix) we show that such result extends to χ^2 limiting distributions.

Therefore, we can compare

$$Pr(g(LM) > z^2_{(\alpha+1)/2}|H_0), \tag{4.1.13}$$

where $g(.)$ is defined according to (4.1.11) or (4.1.12) depending on h, with (4.1.9). Again, (4.1.13) has error $o(h/n)$ compared to the $O(h/n)$ error of (4.1.9).

4.2 Improved LM tests in regressions where the disturbances are spatially correlated

In this section we extend the results derived in Section 4.1 to to the more general model (1.2.15).

From Burridge (1980), Anselin (1988, 2001), the LM statistic for testing (1.2.1) against (2.2.1) is

$$\tilde{LM} = \frac{n^2}{tr(W'W)+tr(W^2)}\left(\frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}}\right)^2 = \frac{n^2}{\tilde{a}^2}\left(\frac{Y'PWPY}{Y'PY}\right)^2, \qquad (4.2.1)$$

where P is defined according to (1.2.17). Indeed, when data are driven by (1.2.15) and for any admissible values of λ, σ and β, the Gaussian log-likelihood for Y is given by

$$l(\lambda,\sigma^2,\beta) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) + ln(det(S(\lambda))) - \frac{1}{2\sigma^2}(Y-X\beta)'S(\lambda)S(\lambda)'(Y-X\beta).$$

Thus, given λ, the MLE for β and σ^2 are

$$\hat{\beta}(\lambda) = (X'S(\lambda)S(\lambda)'X)^{-1}X'S(\lambda)S(\lambda)'Y$$

and

$$\hat{\sigma}^2(\lambda) = \frac{1}{n}(Y-X\beta)'S(\lambda)S(\lambda)'(Y-X\beta).$$

It is straightforward to notice that $\hat{\beta}(0)$ is the OLS estimate of β. We denote $\hat{u} = Y - X\hat{\beta}(0)$, which is the vector of OLS residuals. By standard linear algebra,

$$\frac{\partial l(\lambda,\sigma^2,\beta)}{\partial\lambda}|_{H_0} = \frac{1}{\hat{\sigma}^2(0)}(Y-X\hat{\beta}(0))'W(Y-X\hat{\beta}(0)) = n\frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}}$$

and

$$-E\left(\frac{\partial^2 l(\lambda,\sigma^2)}{\partial\lambda^2}\right)|_{H_0} = tr(W'W)+tr(W^2).$$

Hence, $\tilde{LM}$ is given by (4.2.1).

We impose the following condition on X

Assumption 4 *For all n, each element x_{ij} of X is predetermined and $|x_{i,j}| \leq K$. Moreover, for all sufficiently large n*

$$0 < c_1 < \underline{\eta}\left(\frac{X'X}{n}\right)$$

and the limits of at least one element of each $X'WX/n$, $X'W^2X/n$ and $X'W'WX/n$ are non zero.

Non nullity of the limits of at least one element of each $X'WX/n$, $X'W^2X/n$ and $X'W'WX/n$ is required to ensure that the orders of some of the quantities appearing in the following Theorem hold as exact rates and not only as upper bounds, as will be explained below. We have the following results

Theorem 4.2 *Suppose that model (1.2.15) and Assumptions 1-4 hold. Under (1.2.1),* *the cdf of $\tilde{LM}$ admits the formal Edgeworth expansion*

$$Pr(\tilde{LM} \leq \eta | H_0) = F(\eta) + \left(\frac{\kappa}{4}\eta - \frac{\kappa}{12}\eta^2 + 2\omega_1\eta\right) f(\eta) + o\left(\frac{h}{n}\right) \tag{4.2.2}$$

with

$$\omega_1 = \frac{tr(K_3 - K_2)}{\tilde{a}^2} - \frac{1}{2}\frac{(tr(K_1))^2}{\tilde{a}^2} \sim \frac{h}{n} \tag{4.2.3}$$

if h is divergent, and

$$Pr(\tilde{LM} \leq \eta) = F(\eta) + \left(\frac{\kappa}{4}\eta - \frac{\kappa}{12}\eta^2 + 2\omega_2\eta - \frac{2}{n}\eta^2\right) f(\eta) + o\left(\frac{1}{n}\right) \tag{4.2.4}$$

with

$$\omega_2 = \frac{tr(K_3 - K_2)}{\tilde{a}^2} - \frac{1}{2}\frac{(tr(K_1))^2}{\tilde{a}^2} - \frac{k}{n} \sim \frac{1}{n} \tag{4.2.5}$$

if h is bounded, where κ is given in (4.1.5),

$$K_1 = (X'X)^{-1}X'WX, \tag{4.2.6}$$

$$K_2 = \frac{1}{2}X'(W + W')X(X'X)^{-1}X'(W' + W)X(X'X)^{-1} \tag{4.2.7}$$

and

$$K_3 = X'(W + W')^2X(X'X)^{-1}. \tag{4.2.8}$$

The components of $(X'X)^{-1}$ have order $1/n$ by Assumption 4. On the other hand, the absolute values of the components of $X'WX$, $X'(W + W')X$ and $X'(W + W')^2X$ are $O(n)$ by Lemma 2.2. Assumption 4 imposes that for at least one component of each matrix the latter holds as an exact rate. It follows that $tr(K_1)$, $tr(K_2)$ and $tr(K_3)$ are bounded and non zero. Since $\tilde{a}^2 \sim n/h$ under Assumption 3, ω_1 and ω_2 have exactly order h/n and $1/n$, respectively.

The proof of Theorem 4.2 is the Appendix. Again, both the expansions are formal.

From (4.2.2) and (4.2.4), we can obtain Edgeworth-corrected critical values. Proceeding as described in Section 4.1, the size based on χ^2 critical value is

$$Pr(\tilde{LM} > z^2_{(\alpha+1)/2} | H_0) \tag{4.2.9}$$

while the Edgeworth-corrected critical value is

$$Pr(\tilde{LM} > z^2_{(\alpha+1)/2} + \tilde{p}_1(z^2_{(\alpha+1)/2}) | H_0), \tag{4.2.10}$$

where

$$\tilde{p}_1(z^2_{(\alpha+1)/2}) = -\left(\frac{\kappa}{4}z^2_{(\alpha+1)/2} - \frac{\kappa}{12}z^4_{(\alpha+1)/2} + 2\omega_1 z^2_{(\alpha+1)/2}\right)$$

if h is divergent and

$$\tilde{p}_1(z_{(\alpha+1)/2}) = -\left(\frac{\kappa}{4}z^2_{(\alpha+1)/2} - \frac{\kappa}{12}z^4_{(\alpha+1)/2} + 2\omega_2 z^2_{(\alpha+1)/2} - \frac{2}{n}z^4_{(\alpha+1)/2}\right)$$

if h is bounded. As before, (4.2.9) has error of order h/n, while (4.2.10) has error $o(h/n)$.

As in Section 4.1, we can also consider Edgeworth-corrected test statistics. The size of test of (1.2.1) based on $\tilde{LM}$ is compared with that based on a corrected statistic, i.e.

$$Pr(g(\tilde{LM}) > z^2_{(\alpha+1)/2}|H_0). \tag{4.2.11}$$

The choice of the function g is motivated by Lemma 4.2 and in this case is given by

$$g(x) = x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 + 2\omega_1 x + \frac{1}{4}Q(x),$$

where

$$Q(x) = \left(\left(\frac{\kappa}{4}\right)^2 + 4\omega_1^2 + \kappa\omega_1\right)x - \frac{1}{2}\left(\frac{2}{3}\kappa\omega_1 + \frac{\kappa^2}{12}\right)x^2 + \frac{1}{3}\left(\frac{\kappa}{6}\right)^2 x^3$$

in case h is divergent and

$$g(x) = x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 + 2\omega_2 x - \frac{2}{n}x^2 + \frac{1}{4}Q(x),$$

with

$$Q(x) = \left(\frac{\kappa}{4} + 2\omega_2\right)^2 x - \left(\frac{\kappa}{4} + 2\omega_2\right)\left(\frac{\kappa}{6} + \frac{4}{n}\right)x^2 + \frac{1}{3}\left(\frac{\kappa}{6} + \frac{4}{n}\right)^2 x^3$$

if h is bounded. Similarly to Section 4.1, when $\tilde{LM}$ is used the error of the approximation has order h/n while it is reduced to $o(h/n)$ when the test is based on the Edgeworth-corrected variant.

4.3 Alternative correction

The results derived in Sections 4.1 and 4.2 can be compared with two alternative corrections derived for asymptotically χ^2 statistics in Robinson (2008b). The class of statistics considered in Robinson (2008b) include the LM for testing (1.2.1) in either (1.2.5) or (1.2.15) as special cases. In particular, Robinson (2008b) proposes both mean-adjusted and mean and variance-adjusted variants of (4.1.1) and (4.2.1), which prove to be asymptotically distributed as a χ^2 random variable with one degree of freedom. Such corrected statistics are expected to have better finite sample properties than either (4.1.1) or (4.2.1), even though the magnitude of the gain in accuracy is not explicitly shown. In finite sample the corrected statistic based on mean adjustment might have a larger variance than the non-corrected version, resulting in a partial (or total) offset of the gain in accuracy from the mean standardisation. In such case, a mean and variance standardisation should be performed instead.

It should be stressed that such corrected statistics might be convenient in the present case since the ratios $\epsilon' W \epsilon / \epsilon' \epsilon$ and $\epsilon' P W P \epsilon / \epsilon' P \epsilon$ are independent of their own denominator and therefore the expectation of the ratio is equal to the ratio of expectations (Pitman (1937)). If the latter condition failed, a correction based on mean and variance standardisation be much less feasible, since the evaluation of mean and variance would require some approximation.

We suppose that Assumptions 1-4 hold and focus on the simpler case first, i.e. the statistic given in (4.1.1). Specifically, Robinson (2008b) proposes a mean and variance-adjusted statistic under H_0 as

$$\left(\frac{2}{Var(LM)} \right)^{1/2} (LM - E(LM)) + 1, \tag{4.3.1}$$

where $Var(LM)$ denotes the variance of LM. In order to compare the performance of such corrected statistics with that based on the results presented in Section 4.1, the leading terms of (4.3.1) have to be derived.

As presented in Robinson (2008b),

$$E(LM) = \left(1 + \frac{2}{n} \right)^{-1}, \tag{4.3.2}$$

while

$$Var(LM) = \frac{n^4}{\tilde{a}^4} \frac{E(\epsilon' W \epsilon)^4}{E(\epsilon' \epsilon)^4} - \left(1 + \frac{2}{n} \right)^{-2}.$$

By standard formulae for moments of quadratic forms in normal random variables (see e.g. Ghazal (1996)),

$$\begin{aligned} E(\epsilon' W \epsilon)^4 &= E\left(\frac{1}{2} \epsilon' (W + W') \epsilon \right)^4 \\ &= \frac{1}{16} \sigma^8 (6tr((W + W')^2) E(\epsilon'(W + W')\epsilon)^2 + 48tr((W + W')^4)) \\ &= 3\sigma^8 (\tilde{a}^4 + tr((W + W')^4) \end{aligned}$$

and

$$E(\epsilon' \epsilon)^4 = \sigma^8 (n^4 + 12n^3 + 44n^2 + 48n) = \sigma^8 n^4 \left(1 + \frac{12}{n} + \frac{44}{n^2} + \frac{48}{n^3} \right).$$

Hence,

$$Var(LM) = \frac{n^4}{\tilde{a}^4} \frac{3\tilde{a}^4 + 3tr((W + W')^4)}{n^4 \left(1 + \frac{12}{n} + \frac{44}{n^2} + \frac{48}{n^3}\right)} - \left(1 + \frac{2}{n} \right)^{-2}$$

$$= \quad 2 + \frac{3tr((W+W')^4)}{\tilde{a}^4} - \frac{32}{n} + o\left(\frac{1}{n}\right), \qquad (4.3.3)$$

where the second equality follows by standard Taylor expansion.

Collecting (4.3.2) and (4.3.3), (4.3.1) becomes

$$\begin{aligned} & \left(1 + \frac{3tr((W+W')^4)}{2\tilde{a}^4} - \frac{16}{n} + o\left(\frac{1}{n}\right)\right)^{-1/2} \left(LM - \left(1+\frac{2}{n}\right)^{-1}\right) + 1 \\ = \quad & \left(1 - \frac{3}{4}\frac{tr((W+W')^4)}{\tilde{a}^4} + \frac{8}{n} + o\left(\frac{1}{n}\right)\right)\left(LM - 1 + \frac{2}{n} + o\left(\frac{1}{n}\right)\right) + 1, \end{aligned}$$

where the second equality follows by Taylor expansion. Hence, when h is divergent, we define

$$\bar{LM} = LM - \frac{3}{4}\frac{tr((W+W')^4)}{\tilde{a}^4}(LM-1), \qquad (4.3.4)$$

while

$$\bar{LM} = LM - \frac{3}{4}\frac{tr((W+W')^4)}{\tilde{a}^4}(LM-1) + \frac{8}{n}LM - \frac{6}{n} \qquad (4.3.5)$$

when h is bounded.

For both divergent and bounded h, we consider the size of the test of (1.2.1) against (2.2.1) based on $\bar{LM}$, i.e.

$$Pr(\bar{LM} > z^2_{(\alpha+1)/2}|H_0). \qquad (4.3.6)$$

We expect that when inference is based on $\bar{LM}$ rather than on LM, the error of the approximation is reduced by one order. To this extent, the finite sample performance of $\bar{LM}$ should be similar to that of $g(LM)$, with g defined in (4.1.11) or (4.1.12).

Finally, we consider the mean-adjusted null statistic corresponding to (4.2.1). Since the algebraic burden is larger relative to the previous case, the derivation of the mean and variance-adjusted variant is omitted. At the beginning of this section, we stressed that mean and mean and variance adjustments might be algebraically more convenient than Edgeworth corrections. However, the mean and variance standardisation of (4.2.1) does not entail significant computational advantage and is therefore omitted.

Given (4.2.1), Robinson (2008b) proposes the mean-adjusted null statistic

$$\frac{\tilde{LM}}{E(\tilde{LM})}. \qquad (4.3.7)$$

Using standard formulae, we describe the results of Robinson (2008b) as

$$\begin{aligned} E(\tilde{LM}) \quad &= \quad \frac{n^2}{\tilde{a}^2}\frac{E\left(\frac{1}{2}\epsilon' P(W+W')P\epsilon\right)^2}{E(\epsilon' P\epsilon)^2} \\ &= \quad 1 + \frac{(tr(K_1))^2}{\tilde{a}^2} + \frac{tr(K_2-K_3)}{\tilde{a}^2} - \frac{2(1-k)}{n} + O\left(\frac{1}{n^2}\right), \end{aligned}$$

where K_1, K_2 and K_3 are defined according to (4.2.6), (4.2.7) and (4.2.8), respectively. The second equality follows by a standard Taylor expansion of the denominator. Hence, (4.3.7) becomes

$$L\tilde{M}\left(1-\frac{(tr(K_1))^2}{\tilde{a}^2}-\frac{tr(K_2-K_3)}{\tilde{a}^2}\right)+o\left(\frac{h}{n}\right)$$

in case h is divergent, and

$$L\tilde{M}\left(1-\frac{(tr(K_1))^2}{\tilde{a}^2}-\frac{tr(K_2-K_3)}{\tilde{a}^2}+\frac{2(1-k)}{n}\right)+o\left(\frac{1}{n}\right)$$

if h is bounded. We define

$$L\tilde{\tilde{M}}=L\tilde{M}\left(1-\frac{(tr(K_1))^2}{\tilde{a}^2}-\frac{tr(K_2-K_3)}{\tilde{a}^2}\right) \tag{4.3.8}$$

in case h_n is divergent, and

$$L\tilde{\tilde{M}}=L\tilde{M}\left(1-\frac{(tr(K_1))^2}{\tilde{a}^2}-\frac{tr(K_2-K_3)}{\tilde{a}^2}+\frac{2(1-k)}{n}\right) \tag{4.3.9}$$

when h_n is bounded.

We consider the size of the test of (1.2.1) against (2.2.1) based on $L\tilde{\tilde{M}}$, i.e.

$$Pr(L\tilde{\tilde{M}}>z^2_{(\alpha+1)/2}|H_0). \tag{4.3.10}$$

As previously mentioned, the finite sample variance of the mean-adjusted statistic can be larger than that of the non corrected one. From (4.3.8) and (4.3.9), it is straightforward to notice that this might be the case, depending on the choice of W. By Monte Carlo simulations we can assess whether the mean standardisation correction is worthwhile for any particular choice of W and its performance is therefore comparable with that based on Edgeworth corrections.

4.4 Bootstrap correction and simulation results

In this section we report some Monte Carlo simulations to investigate the finite sample performance of the refined tests derived in Sections 4.1, 4.2 and 4.3. The general setting of the Monte Carlo simulation is identical to that described in Section 1.5. In addition, we construct X as an $n\times 3$ matrix (that is, we set $k=3$) whose first column is a column of ones, while each component of the remaining two columns are generated independently from a uniform distribution with support $[0,1]$ and kept fixed over replications.

For both models (1.2.5) and (1.2.15), the empirical sizes of the test of H_0 in (1.2.1) against (2.2.1) based on the usual normal approximation are compared with the same quantities obtained with both the Edgeworth-corrected critical values and Edgeworth-corrected test statistics. Such values are compared also with the empirical size based on the corrected statistics derived according to the procedure described in Section 4.3. In addition, we consider the simulated sizes based on bootstrap critical values.

Before discussing and comparing the simulation results, we outline how the bootstrap critical values have been obtained in this case. Again, we focus on the implementation of the bootstrap procedure, without addressing validity issues. As described in both Sections 2.5 and 3.2, we generate B pseudo-samples Y_j^*, $j = 1, ...B$, and hence B bootstrap statistics

$$LM_j^* = \frac{n^2}{\tilde{a}^2} \left(\frac{Y_j^{*\prime} W Y_j^*}{Y_j^{*\prime} Y_j^*} \right)^2 \qquad j = 1, ...B.$$

The bootstrap quantile w_α^* is defined such that the proportion of LM_j^* that does not exceed w_α^* is α. The bootstrap test rejects H_0 when $LM > w_\alpha^*$. Hence, the size of the test of (1.2.1) based on bootstrap is

$$Pr(LM > w_\alpha^* | H_0). \tag{4.4.1}$$

When dealing with (4.2.1), we modify the previous algorithm accordingly, i.e. we define

$$\tilde{LM}_j^* = \frac{n^2}{a} \left(\frac{u_j^{*\prime} P W P u_j^*}{u_j^{*\prime} P u_j^*} \right)^2, \quad j = 1, ...B,$$

where u_j^* is a vector of independent observations from the $N(0, Y'PY/n)$ distribution. In this case, we denote $\bar{w}_\alpha^*$ the bootstrap $\alpha-$quantile. The size of the test of (1.2.1) based on the bootstrap procedure is then

$$Pr(\tilde{LM} > \bar{w}_\alpha^* | H_0). \tag{4.4.2}$$

Tables 4.1 and 4.2 display the simulated values corresponding to (4.1.9), (4.1.10), (4.1.13), (4.3.6) and (4.4.1) when m is increased monotonically and kept fixed (i.e. when h is "divergent" and "bounded"), respectively. Moreover, Tables 4.3 and 4.4 display the simulated values corresponding to (4.2.9), (4.2.10), (4.2.11), (4.3.10) and (4.4.2) when h is either "divergent" or "bounded", respectively. All the values in Tables 4.1-4.4 have to be compared with the nominal 5%. Similarly to Chapters 2 and 3, in the Tables we denote by "chi square", "Edgeworth", "transformation", "mean-variance correction" and "bootstrap" the simulated values corresponding to (4.1.9)/(4.2.9),(4.1.10)/(4.2.10), (4.1.13)/(4.2.11), (4.3.6)/(4.3.10) and (4.4.1)/(4.4.2), respectively.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
chi square	0.032	0.036	0.038	0.037
Edgeworth	0.040	0.039	0.041	0.042
transformation	0.045	0.048	0.046	0.048
mean-variance correction	0.035	0.037	0.041	0.042
bootstrap	0.054	0.046	0.047	0.053

Table 4.1: Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.5) when the sequence h is "divergent". The reported values have to be compared with the nominal 0.05.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
chi square	0.034	0.036	0.037	0.037
Edgeworth	0.041	0.042	0.047	0.048
transformation	0.034	0.045	0.048	0.050
mean-variance correction	0.041	0.043	0.046	0.052
bootstrap	0.063	0.052	0.051	0.052

Table 4.2: Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.5) when the sequence h is "bounded". The reported values have to be compared with the nominal 0.05.

From Tables 4.1 and 4.2 we notice that the approximation entailed by the first order asymptotic theory does not work well in practice. Indeed, the nominal 5% is underestimated for all sample sizes and whether h is "divergent" or "bounded", although in the latter case the convergence to the nominal value appears to be faster, as expected. On the other hand, all the corrections we consider improve upon the approximation. In particular, when h is "divergent" (Table 4.1) the corrections based on the Edgeworth-corrected test statistic and bootstrap critical values appear to outperform the others, at least for the sample sizes considered here. On average across sample sizes, the simulated sizes based on Edgeworth-corrected statistic and bootstrap critical values are 77% and 75%, respectively, closer to the nominal 0.05 than the values based on standard LM statistic. Improvements entailed by Edgeworth-corrected critical values and mean-variance correction, instead, are only 32% and 21%, respectively. A similar pattern holds in case h is "bounded" (Table 4.2), although the discrepancy among the performance of the different corrections is less glaring. The difference between the nominal 0.05 and simulated sizes based on Edgeworth-corrected critical values, Edgeworth-corrected statistics, mean-variance corrections and bootstrap critical values are, on average across sample sizes, 62%, 62%, 61% and 70% lower than the difference between the nominal 0.05 and the simulated sizes based on the standard LM statistic. The latter result was expected since, as previously mentioned, the rate of convergence of the cdf of LM to the χ^2 cdf is faster in this case.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
chi square	0.023	0.027	0.036	0.027
Edgeworth	0.044	0.048	0.046	0.047
transformation	0.055	0.049	0.047	0.049
mean-variance correction	0.030	0.034	0.032	0.038
bootstrap	0.045	0.052	0.056	0.051

Table 4.3: Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.15) when the sequence h is "divergent". The reported values have to be compared with the nominal 0.05.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
chi square	0.025	0.035	0.038	0.036
Edgeworth	0.031	0.044	0.047	0.052
transformation	0.033	0.042	0.047	0.052
mean-variance correction	0.027	0.044	0.046	0.052
bootstrap	0.043	0.047	0.048	0.048

Table 4.4: Empirical sizes of the tests of (1.2.1) against (2.2.1) for model (1.2.15) when the sequence h is "bounded". The reported values have to be compared with the nominal 0.05.

From Tables 4.3 and 4.4 we see that the usual test based on first order asymptotic theory performs even worse than in the previous case. Indeed, when inference is based on the standard χ^2 approximation, on average the difference between the simulated sizes and the nominal 0.05 is 52% larger than in the previous case when h is "divergent", and 16% when h is "bounded". However, the corrections give very satisfactory results. In particular, when h is "divergent", both the test based on Edgeworth-corrected critical values and Edgeworth-corrected statistics appear to perform very well, giving results that are comparable to the bootstrap-based procedure. Specifically, when h is "divergent, the improvements entailed by Edgeworth-corrected critical values, Edgeworth-corrected statistics and bootstrap critical values over the standard LM statistic are 82%, 88% and 81%, respectively (on average across sample sizes). The simulated values corresponding to (4.3.6) are closer to the nominal than ones of the standard test for all sample sizes, but not as satisfactory as the Edgeworth-based results (the improvement over the standard LM is only 18%). This might be due to the variance inflation discussed in Section 4.3. Again, when the sequence h is "bounded", the pattern of the results appears to be very similar: on average across sample sizes, the values for the simulated sizes based on Edgeworth-corrected critical values, Edgeworth-corrected statistics, mean-variance standardization and bootstrap critical values are, respectively, 61%, 60%, 55% and 80% closer to 0.05 than those based on the standard LM statistic.

4.5 The exact distribution

In Sections 4.1 and 4.2 we developed refined procedures for testing (1.2.1) against (2.2.1) based on LM statistics, as given in (4.1.1) and (4.2.1), respectively. It must be mentioned that, since λ is a scalar parameter, we could have focused on the square root of the statistics in (4.1.1) and (4.2.1) and test H_0 against a one-sided alternative. We chose to develop the corrected procedure based on (4.1.1) and (4.2.1), and compare its performance in finite samples with that derived in Robinson (2008b), because in several circumstances we might not have any preliminary evidence about the sign of λ and therefore the standard two-sided LM test might be preferred instead. However, it should be stressed that in case a test against a one-sided alternative is justified, suitable Edgeworth-corrections can be derived by a relatively straightforward modification of the proofs of either Theorems 4.1 or 4.2.

In this section we investigate numerically the properties of the distribution under H_0 of the square root of both (4.1.1) and (4.2.1), denoted by T and $\tilde{T}$ respectively, by means of Imhof's procedure and compare the results with those obtained using Edgeworth correction terms. The numerical evaluation of the cdf of T and $\tilde{T}$ and the corresponding quantiles, despite the obvious limitations of numerical algorithms, provides some information about the true distribution of the statistics and, to some extent, confirms the accuracy of Edgeworth corrections.

Since the numerical procedure is implemented using W given in (1.1.2), we describe the algorithm for a symmetric weight matrix, although it can be easily generalised to any choice of W. Moreover, we will describe the numerical procedure for evaluating the cdf of T, but the same argument with minor, obvious, modifications holds for $\tilde{T}$.

As discussed in the proof of Theorem 4.1, we can write $Pr(T \leq \zeta) = Pr(\epsilon' C \epsilon \leq 0)$, where $C = W - I\zeta\tilde{a}/n$ (that is (4.A.2) with $x = \tilde{a}\zeta$).

When the cdf can be written in terms of a quadratic form in normal random variables, as is the case in the last displayed expression, a procedure to evaluate it by numerical inversion of the characteristic function has been developed by Imhof (1961) and then improved and extended to different contexts by several authors. For the purpose of our implementation, we rely on the work by Imhof (1961), Davies (1973), Davies (1980), Ansley et al. (1992) and on the survey of Lu and King (2002).

Let s be the number of distinct eigenvalues of $\sigma^2 C$, which are denoted by μ_j for $j = 1, .., s$, while n_j for $j = 1, ..., s$ is their order of algebraic multiplicity. Starting from the inversion formula of Gil-Pelaez (1951), Imhof (1961) suggests to evaluate the cdf of $\epsilon' C \epsilon$ as

$$Pr(\epsilon' C \epsilon \leq 0) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{sin\theta(u)}{u\gamma(u)} du, \tag{4.5.1}$$

where

$$\theta(u) = \sum_{j=0}^{s} \left(\frac{n_j}{2} tg^{-1}(2u\mu_j) \right) \quad \text{and} \quad \gamma(u) = \prod_{j=1}^{s} (1 + 4u^2\mu_j^2)^{n_j/4}.$$

The integral on the RHS of (4.5.1) cannot be evaluated using standard analytical methods because of the oscillatory nature of the integrand function and numerical procedures should be employed instead.

As suggested in Lu and King (2002), we rely on the discretisation rule provided by Davies (1973), which is based on a trapezoidal approximation for the integral on the RHS of (4.5.1), i.e.

$$Pr(\epsilon' C\epsilon \leq 0) = \frac{1}{2} - \sum_{m=0}^{M} \frac{sin\theta((m+\frac{1}{2})\Delta)}{\pi(m+\frac{1}{2})\gamma((m+\frac{1}{2})\Delta)}, \tag{4.5.2}$$

where Δ is the step interval and M is related to the truncation point, denoted by U henceforth, by the relationship $U = (M+1/2)\Delta$. Both Δ and U need to be determined numerically.

We denote by $MGF(t)$ the moment generating function of $\epsilon' C\epsilon$. In order to evaluate Δ, we solve numerically the equation

$$MGF(t) - tMGF^{(1)}(t) - ln(E_I) = 0, \tag{4.5.3}$$

where $MGF^{(1)}(t) = dMGF(t)/dt$ and E_I is the maximum allowable integration error. It can be shown (see e.g. Ansley et al.(1992)) that the last displayed equation has always two solutions $t_1 > 0$ and $t_2 < 0$, both satisfying the constraint $(1 - 2t_i\mu_j) > 0, \ \ \forall j = 1, ...s$, and $i = 1, 2$. For $i = 1, 2$, we define

$$\Delta_i = sign(t_i)\frac{2\pi}{MGF^{(1)}(t)|t = t_i}.$$

We choose Δ appearing in the RHS of (4.5.2) as the minimum value of Δ_i, for $i = 1, 2$.

We briefly mention the algorithm to determine U, for more details we see Lu and King (2002). It is possible to show that the function $u\gamma(u)$ in (4.5.1) is strictly increasing, while $|sin\theta(u)|$ is bounded. Hence, there exists a function $\xi(U)$ such that

$$|\frac{1}{\pi}\int_{U}^{\infty} \frac{sin\theta(u)}{u\gamma(u)}du| \leq \xi(U) \leq E_T,$$

where E_T is the maximum allowable truncation error. U is then derived as the numerical solution of

$$ln\xi(U) - lnE_T = 0. \tag{4.5.4}$$

Several functional forms for $\xi(U)$ have been proposed in the literature. In the present case, we implement the procedure using Imhof's truncation bound, that is

$$\xi(U) = \frac{2}{\pi n}\prod_{j=1}^{m}|\mu_j|^{-f_j/2}(2U)^{-n/2}.$$

Our results seem to be insensitive to the choice of $\xi(U)$.

Once both Δ and U are obtained, the cdf of $\epsilon' C \epsilon$ using (4.5.2) can be evaluated. As suggested in Davies (1973), we set tolerance $E = 10^{-6}$ and choose $E_I = 0.1E$ and $E_T = 0.9E$.

In order to calculate the α-quantile of the cdf of T, we need to find ζ so that

$$Pr(T \leq \zeta) = \alpha,$$

where the LHS of the last displayed expression can be obtained, as a function of ζ, by the algorithm described above. However, in the present case, the numerical solution to calculate ζ is particularly troublesome since the approximated cdf of T is almost flat as ζ varies.

Although Imhof's framework to obtain the cdf and its quantiles is useful to some extent, it obviously relies heavily on several numerical solutions of highly non-linear equations, such as (4.5.3) and (4.5.4). Hence it cannot be preferred to analytical procedures that improve upon the approximation given by the central limit theorem, such as those based on Edgeworth expansions or on mean and variance standardization. However, despite being not fully reliable, quantiles obtained with Imhof's procedure can be compared with Edgeworth-corrected ones, to provide further evidence that the latter are closer to the true values than those of the normal cdf.

Edgeworth-corrected quantiles of the cdf of T can be obtained from intermediate results reported in the proof of Theorem 4.1 and a procedure similar to that described in Section 4.1. Specifically, in the Appendix we derive the Edgeworth expansion for the cdf of T as

$$Pr\left(T \leq \zeta\right) \quad = \quad \Phi(\zeta) - \frac{\bar{\kappa}}{3!} H_2(\zeta)\phi(\zeta) + o\left(\sqrt{\frac{h}{n}}\right),$$

where $\bar{\kappa} = tr(W' + W)^3 / \tilde{a}^3$. From the last displayed expression we can derive a corresponding expansion for the $\alpha-$quantile by a straightforward modification of the argument presented in Section 4.1. We denote the true $\alpha-$quantile of the cdf of T by w_α^T and write

$$w_\alpha^T = z_\alpha + \frac{\bar{\kappa}}{3!} H_2(\zeta) + o\left(\sqrt{\frac{h}{n}}\right),$$

whether h is either divergent or bounded.

		$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$
$m = 8$	Edgeworth	1.9334	2.4403	3.0715
$r = 5$	Imhof	1.8620	2.3250	2.9000
$m = 12$	Edgeworth	1.8925	2.3722	2.9658
$r = 8$	Imhof	1.8430	2.3100	2.8850
$m = 18$	Edgeworth	1.8668	2.3294	2.8994
$r = 11$	Imhof	1.8310	2.2880	2.8550
$m = 28$	Edgeworth	1.8482	2.2985	2.8514
$r = 14$	Imhof	1.8200	2.2700	2.8250

Table 4.5: Edgeworth-corrected and Imhof's α-quantiles of the cdf of T in when h is "divergent".

		$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$
$m = 5$	Edgeworth	1.8357	2.2777	2.8191
$r = 8$	Imhof	1.7840	2.1920	2.6800
$m = 5$	Edgeworth	1.7656	2.1609	2.6379
$r = 20$	Imhof	1.7450	2.1280	2.5850
$m = 5$	Edgeworth	1.7303	2.1021	2.5465
$r = 40$	Imhof	1.7200	2.0860	2.5200
$m = 5$	Edgeworth	1.7053	2.0605	2.4819
$r = 80$	Imhof	1.7010	2.0530	2.4730

Table 4.6: Edgeworth-corrected and Imhof's α-quantiles of the cdf of T when h is "bounded".

As expected, from Tables 4.5 and 4.6 we notice that for all sample sizes and for h being either divergent or bounded, the Edgeworth-corrected quantiles for $\alpha = 0.95, 0.975, 0.99$ are closer to those obtained by Imhof's procedure than ones of the standard normal cdf. Indeed, the standard normal quantiles are significantly lower than Imhof's ones for all sample sizes. To some extent, this confirms that tests based on Edgeworth-corrected critical values should be more reliable than those based on the standard normal approximation.

Imhof's algorithm was also implemented to obtain the cdf of $\tilde{T}$. Unfortunately, in this case, the numerical procedure does not work well and it appears to be too sensitive to both the choice of the initial values for the numerical solution of non-linear equations and the choice of X. This give strong motivation to the practitioner to rely on the analytical corrections based on Edgeworth expansions, rather than on numerical procedures to evaluate the exact cdf.

A Appendix

Proof of Theorem 4.1

Given (4.1.2), we start by deriving the formal Edgeworth expansion of the cdf of

$$n\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}. \tag{4.A.1}$$

The development is standard and similar to that presented for the proofs of Theorems 2.1, 2.2 and 2.3. Hence, some of the details are omitted. The cdf of (4.A.1) can be written in terms of a quadratic form in ϵ, i.e.

$$Pr(n\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \leq x) = Pr(\epsilon' C\epsilon \leq 0),$$

where

$$C = \frac{1}{2}(W + W') - \frac{x}{n}I \tag{4.A.2}$$

and x is any real number.

Proceeding as described in detail in the proof of Theorem 2.1, under Assumption 1, we derive the s-th cumulant, κ_s, of $\epsilon' C\epsilon$ as

$$\kappa_1 = \sigma^2 tr(C), \tag{4.A.3}$$

$$\kappa_2 = 2\sigma^4 tr(C^2), \tag{4.A.4}$$

$$\kappa_s = \frac{\sigma^{2s} s! 2^{s-1} tr(C^s)}{s}, s > 2. \tag{4.A.5}$$

From (4.A.3), (4.A.4) and given (4.A.2),

$$\kappa_1 = -\sigma^2 x, \qquad \kappa_2 = \sigma^4\left(tr(W^2 + W'W) + \frac{2}{n}x^2\right) = \sigma^4\left(\tilde{a}^2 + \frac{2}{n}x^2\right)$$

and hence the first centred cumulant, denoted κ_1^c, becomes

$$\kappa_1^c = \frac{-x}{\tilde{a}\left(1 + \frac{2}{n\tilde{a}^2}x^2\right)^{1/2}}. \tag{4.A.6}$$

We set

$$x = \tilde{a}\zeta, \tag{4.A.7}$$

where ζ, as usual, denotes any real number. Under Assumption 3, $x \sim \sqrt{n/h}$. By Taylor expansion of the denominator of (4.A.6) we obtain

$$\kappa_1^c \;=\; -\zeta\left(1 - \frac{1}{n}\zeta^2\right) + o\left(\frac{1}{n}\right).$$

Moreover, under Assumption 3,

$$\kappa_3^c = \frac{8\sigma^6 tr(C^3)}{\kappa_2^{3/2}} \sim \frac{tr(W'+W)^3}{\tilde{a}^3} \sim \sqrt{\frac{h}{n}}$$

and

$$\kappa_4^c = \frac{48\sigma^8 tr(C^4)}{(\kappa_2)^2} \sim \frac{3tr(W'+W)^4}{\tilde{a}^4} \sim \frac{h}{n}. \tag{4.A.8}$$

By Taylor expansion we have

$$\Phi(-\kappa_1^c) = \Phi(\zeta) + O\left(\frac{1}{n}\right) = \Phi(\zeta) + o\left(\frac{h_n}{n}\right)$$

when h is divergent and

$$\Phi(-\kappa_1^c) \quad = \quad \Phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) + o\left(\frac{1}{n}\right)$$

when h is bounded.

Proceeding as in the proof of Theorem 2.1, for x given in (4.A.7) and when h is divergent, the Edgeworth expansion of the cdf of (4.A.1) under H_0 is

$$\begin{aligned} Pr\left(n\tilde{a}^{-1}\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \le \zeta\right) =& \Phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{h}{n}\right) \\ =& \Phi(\zeta) - \frac{\kappa_3^c}{3!}H_2(\zeta)\phi(\zeta) - \frac{\kappa_4^c}{4!}H_3(\zeta)\phi(\zeta) + o\left(\frac{h}{n}\right), \end{aligned} \tag{4.A.9}$$

where the last equality follows by (2.1.9). Similarly, when h is bounded,

$$\begin{aligned} Pr\left(n\tilde{a}^{-1}\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \le \zeta\right) =& \Phi(\zeta) - \frac{\zeta^3}{n}\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{1}{n}\right) \\ =& \Phi(\zeta) - \frac{\kappa_3^c}{3!}H_2(\zeta)\phi(\zeta) - \left(\frac{\zeta^3}{n} + \frac{\kappa_4^c}{4!}H_3(\zeta)\right)\phi(\zeta) + o\left(\frac{1}{n}\right). \end{aligned} \tag{4.A.10}$$

For notational simplicity, let $T = n\tilde{a}^{-1}\epsilon' W\epsilon/\epsilon'\epsilon$, so that $LM = T^2$. Term by term differentiation of (4.A.9) and (4.A.10) gives the corresponding expressions for the pdf of T, $f_T(\zeta)$, i.e.

$$f_T(\zeta) = \phi(\zeta) - \frac{\kappa_3^c}{3!}(-\zeta^3 + 3\zeta)\phi(\zeta) - \frac{\kappa_4^c}{4!}(-\zeta^4 + 6\zeta^2 - 3)\phi(\zeta) + o\left(\frac{h}{n}\right) \tag{4.A.11}$$

and

$$f_T(\zeta) = \phi(\zeta) + \frac{1}{n}(\zeta^4 - 3\zeta^2)\phi(\zeta) - \frac{\kappa_3^c}{3!}(-\zeta^3 + 3\zeta)\phi(\zeta) - \frac{\kappa_4^c}{4!}(-\zeta^4 + 6\zeta^2 - 3)\phi(\zeta) + o(\frac{1}{n}), \tag{4.A.12}$$

respectively.

For divergent h, using (4.A.11), we can derive an approximate expression for the characteristic function of T^2 as

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{itv^2}e^{-\frac{v^2}{2}}(1-\frac{\kappa_3^c}{3!}(-v^3+3v)-\frac{\kappa_4^c}{4!}(-v^4+6v^2-3))dv\\ =&\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}(1-\frac{\kappa_3^c}{3!}(-v^3+3v)-\frac{\kappa_4^c}{4!}(-v^4+6v^2-3))dv. \qquad (4.A.13)\end{aligned}$$

We notice that the first term of the last displayed integral is

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}dv=(1-2it)^{-1/2},$$

which is the χ^2 characteristic function. By Gaussian integration, the second and third terms are, respectively,

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}\frac{\kappa_3^c}{3!}v^3dv=0$$

and

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}\frac{\kappa_3^c}{3!}3vdv=0,$$

while

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}v^4dv=\frac{3}{(1-2it)^{5/2}},\quad \frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}v^2dv=\frac{1}{(1-2it)^{3/2}}.$$

Collecting the previously displayed results, (4.A.13) becomes

$$\frac{1}{\sqrt{1-2it}}+\frac{\kappa_4^c}{8}\frac{1}{\sqrt{1-2it}}-\frac{\kappa_4^c}{4}\frac{1}{(1-2it)^{3/2}}+\frac{\kappa_4^c}{8}\frac{1}{(1-2it)^{5/2}}. \qquad (4.A.14)$$

Term by term Fourier inversion of (4.A.14) gives

$$\begin{aligned}Pr(LM\le\eta)=&F(\eta)+\frac{\kappa_4^c}{8}F(\eta)-\frac{\kappa_4^c}{4}F_3(\eta)+\frac{\kappa_4^c}{8}F_5(\eta)+o\left(\frac{h}{n}\right)\\ =&F(\eta)+\frac{\kappa_4^c}{4}\eta f(\eta)-\frac{\kappa_4^c}{12}\eta^2 f(\eta)+o\left(\frac{h}{n}\right). \qquad (4.A.15)\end{aligned}$$

The last displayed equality follows from the recursions (see e.g. Harris (1985))

$$\begin{aligned}f_{k+2}(x)&=xk^{-1}f_k(x),\\ F_{k+2}(x)&=F_k(x)-2xk^{-1}f_k(x), \qquad (4.A.16)\end{aligned}$$

where f_k and F_k denote the χ^2 pdf and cdf with k degrees of freedom, respectively. When no subscript is specified, $k=1$.

Similarly, for bounded h, from (4.A.12) we obtain an approximation for the characteristic function as

$$\frac{1}{\sqrt{1-2it}}+\frac{\kappa_4^c}{8}\frac{1}{\sqrt{1-2it}}-\frac{\kappa_4^c}{4}\frac{1}{(1-2it)^{3/2}}+\frac{\kappa_4^c}{8}\frac{1}{(1-2it)^{5/2}}$$
$$+\frac{1}{n}\frac{3}{(1-2it)^{5/2}}-\frac{3}{n}\frac{1}{(1-2it)^{3/2}}$$

and thus, term by term Fourier inversion gives

$$\begin{aligned}Pr(LM\le\eta)=&F(\eta)+\frac{\kappa_4^c}{8}F(\eta)-\frac{\kappa_4^c}{4}F_3(\eta)+\frac{\kappa_4^c}{8}F_5(\eta)+\frac{3}{n}(-F_3(\eta)+F_5(\eta))+o\left(\frac{1}{n}\right)\\ =&F(\eta)+\frac{\kappa_4^c}{4}\eta f(\eta)-\frac{\kappa_4^c}{12}\eta^2 f(\eta)-\frac{2}{n}\eta^2 f(\eta)+o\left(\frac{1}{n}\right).\end{aligned}\tag{4.A.17}$$

The claim in Theorem 4.1 follows from (4.A.15) and (4.A.17) by letting $\kappa=3tr(W'+W)^4/\tilde{a}^4$, which is the leading term of κ_4^c, as given in (4.A.8).

Proof of Theorem 4.2

Parts of the proof of Theorem 4.2 are similar to Theorem 4.1 and are omitted. We derive the third order Edgeworth expansion of the cdf of

$$n\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon},\tag{4.A.18}$$

where P is defined according to (1.2.17). The cdf of (4.A.18) can be written in terms of a quadratic form in ϵ, i.e.

$$Pr\left(\frac{\epsilon' PWP\epsilon}{\frac{1}{n}\epsilon' P\epsilon}\le z\right)=Pr(\epsilon' C\epsilon\le 0),$$

where

$$C=\frac{1}{2}P(W+W')P-\frac{1}{n}Pz\tag{4.A.19}$$

and z is any real number.

The same argument presented in the proof of Theorem 4.1 for the evaluation of both characteristic and cumulant generating functions holds here with C defined according to (4.A.19) instead of (4.A.2). From (4.A.19),

$$\kappa_1 \;=\; \sigma^2 tr(PW)-\sigma^2\frac{1}{n}tr(P)z=-\sigma^2(tr((X'X)^{-1}X'WX)-\frac{n-k}{n}z).$$

Also, by straightforward algebra,

$$\begin{aligned}\kappa_2&=\sigma^4(tr(WPWP)+tr(W'PWP)+2\frac{n-k}{n}z^2-\frac{4}{n}tr(PW)z)\\ &=\sigma^4(tr((W+W')PWP)+2\frac{n-k}{n^2}z^2-\frac{4}{n}tr(PW)z)\end{aligned}$$

$$= \sigma^4(tr(W^2) + tr(W'W) + \frac{1}{2}tr(X'(W+W')X(X'X)^{-1}X'(W'+W)X(X'X)^{-1})$$
$$- tr(X'(W+W')^2X(X'X)^{-1}) + 2\frac{n-k}{n^2}z^2 + \frac{4}{n}tr((X'X)^{-1}X'WX)z).$$

By (4.2.6), (4.2.7), and (4.2.8), we write

$$\kappa_1 = -\sigma^2 tr(K_1) - \sigma^2 z + \sigma^2\frac{k}{n}z \tag{4.A.20}$$

and

$$\kappa_2 = \sigma^4(\tilde{a}^2 + tr(K_2 - K_3) + 2\frac{n-k}{n^2}z^2 + \frac{4}{n}tr(K_1)z). \tag{4.A.21}$$

Similarly to the proof of Theorem 4.1, we define $f^c = (\epsilon' C\epsilon - \kappa_1)/\kappa_2^{1/2}$ and derive the centred cumulants as $\kappa_s^c = \kappa_s/\kappa_2^{s/2}$. From (4.A.20) and (4.A.21),

$$\kappa_1^c = \frac{-\sigma^2 tr(K_1) - \sigma^2 z + \sigma^2\frac{k}{n}z}{\sigma^2\tilde{a}\left(1 + \frac{tr(K_2)}{\tilde{a}^2} - \frac{tr(K_3)}{\tilde{a}^2} + \frac{2}{\tilde{a}^2}\frac{n-k}{n^2}z^2 + \frac{4}{n}\frac{tr(K_1)z}{\tilde{a}^2}\right)^{1/2}}. \tag{4.A.22}$$

We choose $z = \tilde{a}\zeta$. Under Assumptions 3 and 4, we have $\tilde{a} \sim \sqrt{n/h}$ and

$$z \sim \sqrt{\frac{n}{h}}, \quad \frac{tr(K_1)}{\tilde{a}^2} \sim \frac{h}{n}, \quad \frac{tr(K_2)}{\tilde{a}^2} \sim \frac{h}{n}, \quad \frac{tr(K_3)}{\tilde{a}^2} \sim \frac{h}{n}.$$

Hence, substituting the expression for z in (4.A.22) and performing a standard Taylor expansion of the denominator we obtain

$$\begin{aligned}\kappa_1^c &= -\left(\zeta + \frac{tr(K_1)}{\tilde{a}}\right)\left(1 + \frac{tr(K_3 - K_2)}{2\tilde{a}^2} + o\left(\frac{h}{n}\right)\right)\\ &= -\zeta - \frac{tr(K_1)}{\tilde{a}} - \frac{tr(K_3 - K_2)}{2\tilde{a}^2}\zeta + o\left(\frac{h}{n}\right)\end{aligned}$$

in case h is divergent, and

$$\begin{aligned}\kappa_1^c &= -\left(\zeta + \frac{tr(K_1)}{\tilde{a}} - \frac{k}{n}\zeta\right)\left(1 + \frac{tr(K_3 - K_2)}{2\tilde{a}^2} - \frac{1}{n}\zeta^2 + o\left(\frac{1}{n}\right)\right)\\ &= -\zeta - \frac{tr(K_1)}{\tilde{a}} + \frac{k}{n}\zeta - \frac{tr(K_3 - K_2)}{2\tilde{a}^2}\zeta + \frac{1}{n}\zeta^3 + o\left(\frac{1}{n}\right)\end{aligned}$$

if h is bounded.

Moreover,

$$\kappa_3^c = \frac{8\sigma^6 tr(C^3)}{\kappa_2^{3/2}} \sim \frac{tr(((W+W')P)^3)}{\tilde{a}^3} \sim \sqrt{\frac{h}{n}}$$

and

$$\kappa_4^c = \frac{48\sigma^8 tr(C^4)}{\tilde{\kappa}_2^2} \sim \frac{3tr(((W+W')P)^4)}{\tilde{a}^4} \sim \frac{h}{n}.$$

Therefore,

$$Pr(n\tilde{a}^{-1}\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon} \leq \zeta | H_0) = Pr(\epsilon' C\epsilon \leq 0 | H_0)$$
$$= Pr(f^c \kappa_2^{1/2} + \kappa_1 \leq 0 | H_0) = Pr(f^c \leq -\kappa_1^c)$$
$$= \Phi(-\kappa_1^c) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(-\kappa_1^c) + \ldots. \quad (4.A.23)$$

By Taylor expansion we have

$$\Phi(-\kappa_1^c) = \Phi(\zeta) + \frac{tr(K_1)}{\tilde{a}}\phi(\zeta) + \frac{tr(K_3 - K_2)}{\tilde{a}^2}\zeta\phi(\zeta) + \frac{1}{2}\left(\frac{tr(K_1)}{\tilde{a}}\right)^2 \Phi^{(2)}(\zeta) + o\left(\frac{h}{n}\right)$$

when h is divergent and

$$\begin{aligned}\Phi(-\kappa_1^c) &= \Phi(\zeta) + \frac{tr(K_1)}{\tilde{a}}\phi(\zeta) + \frac{tr(K_3 - K_2)}{\tilde{a}^2}\zeta\phi(\zeta) \\ &- \frac{k}{n}\zeta\phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) + \frac{1}{2}\left(\frac{tr(K_1)}{\tilde{a}}\right)^2 \Phi^{(2)}(\zeta) + o\left(\frac{1}{n}\right)\end{aligned}$$

when h is bounded. Therefore, (4.A.23) becomes

$$\begin{aligned}Pr(n\tilde{a}^{-1}\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon} \leq \zeta | H_0) =& \Phi(\zeta) + \frac{tr(K_1)}{\tilde{a}}\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{tr(K_3 - K_2)}{\tilde{a}^2}\zeta\phi(\zeta) \\ &+ \frac{1}{2}\left(\frac{tr(K_1)}{\tilde{a}}\right)^2 \Phi^{(2)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{h}{n}\right) \\ =& \Phi(\zeta) + \frac{tr(K_1)}{\tilde{a}}\phi(\zeta) - \frac{\kappa_3^c}{3!}H_2(\zeta)\phi(\zeta) + \frac{tr(K_3 - K_2)}{\tilde{a}^2}\zeta\phi(\zeta) \\ &- \frac{1}{2}\left(\frac{tr(K_1)}{\tilde{a}}\right)^2 H_1(\zeta)\phi(\zeta) - \frac{\kappa_4^c}{4!}H_3(\zeta)\phi(\zeta) + o\left(\frac{h}{n}\right),\end{aligned} \quad (4.A.24)$$

where the last equality follows by (2.1.9). Similarly, when h is bounded,

$$\begin{aligned}Pr(n\tilde{a}^{-1}\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon} \leq \zeta | H_0) =& \Phi(\zeta) + \frac{tr(K_1)}{\tilde{a}}\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{tr(K_3 - K_2)}{\tilde{a}^2}\zeta\phi(\zeta) \\ &+ \frac{1}{2}\left(\frac{tr(K_1)}{\tilde{a}}\right)^2 \Phi^{(2)}(\zeta) - \frac{k}{n}\zeta\phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) \\ &+ \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{h}{n}\right) \\ =& \Phi(\zeta) + \frac{tr(K_1)}{\tilde{a}}\phi(\zeta) - \frac{\kappa_3^c}{3!}H_2(\zeta)\phi(\zeta) + \frac{tr(K_3 - K_2)}{\tilde{a}^2}\zeta\phi(\zeta) \\ &- \frac{1}{2}\left(\frac{tr(K_1)}{\tilde{a}}\right)^2 H_1(\zeta)\phi(\zeta) - \frac{k}{n}\zeta\phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) \\ &- \frac{\kappa_4^c}{4!}H_3(\zeta)\phi(\zeta) + o\left(\frac{1}{n}\right).\end{aligned} \quad (4.A.25)$$

For notational convenience, we write $\tilde{T} = n\tilde{a}^{-1}\epsilon' PWP\epsilon / \epsilon' P\epsilon$, so that $\tilde{LM} = \tilde{T}^2$.

Moreover, we recall that $H_1(\zeta)=\zeta$, $H_2(\zeta)=\zeta^2-1$ and $H_3(\zeta)=\zeta^3-3\zeta$.

As discussed in detail in the proof of Theorem 4.1, term by term differentiation of (4.A.24) and (4.A.25) gives

$$\begin{aligned} f_{\tilde{T}}(\zeta) =& \phi(\zeta) - \frac{tr(K_1)}{\tilde{a}}\zeta\phi(\zeta) - \frac{\kappa_3^c}{3!}(-\zeta^3+3\zeta)\phi(\zeta) + \frac{tr(K_3-K_2)}{\tilde{a}^2}(1-\zeta^2)\phi(\zeta) \\ &- \frac{1}{2}\frac{(tr(K_1))^2}{\tilde{a}^2}(1-\zeta^2)\phi(\zeta) - \frac{\kappa_4^c}{4!}(-\zeta^4+6\zeta^2-3)\phi(\zeta) + o\left(\frac{h}{n}\right) \end{aligned} \quad (4.A.26)$$

and

$$\begin{aligned} f_{\tilde{T}}(\zeta) =& \phi(\zeta) - \frac{tr(K_1)}{\tilde{a}}\zeta\phi(\zeta) - \frac{\kappa_3^c}{3!}(-\zeta^3+3\zeta)\phi(\zeta) \\ &+ \frac{tr(K_3-K_2)}{\tilde{a}^2}(1-\zeta^2)\phi(\zeta) - \frac{1}{2}\frac{(tr(K_1))^2}{\tilde{a}^2}(1-\zeta^2)\phi(\zeta) \\ &- \frac{k}{n}(1-\zeta^2)\phi(\zeta) - \frac{1}{n}(3\zeta^2-\zeta^4)\phi(\zeta) - \frac{\kappa_4^c}{4!}(-\zeta^4+6\zeta^2-3)\phi(\zeta) + o\left(\frac{1}{n}\right), \end{aligned} \quad (4.A.27)$$

respectively.

In order to simplify the notation, we define

$$\omega_1 = \frac{tr(K_3-K_2)}{\tilde{a}^2} - \frac{1}{2}\frac{(tr(K_1))^2}{\tilde{a}^2}, \quad \omega_2 = \frac{tr(K_3-K_2)}{\tilde{a}^2} - \frac{1}{2}\frac{(tr(K_1))^2}{\tilde{a}^2} - \frac{k}{n}$$

and

$$\omega_3 = \frac{tr(K_1)}{\tilde{a}} + \frac{\kappa_3^c}{2}.$$

Proceeding as described in the proof of Theorem 4.1, when h is divergent we approximate the characteristic function of $\tilde{T}$ as

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{itv^2}e^{-\frac{v^2}{2}}\left(1-\omega_3 v+\frac{\kappa_3^c}{3!}v^3+\omega_1(1-v^2)-\frac{\kappa_4^c}{4!}(-v^4+6v^2-3)\right)dv \\ =&\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}\left(1-\omega_3 v+\frac{\kappa_3^c}{3!}v^3+\omega_1(1-v^2)-\frac{\kappa_4^c}{4!}(-v^4+6v^2-3)\right)dv \\ =&\frac{1}{\sqrt{1-2it}}\left(1+\omega_1-\frac{\omega_1}{1-2it}+\frac{\kappa_4^c}{8}\frac{1}{(1-2it)^2}-\frac{\kappa_4^c}{4}\frac{1}{1-2it}+\frac{\kappa_4^c}{8}\right). \end{aligned} \quad (4.A.28)$$

By term by term Fourier inversion of (4.A.28) and some standard algebraic manipulation,

$$\begin{aligned} Pr(\tilde{LM}\leq\eta|H_0) =& F(\eta) + \left(\frac{\kappa_4^c}{8}+\omega_1\right)F(\eta) - \left(\omega_1+\frac{\kappa_4^c}{4}\right)F_3(\eta) + \frac{\kappa_4^c}{8}F_5(\eta) + o\left(\frac{h}{n}\right) \\ =& F(\eta) + \left(\frac{\kappa_4^c}{4}\eta - \frac{\kappa_4^c}{12}\eta^2 + 2\omega_1\eta\right)f(\eta) + o\left(\frac{h}{n}\right). \end{aligned} \quad (4.A.29)$$

Similarly, when h is bounded, we have

$$\begin{aligned}
Pr(\tilde{LM} \leq \eta | H_0) &= F(\eta) + \left(\frac{\kappa_4^c}{8} + \omega_2 \right) F(\eta) - \left(\omega_2 + \frac{\kappa_4^c}{4} + \frac{3}{n} \right) F_3(\eta) \\
&+ \left(\frac{\kappa_4^c}{8} + \frac{3}{n} \right) F_5(\eta) + o\left(\frac{1}{n} \right) \\
&= F(\eta) + \left(\frac{\kappa_4^c}{4}\eta - \frac{\kappa_4^c}{12}\eta^2 + 2\omega_2\eta - \frac{2}{n}\eta^2 \right) f(\eta) + o\left(\frac{1}{n} \right). \qquad (4.A.30)
\end{aligned}$$

The claim in Theorem 4.2 follows from (4.A.29) and (4.A.30) by observing that the leading term of κ_4^c is $\kappa = 3tr(W' + W)^4/\tilde{a}^4$. Indeed, each term in $(tr(W + W')^4 P)$ other than $tr((W + W')^4) \sim n/h$ is $O(1)$ by Assumption 4 and Lemma 2.2, and is therefore $o(n/h)$.

Lemma 4.1 *Suppose that for all n, each element x_{ij} of X is non-stochastic and $|x_{ij}| < K$. Moreover, for all sufficiently large n,*

$$0 < c_1 < \underline{\eta}\left(\frac{X'X}{n}\right).$$

It follows that

$$||P||_r + ||P||_c \leq K,$$

where P is defined according to (1.2.17) .

Proof We show that $||X(X'X)^{-1}X'||_r \leq K$. Let x_i' be the ith row of X.

Hence,

$$||X(X'X)^{-1}X'||_r = \max_{1\leq i\leq n}\sum_{j=1}^{n}|(X(X'X)^{-1}X')_{ij}| = \max_{1\leq i\leq n}\sum_{j=1}^{n}|(x_i'(X'X)^{-1}x_j)|$$

$$\leq \max_{1\leq i\leq n}\sum_{j=1}^{n}||x_i'||||(X'X)^{-1}||||x_j|| \leq \max_{1\leq i,j\leq n}||x_i'||||(\frac{1}{n}X'X)^{-1}||||x_j|| \leq \frac{1}{c_1}kK^2 < \infty,$$

since

$$||(\frac{1}{n}X'X)^{-1}|| = \bar{\eta}\left(\left(\frac{1}{n}X'X\right)^{-1}\right) = \frac{1}{\underline{\eta}\left(\frac{1}{n}X'X\right)} \leq \frac{1}{c_1}$$

and

$$\max_{0<i\leq n}||x_i|| = \max_{0<i\leq n}(x_i'x_i)^{1/2} \leq (kK^2)^{1/2} \leq K.$$

By symmetry, $||X(X'X)^{-1}X'||_c \leq K$. Trivially, the same property holds for P in (1.2.17).

Lemma 4.2 *Let ξ be a statistic whose cdf admits the expansion*

$$Pr(\xi \leq \eta) = F(\eta) + \frac{h}{n}s(\eta)f(\eta) + o\left(\frac{h}{n}\right), \tag{4.A.31}$$

where h can be either divergent or bounded and $s(\eta)$ is a polynomial in η, whose leading coefficients are finite and non-zero as $n \to \infty$. We define the function $g(.)$ as

$$g(x) = x + \frac{h}{n}s(x) + \left(\frac{h}{n}\right)^2 Q(x), \quad \text{with} \quad Q(x) = \frac{1}{4}\int\left(\frac{d}{dx}s(x)\right)^2 dx. \tag{4.A.32}$$

We have

$$Pr(g(\xi) \leq \eta) = o\left(\frac{h}{n}\right).$$

Proof It is straightforward to verify that $g(x)$ is strictly increasing, its first derivative

being
$$1+\frac{h}{n}\frac{ds(x)}{dx}+\frac{1}{4}\left(\frac{h}{n}\right)^2\left(\frac{ds(x)}{dx}\right)^2=\left(1+\frac{1}{2}\frac{h}{n}\frac{ds(x)}{dx}\right)^2.$$

Since $g(.)$ is monotonic,

$$Pr(g(\xi)\leq\eta)=Pr(\xi\leq g^{-1}(\eta))=F_1(g^{-1}(\eta))+\frac{h}{n}s(g^{-1}(\eta))f_1(g^{-1}(\eta))+o\left(\frac{h}{n}\right). \tag{4.A.33}$$

Now, by (4.A.32),

$$\eta=g^{-1}\left(\eta+\frac{h}{n}s(\eta)+\left(\frac{h}{n}\right)^2Q(\eta)\right)=g^{-1}(\eta)+\frac{h}{n}\frac{dg^{-1}(x)}{dx}|_{|x=\eta}s(\eta)+o\left(\frac{h}{n}\right), \tag{4.A.34}$$

where the second equality follows by a standard Taylor expansion. We define $q=g^{-1}(x)$. Therefore,

$$\frac{dg^{-1}(x)}{dx}|_{x=\eta}=\left(\frac{dg(q)}{dq}\right)^{-1}|_{x=\eta}=1+O\left(\frac{h}{n}\right), \tag{4.A.35}$$

where the last equality follows by total differentiation of the function $g(.)$ and Taylor expansion. Collecting (4.A.34) and (4.A.35),

$$\eta=g^{-1}(\eta)+\frac{h}{n}s(\eta)+o\left(\frac{h}{n}\right)$$

and hence

$$g^{-1}(\eta)=\eta-\frac{h}{n}s(\eta)+o\left(\frac{h}{n}\right). \tag{4.A.36}$$

Finally, by substitution of (4.A.36) into (4.A.33) and using

$$F(g^{-1}(\eta))=F(\eta)-\frac{h}{n}s(\eta)f(\eta)+o\left(\frac{h}{n}\right),$$

$$f(g^{-1}(\eta))=f(\eta)+O\left(\frac{h}{n}\right),\qquad s(g^{-1}(\eta))=s(\eta)+O\left(\frac{h}{n}\right),$$

we obtain

$$Pr(g(\xi)\leq\eta) = F(\eta)-\frac{h}{n}s(\eta)f(\eta)+\frac{h}{n}s(\eta)f(\eta)+o\left(\frac{h}{n}\right)=F(\eta)+o\left(\frac{h}{n}\right).$$

5 Refined Tests for Mixed SAR Models

Throughout this chapter we suppose that the data follow model (1.1.3). However, we rule out the case where only the intercept and none of the other regressors is relevant (this case was indeed discussed in Section 2.3). Hence, throughout this chapter, $\beta \neq 0$ indicates that at least one component of β other than the intercept is non zero. The main focus of this chapter is to derive finite sample corrections for standard tests of (1.2.1) against (2.1.1) when λ in (1.1.3) is estimated by OLS and IV. We also consider finite sample corrections for testing linear restrictions on β in (1.1.3).

As discussed in Chapter 1, unlike in model (1.2.5) the estimates of λ and β in (1.1.3) converge to the true values at the standard $\sqrt{n}$ rate, regardless of the choice of W. Therefore, the error when the approximation of the cdf of relevant test statistics is based on the normal is the usual $1/\sqrt{n}$. However, in much empirical work, only small or moderately-sized samples are available and this motivates refined tests.

In Sections 5.1 and 5.2 we derive Edgeworth-corrected tests for (1.2.1) based on OLS and IV estimates, respectively. In Section 5.3 we propose a refined procedure to test linear restrictions on β in (1.1.3) based on IV estimates. In Section 5.4 the small sample performance of the new tests is investigated with a Monte Carlo simulation and compared with bootstrap-corrected ones.

5.1 Refined tests based on OLS estimates

In this section we focus on testing (1.2.1) against (2.1.1) when λ and β in (1.1.3) are estimated by OLS. We modify Assumption 2(iii) to

Assumption 2(iv) *Uniformly in* $i, j = 1, ..., n$, $w_{ij} = O(1/h)$, *where* $h = h_n$ *is bounded away from zero for all* n. *Moreover,* $\sqrt{n}/h = O(1)$ *and and* $h = O(n)$.

Assumption 2(iv) provides conditions on W under which OLS estimates of λ and β in (1.1.3) are consistent (Lee (2002)), as discussed in Chapter 1. For instance, when W is chosen as in (1.1.2), Assumption 2(iv) rules out not only the case where each household has a finite number of neighbors in the limit, but also the case where the number of neighbors increases slower than $\sqrt{n}$. It's worth stressing that, unlike the case of pure SAR considered in Chapter 2, under Assumptions 2(i), 2(ii) and 2(iv), $\hat{\lambda}$ in (1.1.3) is consistent and asymptotically normal for every $\lambda \in (-1, 1)$. However, we will show that it is not possible to derive Edgeworth-based corrections for testing the general null hypothesis given in (3.0.1) against (3.0.2) and hence we will focus here on the test of (1.2.1).

$(\hat{\lambda}, \hat{\beta}')'$ can be conveniently written as

$$\begin{pmatrix} \hat{\lambda} - \lambda \\ \hat{\beta} - \beta \end{pmatrix} = M^{-1}u, \tag{5.1.1}$$

where

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix},$$

with

$$m_{11} = Y'W'WY, \quad m_{12=}m_{21}' = Y'W'X, \quad m_{22} = X'X, \tag{5.1.2}$$

and

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

with $u_1 = Y'W'\epsilon$ and $u_2 = X'\epsilon$. We denote

$$M^{-1} = \begin{bmatrix} m^{11} & m^{12} \\ m^{21} & m^{22} \end{bmatrix},$$

where m^{11}, m^{12}, m^{21} and m^{22} are obtained by the standard formulae for the inverse of a partitioned matrix, i.e.

$$\begin{aligned} m^{11} &= (m_{11} - m_{12}m_{22}^{-1}m_{21})^{-1}, \qquad m^{12} = m^{21'} = -m^{11}m_{12}m_{22}^{-1}, \\ m^{22} &= m_{22}^{-1}(I + m_{21}m^{11}m_{12}m_{22}^{-1}). \end{aligned}$$

Let P be the projection matrix defined in (1.2.17). We impose the following assumption on X.

Assumption 5 *For all n, each element x_{ij} of X is predetermined and $|x_{ij}| \leq K$. Moreover,*

$$0 < c_1 < \underline{\eta}(\left(\frac{X'X}{n}\right)$$

for all sufficiently large n and at least one element of $X'WX/n$ is non zero in the limits. Finally, $\forall \beta \neq 0$, the limits of $\beta'X'W'PWX\beta/n$ and $\beta'X'W'PWPWX\beta/n$ are non zero.

Assumption 5 is similar to Assumption 4 in Chapter 4 We should stress that, although the first part of Assumption 5 is the same as Assumption 4, the second part of the latter would not be sufficient to guarantee $\lim \beta'X'W'PWX\beta/n \neq 0$ and $\lim \beta'X'W'PWPWX\beta/n \neq 0$ as $n \to \infty$. In particular, the limits displayed in Assumption 5 are finite by Lemma 1.2, after observing that $||P||_r + ||P||_c \leq K$ by Lemma

4.1. Assumption 5 requires that such limits are also non zero. We also stress that Assumption 5 is sufficient to guarantee that M/n has a finite limit under H_0 and is therefore invertible. Indeed, since $Y = X\beta + \epsilon$ under H_0, under Assumption 2(iv) it is easy to verify (Lee (2002)) that

$$\frac{1}{n}M = \frac{1}{n}\begin{pmatrix} E(Y')W' \\ X' \end{pmatrix}(WE(Y), X) + o_p(1).$$

Finiteness of $(M/n)^{-1}$ can be concluded under Assumption 5 by applying standard formulae for the inverse of a partitioned matrix.

Theorem 5.1 *Suppose that model (1.1.3) and Assumptions 1, 2(i), 2(ii), 2(iv), 3 and 5 hold. Under H_0 in (1.2.1), the cdf of $\hat{\lambda}$, admits the second order formal Edgeworth expansion*

$$Pr\left(\frac{\delta^{1/2}}{\sigma}\hat{\lambda} \leq \zeta | H_0\right) = \Phi(\zeta) - \vartheta_1(\zeta)\phi(\zeta) - \frac{\kappa}{3!}\Phi^{(3)}(\zeta) + O\left(\frac{1}{h}\right) \tag{5.1.3}$$

when $\sqrt{n}/h = o(1)$, and

$$Pr\left(\frac{\delta^{1/2}}{\sigma}\hat{\lambda} \leq \zeta | H_0\right) = \Phi(\zeta) - \vartheta_2(\zeta)\phi(\zeta) - \frac{\kappa}{3!}\Phi^{(3)}(\zeta) + o\left(\frac{1}{\sqrt{n}}\right), \tag{5.1.4}$$

when $h \sim \sqrt{n}$, where

$$\delta = \beta' X' W' P W X \beta \sim n,$$

$$\vartheta_1(\zeta) = -\frac{\sigma tr(X'WX(X'X)^{-1})}{\delta^{1/2}} - \frac{2\sigma\zeta^2}{\delta^{3/2}}\beta' X' W' PWPWX\beta \sim \frac{1}{\sqrt{n}},$$

$$\begin{aligned} \vartheta_2(\zeta) &= -\frac{\sigma tr(X'WX(X'X)^{-1})}{\delta^{1/2}} - \frac{tr(W'PW)}{\delta}\zeta - \frac{2\sigma\zeta^2}{\delta^{3/2}}\beta' X' W' PWPWX\beta \\ &+ \frac{\sigma^2}{2\delta}tr((W'P)^2 + W'PW)\zeta \sim \frac{1}{h}, \end{aligned}$$

$$\kappa = 6\sigma\frac{\beta' X' W' PWPWX\beta}{\delta^{3/2}} \sim \frac{1}{\sqrt{n}}.$$

The proof of Theorem 5.1 in the Appendix.

We should notice that (5.1.4) is justified also when $\sqrt{n}/h = o(1)$, since the extra term $(tr(W'PW)/\delta)\zeta$ and $(\sigma^2/2\delta)tr((W'P)^2 + W'PW)\zeta$ would be $O(1/h)$. The rates of δ, ϑ_1, ϑ_2 and κ follow by Assumption 5 and Lemmas 2.1 and 2.2. Specifically, δ and $\beta' X' W' PWPWX\beta$ have exact rate n under Assumption 5. Also, under Assumption 5, each element of $(X'X)^{-1}$ has rate $1/n$ while at least one component of $X'WX$ has exact rate n. It follows that $tr(X'WX(X'X)^{-1}) = O(1)$ and is non zero. Also,

$tr(W'PW)$and $tr((W'P)^2 + W'PW)$ have exact rate n/h under Assumption 3, their leading terms being $tr(W'W)$ and $tr(W^2 + W'W)$, respectively. For sake of clarity we should mention that, to the extent of Theorem 5.1, Assumption 3 is only needed for establishing the exact rate of the latter two quantities.

We must stress that (5.1.3) and (5.1.4) are infeasible, since σ and β are unknown. In order to apply Theorem 5.1 to derive refined tests for (1.2.1), we substitute $\sqrt{n}-$consistent estimates for σ and β in (5.1.3) and (5.1.4). Heuristically, we expect that a second order Edgeworth expansion is not affected when $\sqrt{n}-$ consistent estimates are used instead of true values, but theoretical justification of such conjecture is beyond the scope of this project and is left for future investigation.

Let

$$\hat{\delta} = \hat{\beta}'X'W'PWX\hat{\beta},$$

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n}, \quad \text{with} \quad \hat{\epsilon} = S(\hat{\lambda})Y - X\hat{\beta},$$

$$\hat{\vartheta}_1(\zeta) = -\frac{\hat{\sigma} tr(X'WX(X'X)^{-1})}{\hat{\delta}^{1/2}} - \frac{2\hat{\sigma}\zeta^2}{\hat{\delta}^{3/2}}\hat{\beta}'X'W'PWPWX\hat{\beta},$$

$$\begin{aligned} \hat{\vartheta}_2(\zeta) &= -\frac{\hat{\sigma} tr(X'WX(X'X)^{-1})}{\hat{\delta}^{1/2}} - \frac{2\hat{\sigma}\zeta^2}{\hat{\delta}^{3/2}}\hat{\beta}'X'W'PWPWX\hat{\beta} \\ &- \left(\frac{tr(W'PW)}{\hat{\delta}} - \frac{\hat{\sigma}^2}{2\hat{\delta}} tr((W'P)^2 + W'PW)\right)\zeta \end{aligned}$$

and

$$\hat{\kappa} = 6\hat{\sigma}\frac{\hat{\beta}'X'W'PWPWX\hat{\beta}}{\hat{\delta}^{3/2}}.$$

Similarly to what discussed in the previous chapters, we can compare the size of test (1.2.1) against (2.1.1) using the standard normal $\alpha-$quantile,

$$Pr\left(\frac{\hat{\delta}^{1/2}}{\hat{\sigma}}\hat{\lambda} > z_\alpha | H_0\right), \tag{5.1.5}$$

with the same quantity based on Edgeworth-corrected $\alpha-$quantile, i.e

$$Pr\left(\frac{\hat{\delta}^{1/2}}{\hat{\sigma}}\hat{\lambda} > z_\alpha + p_1(z_\alpha) | H_0\right) \tag{5.1.6}$$

where $p_1(z_\alpha)$ can be determined from Theorem 5.1. The derivation of the Edgeworth-corrected $\alpha-$ quantile is identical to what was discussed in detail in Chapter 2 and is omitted. We obtain

$$p_1(z_\alpha) = \hat{\vartheta}_1(z_\alpha) + \frac{\hat{\kappa}}{3!}H_2(z_\alpha),$$

when $\sqrt{n}/h = o(1)$, and

$$p_1(z_\alpha) = \hat{\vartheta}_2(z_\alpha) + \frac{\hat{\kappa}}{3!}H_2(z_\alpha),$$

when $h \sim \sqrt{n}$. When inference is based on the corrected quantile, we expect that the error of the approximation decreases from $O(1/\sqrt{n})$ to $o(1/\sqrt{n})$ (or, more precisely, to $O(1/h)$ when $\sqrt{n}/h = o(1)$).

Alternatively, from Theorem 5.1 we can derive the Edgeworth-corrected statistic. Again, the details of derivation are identical to those presented in Chapter 2. For notational convenience, let

$$g_a = \frac{4\hat{\sigma}}{\hat{\delta}^{3/2}}\hat{\beta}'X'W'PWPWX\hat{\beta} + \frac{\hat{\kappa}}{3}$$

and

$$g_b = \frac{tr(W'PW)}{\hat{\delta}} - \frac{\hat{\sigma}^2}{2\hat{\delta}}tr((W'P)^2 + W'PW).$$

From Theorem 5.1, the transformation in (2.1.14) becomes

$$g(x) = x - \hat{\vartheta}_1(x) - \frac{\hat{\kappa}}{3!}(x^2 - 1) + \frac{1}{12}g_a^2 x^3,$$

when $\sqrt{n}/h = o(1)$ and

$$g(x) = x - \hat{\vartheta}_2(x) - \frac{\hat{\kappa}}{3!}(x^2 - 1) + \frac{1}{12}g_a^2 x^3 + \frac{1}{4}g_a g_b x^2 + \frac{1}{4}g_b^2 x,$$

when $h \sim \sqrt{n}$. Therefore we can compare (5.1.5) with

$$Pr\left(g\left(\frac{\hat{\delta}^{1/2}}{\hat{\sigma}}\hat{\lambda}\right) > z_\alpha | H_0\right). \tag{5.1.7}$$

The error when the standard statistic is used has order $1/\sqrt{n}$, while it is reduced to $o(1/\sqrt{n})$ (or to $O(1/h)$ when $\sqrt{n}/h = o(1)$) when considering the corrected variant.

In Section 5.4 we will discuss some Monte Carlo simulation to assess the finite sample performance of the proposed new tests.

Before concluding this section, a remark on the limitations of the OLS estimator is necessary. We have mentioned at the beginning of this section that, despite $\hat{\lambda}$ being consistent and asymptotically normal $\forall\lambda \in (-1, 1)$, it is not possible to derive a standard Edgeworth expansion when $\lambda \neq 0$. A more detailed explanation is outlined in the Remark after the proof of Theorem 5.1 (reported in the Appendix). Thus, not only we are unable to derive Edgeworth-corrected tests for (3.0.1), but also we cannot derive improved procedures for testing linear (or possibly non linear) restrictions on β in (1.1.3), such as

$$H_0 : R'\beta = r \qquad H_1 : R'\beta \neq r,$$

where R is a non zero $k-$dimensional vector and r is some constant. Indeed, it is known (Lee (2002)) that OLS is inconsistent in case $\beta = 0$ and $\lambda \neq 0$. Thus, it is impossible to test the joint significance of the components of β, $\beta_1, ...\beta_k$, based on OLS estimates even when we rely on standard first order asymptotic theory. However, we might be interested in testing the joint significance of a subset of $\beta_1, ...\beta_k$. In this case, under the null, OLS estimates would be consistent and asymptotically normal and therefore standard tests based on the normal approximation could be performed. However, since a standard Edgeworth expansion cannot be derived when $\lambda \neq 0$, finite sample improvements of the performance of such tests are not possible.

Another drawback of the OLS estimation is the impossibility of deriving a standard Edgeworth expansion of order higher than two. Details are again outlined in the Remark reported in the Appendix. Consequently, we cannot improve upon the approximation offered by the central limit theorem in case the alternative hypothesis is given by (2.2.1).

Finally, we should stress that the proposed procedure does not extend to the case of $h/\sqrt{n} = o(1)$. It is therefore clear that, even though the proposed finite sample corrections based on OLS might perform well, they are possible only in very specific settings. In the following section we will discuss improved tests based on IV estimates, which are expected to be more general.

5.2 Refined tests based on IV estimates

We suppose that the data follow model (1.1.3). In Section 5.1, we proposed Edgeworth-corrected tests for H_0 in (1.2.1) when $(\lambda, \beta')'$ are estimated by OLS and we noted that such corrected statistics can be derived only in very specific cases. In this section we focus on statistics based on $\hat{\lambda}_{IV}$ and $\hat{\beta}_{IV}$.

Similarly to what was outlined in Chapter 1, let $Z = (Z_1, Z_2)$ be an $n \times (k+1)$ matrix of instruments. $(\hat{\lambda}_{IV}, \hat{\beta}'_{IV})'$ is defined as

$$\begin{pmatrix} \hat{\lambda}_{IV} \\ \hat{\beta}_{IV} \end{pmatrix} = \left(\begin{pmatrix} Z_1' \\ Z_2' \end{pmatrix} (WY, X) \right)^{-1} \begin{pmatrix} Z_1' \\ Z_2' \end{pmatrix} Y.$$

As discussed in Chapter 1, the "ideal" choice of Z is $(E(WY), X)$ (see e.g. Kelejian and Prucha (1998) or Lee (2003)). A discussion about the optimal construction of Z_1 is beyond the scope of this work (and it would be, generally, data-dependent) and we refer to Kelejian and Prucha (1998) and Lee (2003) for further references.

For simplicity, let

$$\begin{aligned} &\delta_{IV,1} = Z_1' PWS^{-1}(\lambda)X\beta \quad \delta_{IV,2} = Z_1' PWS^{-1}(\lambda)PZ_1, \\ &\delta_{IV,3} = Z_1' PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1. \end{aligned} \tag{5.2.1}$$

We introduce the following Assumptions.

Assumption 6 $\forall\lambda \in (-1,1)$, $\||S^{-1}(\lambda)||_r + ||S^{-1}(\lambda)||_c \leq K$, *i.e.* $S^{-1}(\lambda)$ *is uniformly bounded in both row and column sums in absolute value.*

Existence of $S^{-1}(\lambda)$ $\forall\lambda \in (-1,1)$ follows from (2.4.2) under Assumption 2(i).

Assumption 7

(i) *For all n, each element x_{ij} of X is predetermined and $|x_{i,j}| \leq K$. For all n, each element $z_{1,i}$ of the $n-$dimensional, non-null column vector Z_1 is constant and $|z_{1,i}| \leq K$. Let $Z = (Z_1, X)$. Also,*

$$0 < c_1 < \underline{\eta}\left(\frac{Z'Z}{n}\right)$$

for all sufficiently large n.

(ii) $\forall\beta \neq 0$, $\forall Z_1 \neq 0$,

$$\lim_{n\to\infty}\frac{1}{n}\delta_{IV,1}, \quad \lim_{n\to\infty}\frac{1}{n}\delta_{IV,2}, \quad \lim_{n\to\infty}\frac{1}{n}\delta_{IV,3} \tag{5.2.2}$$

are non zero. Moreover, $\forall\beta \neq 0$, at least one element of $X'WS^{-1}X\beta/n$ has a non-zero limit.

The limits displayed in (5.2.2) are finite by Lemma 2.2. Assumption 7(ii) imposes that they are also non-zero.

By standard algebra, we obtain

$$(\hat{\lambda}_{IV} - \lambda, (\hat{\beta}_{IV} - \beta)')' = M_{IV}^{-1}u_{IV}, \tag{5.2.3}$$

where

$$M_{IV} = \begin{pmatrix} m_{IV,11} & m_{IV,12} \\ m_{IV,21} & m_{IV,22} \end{pmatrix},$$

with $m_{IV,11} = Z_1'WY$, $m_{IV,12} = Z_1'X$, $m_{IV,21} = X'WY$, $m_{IV,22} = X'X$, and

$$u_{IV} = \begin{pmatrix} Z_1'\epsilon \\ X'\epsilon \end{pmatrix}.$$

Also,

$$M_{IV}^{-1} = \begin{pmatrix} m_{IV}^{11} & m_{IV}^{12} \\ m_{IV}^{21} & m_{IV}^{22} \end{pmatrix},$$

where m_{IV}^{11}, m_{IV}^{12}, m_{IV}^{21} and m_{IV}^{22} can be obtained by standard results on the inverse of a partitioned matrix.

Similarly to Section 5.1, Assumption 7 is sufficient to guarantee invertibility of

M_{IV}/n as $n \to \infty$. Indeed,

$$\frac{1}{n}M_{IV} = \frac{1}{n}\left(\begin{array}{c} Z_1' \\ X' \end{array}\right)(WEY, X) + o_p(1).$$

Finiteness of $(M_{IV}/n)^{-1}$ can then be established under Assumption 7 by applying formulae for the inverse of a partitioned matrix to

$$\left(\begin{array}{c} Z_1' \\ X' \end{array}\right)(WEY, X).$$

We have the following result

Theorem 5.2 *Suppose that model (1.1.3) and Assumptions 1, 2(i), 2(ii), 6 and 7 hold. The cdf of* $\hat{\lambda}_{IV} - \lambda$*, admits the third order formal Edgeworth expansion*

$$\begin{aligned} Pr\left(\frac{|\delta_{IV,1}|}{\sigma\sqrt{Z_1'PZ_1}}(\hat{\lambda}_{IV} - \lambda) \leq \zeta\right) &= \Phi(\zeta) + \frac{\sigma\delta_{IV,2}}{|\delta_{IV,1}|(Z_1'PZ_1)^{1/2}}\zeta^2\phi(\zeta) \\ &+ \left(\frac{3}{2}\frac{\delta_{IV,2}^2}{\delta_{IV,1}^2 Z_1'PZ_1} - \frac{1}{2}\frac{\delta_{IV,3}}{\delta_{IV,1}^2}\right)\sigma^2\zeta^3\phi(\zeta) \\ &+ \frac{\sigma^2}{2}\frac{\delta_{IV,2}^2}{\delta_{IV,1}^2(Z_1'PZ_1)}\zeta^4\Phi^{(2)}(\zeta) + O\left(\left(\frac{1}{n}\right)^{3/2}\right), \end{aligned} \tag{5.2.4}$$

where $\delta_{IV,1}$*,* $\delta_{IV,2}$ *and* $\delta_{IV,3}$ *are defined according to (5.2.1).*

The proof of Theorem 5.2 is in the Appendix. Under Assumption 7(ii), $\delta_{IV,1}$, $\delta_{IV,2}$ and $\delta_{IV,3}$ have exact rate n. Moreover, under Assumption 7(i),

$$Z_1'PZ_1 = Z_1'Z_1 - Z_1'X(X'X)^{-1}X'Z_1 \sim n.$$

Hence, the second term in (5.2.4) has exact rate $1/\sqrt{n}$, while third and fourth ones have exact rate $1/n$.

We notice that Theorem 5.2 holds $\forall\lambda \in (-1, 1)$ and therefore, in principle, we are able to use a feasible version of (5.2.4) to derive improved tests of the general null hypothesis given in (3.0.1) against a one-sided alternative. However, $\lambda = 0$ is generally the most interesting value one wishes to test and thus we focus on (1.2.1) against (2.1.1). Nevertheless, it should be stressed that the following results can be extended with very minor modifications to the test of (3.0.1). Under H_0 in (1.2.1) $S^{-1}(\lambda) = I$, and $\delta_{IV,1}$ and $\delta_{IV,2}$ can be simplified accordingly.

In order to obtain a feasible version of (5.2.4), we substitute $\sqrt{n}-$consistent estimates of β and σ into (5.2.4) without providing theoretical justification, similarly to what we discussed in Section 5.1. Intuitively, we expect the second order expansion being unaffected by replacing true quantities by their estimates, but such conjecture might not hold true for higher order terms of the expansion. Let

$$\hat{\delta}_{IV,1} = Z_1' PWX\hat{\beta}_{IV} \tag{5.2.5}$$

and

$$\hat{\sigma}^2_{IV} = \frac{\hat{\epsilon}'_{IV}\hat{\epsilon}_{IV}}{n}, \quad \text{with} \quad \hat{\epsilon}_{IV} = S(\hat{\lambda}_{IV})Y - X\hat{\beta}_{IV}. \tag{5.2.6}$$

Hence, we can use the feasible, second order, version of (5.2.4), i.e.

$$Pr\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1}}\hat{\lambda}_{IV} \leq \zeta|H_0\right) = \Phi(\zeta) + \frac{\hat{\sigma}_{IV}\delta_{IV,2}}{|\hat{\delta}_{IV,1}|(Z_1'PZ_1)^{1/2}}\zeta^2\phi(\zeta) + O\left(\frac{1}{n}\right), \tag{5.2.7}$$

to derive improved tests of H_0 in (1.2.1).

Let w^{IV}_α be the true $\alpha-$ quantile of the cdf of $(|\hat{\delta}_{IV,1}|/\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1})\hat{\lambda}_{IV}$. Similarly to Section 5.1, (5.2.7) can be used to derive a more accurate approximation for w^{IV}_α than that based on z_α. The derivation is identical to that outlined in Chapter 2 and is omitted. We compare the size based on the standard normal approximation, i.e.

$$Pr\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1}}\hat{\lambda}_{IV} > z_\alpha|H_0\right) = 1 - \alpha + O\left(\frac{1}{\sqrt{n}}\right), \tag{5.2.8}$$

with that obtained with the Edgeworth-corrected $\alpha-$quantile, i.e.

$$Pr\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1}}\hat{\lambda}_{IV} > z_\alpha - \frac{\hat{\sigma}_{IV}\delta_{IV,2}}{|\hat{\delta}_{IV,1}|(Z_1'PZ_1)^{1/2}}z_\alpha^2|H_0\right) = 1 - \alpha + O\left(\frac{1}{n}\right). \tag{5.2.9}$$

From (5.2.7) and (2.1.14) we can derive a corrected test statistic

$$g\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1}}\hat{\lambda}_{IV}\right),$$

where

$$g(x) = x + \frac{\hat{\sigma}_{IV}\delta_{IV,2}}{|\hat{\delta}_{IV,1}|(Z_1'PZ_1)^{1/2}}x^2 + \frac{1}{3}\left(\frac{\hat{\sigma}_{IV}\delta_{IV,2}}{|\hat{\delta}_{IV,1}|(Z_1'PZ_1)^{1/2}}\right)^2 x^3,$$

so that

$$Pr\left(g\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1}}\hat{\lambda}_{IV}\right) \leq \zeta|H_0\right) = \Phi(\zeta) + O\left(\frac{1}{n}\right).$$

Thus, the error in the size when inference is based on the transformed statistic, i.e.

$$Pr\left(g\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1}}\hat{\lambda}_{IV}\right) > z_\alpha|H_0\right), \tag{5.2.10}$$

is expected to be reduced from $O(1/\sqrt{n})$ to $O(1/n)$.

5.3 Finite sample corrections for testing β.

In this section we propose an Edgeworth-improved procedure for testing

$$H_0 : R'\beta = 0, \tag{5.3.1}$$

where R is a $k-$dimensional, non-null column vector, based on IV estimates for λ and β in (1.1.3). For algebraical simplicity, we focus on tesing (5.3.1) against a one sided alternative $H_1 : R'\beta > 0 \quad (< 0)$, even though our procedure can be easily extended to the case of a two-sided alternative. The proposed test can be extended to $H_0 : R'\beta = r$, where r is a non zero constant.

As mentioned in Chapter 1, non nullity of at least one component of β (other than the intercept) is required for the IV estimates to be well defined. Thus, in this framework, we cannot test $H_0 : \beta = 0$. The results derived in this section can therefore be applied to test (5.3.1) provided that $R \neq l_k$. Also, we rule out a test for the joint significance of the components of β other than the intercept, e.g. $H_0 : \beta_2 = \beta_3 = ...\beta_k = 0$, where β_1 corresponds to the intercept.

By (5.2.3),

$$R'(\hat{\beta}_{IV} - \beta) = R'm_{IV}^{21}Z_1'\epsilon + R'm_{IV}^{22}X'\epsilon,$$

where

$$m_{IV}^{21} = -(m_{IV,22})^{-1}m_{IV,21}m_{IV}^{11} = -(X'X)^{-1}X'WY(Z_1'PWY)^{-1},$$

$$\begin{aligned} m_{IV}^{22} &= (m_{IV,22})^{-1} + (m_{IV,22})^{-1}m_{IV,21}m_{IV}^{11}m_{IV,12}(m_{IV,22})^{-1} \\ &= (X'X)^{-1} + (X'X)^{-1}X'WY(Z_1'PWY)^{-1}Z_1'X(X'X)^{-1}, \end{aligned}$$

and

$$m_{IV}^{11} = (m_{IV,11} - m_{IV,12}m_{IV,22}^{-1}m_{IV,21})^{-1} = (Z_1'PWY)^{-1}.$$

Therefore,

$$\begin{aligned} R'(\hat{\beta}_{IV} - \beta) &= -R'(X'X)^{-1}X'WY(Z_1'PWY)^{-1}Z_1'\epsilon + R'(X'X)^{-1}X'\epsilon \\ &+ R'(X'X)^{-1}X'WY(Z_1'PWY)^{-1}Z_1'X(X'X)^{-1}X'\epsilon \\ &= -R'(X'X)^{-1}X'WY(Z_1'PWY)^{-1}Z_1'P\epsilon + R'(X'X)^{-1}X'\epsilon. \end{aligned} \tag{5.3.2}$$

We define

$$Q = X(X'X)^{-1}RZ_1'P, \quad G' = \beta'X'S^{-1}(\lambda)'W'(Q' - Q). \tag{5.3.3}$$

We introduce the following Assumption.

Assumption 8

(i) $R \in \Re^k/\{0\}$.

(ii) $\forall \beta \neq 0,\ Z_1 \neq 0,\ R \neq 0$ *and* $\lambda \in (-1,1)$,

$$\lim_{n\to\infty} \frac{G'G}{n}, \quad \lim_{n\to\infty} \frac{G'(Q-Q')WS^{-1}(\lambda)G}{n}, \quad \lim_{n\to\infty} \frac{G'S^{-1}(\lambda)W'PZ_1}{n}$$

exist and are non zero. Moreover, at least one component of each of the $k-$dimensional vectors $X'WS^{-1}(\lambda)PZ_1/n$ and $Z_1'PWS^{-1}(\lambda)X/n$ is non zero.

Assumption 8(ii) guarantees that the orders of some of the coefficients in the following Theorem hold as exact rates.

We have the following result.

Theorem 5.3 *Suppose that model (1.1.3) and Assumptions 1, 2(i), 2(ii), 6, 7 and 8 hold. Under H_0 in (5.3.1), the cdf of $R'\hat{\beta}_{IV}$ admits the formal second order Edgeworth expansion*

$$\begin{aligned} Pr\left(\frac{|\delta_{IV,1}|}{\sigma(G'G)^{1/2}}R'\hat{\beta}_{IV} \leq \zeta | H_0\right) &= \Phi(\zeta) - \frac{\delta_{IV,1}}{|\delta_{IV,1}|}\frac{\sigma tr(Q-Q')WS^{-1}(\lambda)}{(G'G)^{1/2}}\phi(\zeta) \\ &+ \frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{|\delta_{IV,1}|(G'G)^{1/2}}\zeta^2\phi(\zeta) \\ &- \frac{\delta_{IV,1}}{|\delta_{IV,1}|}\frac{\kappa_{IV,\beta}}{3!}H_2(\zeta)\phi(\zeta) + O\left(\frac{1}{n}\right), \end{aligned} \qquad (5.3.4)$$

where

$$\kappa_{IV,\beta} = \frac{6\sigma^2 G'(Q-Q')WS^{-1}(\lambda)G}{\sigma^2(G'G)^{3/2}} \sim \frac{1}{\sqrt{n}}.$$

The proof of Theorem 5.3 is in the Appendix. Assumption 7(ii) is relevant here only to the extent of imposing an exact rates for $\delta_{IV,1}$ and $X'WS^{-1}(\lambda)X\beta$ (the latter is required to guarantee existence of $(M_{IV}/n)^{-1}$ and is therefore relevant). We notice that under Assumptions 7(ii) and 8(ii) the third and fourth terms in (5.3.4) have exact rate $1/\sqrt{n}$. As discussed in detail in the proof, the second term vanishes when $WS^{-1}(\lambda)$ is symmetric (e.g. when W is symmetric and (2.4.2) holds). In the Appendix we show that, in case W is not symmetric, so that

$$tr(QWS^{-1}(\lambda)) \neq tr(Q'WS^{-1}(\lambda)),$$

$tr((Q-Q')WS^{-1}(\lambda)) = O(1)$. Assumptions 7(i) and 8(ii) guarantee that the latter does not vanish in the limit, so that the second term in (5.3.4) has exact rate $1/\sqrt{n}$.

Similarly to what discussed in Sections 5.1 and 5.2, (5.3.4) depends on the unknown σ, G and $\delta_{IV,1}$. In order to use (5.3.4) to derive improved tests, we replace unknowns with consistent estimates without providing theoretical justification. Specifically, given $\hat{\lambda}_{IV}$ and $\hat{\beta}_{IV}$, let

$$\hat{G} = \hat{\beta}_{IV}' X' S^{-1}(\hat{\lambda}_{IV})' W'(Q' - Q),$$

$$\hat{\delta}_{IV,1} = Z_1' P W S^{-1}(\hat{\lambda}_{IV}) X \hat{\beta}_{IV}$$

and

$$\hat{\kappa}_{IV,\beta} = \frac{6\hat{\sigma}_{IV}^2 \hat{G}'(Q - Q')WS^{-1}(\hat{\lambda}_{IV})\hat{G}}{\hat{\sigma}_{IV}^2(\hat{G}'\hat{G})^{3/2}},$$

where $\hat{\sigma}_{IV}^2$ is given by (5.2.6). Thus, we can use the feasible version of (5.3.4) to derive a more accurate approximation for the $\alpha-$quantile of the cdf of $(|\hat{\delta}_{IV,1}|/\hat{\sigma}_{IV}(\hat{G}'\hat{G})^{1/2})R'\hat{\beta}_{IV}$.

We can compare the size of the test of (5.3.1) based on the normal approximation, i.e.

$$Pr\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}(\hat{G}'\hat{G})^{1/2}} R'\hat{\beta}_{IV} > z_\alpha | H_0\right) = 1 - \alpha + O\left(\frac{1}{\sqrt{n}}\right), \tag{5.3.5}$$

with

$$Pr\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}(\hat{G}'\hat{G})^{1/2}} R'\hat{\beta}_{IV} > z_\alpha + p(z_\alpha) | H_0\right) = 1 - \alpha + O\left(\frac{1}{n}\right), \tag{5.3.6}$$

$$p(x) = \frac{\hat{\delta}_{IV,1}}{|\hat{\delta}_{IV,1}|}\left(\frac{\hat{\sigma}_{IV} tr(Q - Q')WS^{-1}(\hat{\lambda}_{IV})}{(\hat{G}'\hat{G})^{1/2}} - \frac{\hat{\sigma}_{IV}\hat{G}'S^{-1}(\hat{\lambda}_{IV})'W'PZ_1}{\hat{\delta}_{IV,1}(\hat{G}'\hat{G})^{1/2}} x^2 + \frac{\hat{\kappa}_{IV,\beta}}{3!} H_2(x)\right).$$

Similarly, from (2.1.14), we can derive an Edgeworth-corrected statistic so that

$$Pr\left(g\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}(\hat{G}'\hat{G})^{1/2}} R'\hat{\beta}_{IV}\right) > z_\alpha | H_0\right) = 1 - \alpha + O\left(\frac{1}{n}\right), \tag{5.3.7}$$

where

$$\begin{aligned} g(x) &= x - \frac{\hat{\delta}_{IV,1}}{|\hat{\delta}_{IV,1}|}\left(\frac{\hat{\sigma}_{IV} tr(Q - Q')WS^{-1}(\hat{\lambda})}{(\hat{G}'\hat{G})^{1/2}} - \frac{\hat{\sigma}_{IV}\hat{G}'S^{-1}(\hat{\lambda})'W'PZ_1}{\hat{\delta}_{IV,1}(\hat{G}'\hat{G})^{1/2}} x^2 + \frac{\hat{\kappa}_{IV,\beta}}{3!}(x^2 - 1)\right) \\ &+ \frac{1}{12}\left(\frac{2\hat{\sigma}_{IV}\hat{G}'S^{-1}(\hat{\lambda}_{IV})'W'PZ_1}{\hat{\delta}_{IV,1}(\hat{G}'\hat{G})^{1/2}} - \frac{\hat{\kappa}_{IV,\beta}}{3}\right)^2 x^3. \end{aligned}$$

The small sample performance of the proposed tests will be analysed by Monte Carlo in the next section.

5.4 Bootstrap and Monte Carlo results

In this section we assess the finite sample performance of the new tests presented in Sections 5.1, 5.2 and 5.3 by means of a Monte Carlo study. The setting of the simulation is identical to the one described in Section 2.5. X is a $n \times 2$ matrix whose elements are generated from a uniform distribution with support $[0, 1]$ and kept fixed over replications. Furthermore, Z_1 is chosen as one of the column of WX so that $rank(Z_1, X) = 3$.

As already discussed, the size based on Edgeworth-corrected tests can also be compared with the results obtained by the implementation of a bootstrap algorithm. The procedure to obtain the bootstrap critical value based on OLS statistic is very similar to one described in Section 2.5: after imposing H_0 in (1.1.3), we estimate β by OLS and obtain the "restricted" residuals $\hat{\epsilon}_r = Y - X\hat{\beta}$. We then generate B $n-$dimensional vectors $\hat{\epsilon}^*_j$, $j = 1, ...B$ from $N(0, \hat{\epsilon}_r'\hat{\epsilon}_r/n)$ and hence B pseudo-data $Y^*_j = X\hat{\beta} + \hat{\epsilon}^*_j$, $j = 1, ...B$. For every $j = 1, ...B$, we compute bootstrap estimates λ^*_j and β^*_j by regressing Y^*_j on WY^*_j and X, and we construct

$$\sigma_j^{2*} = \frac{(S(\lambda^*_j)Y^*_j - X\beta^*)'(S(\lambda^*_j)Y^*_j - X\beta^*)}{n}.$$

We then obtain B bootstrap OLS null statistics as

$$\frac{\delta_j^{*1/2}}{\sigma^*_j}\frac{Y_j^{*\prime}W'P\hat{\epsilon}^*_j}{Y_j^{*\prime}W'PWY^*_j},$$

where $\delta^*_j = \beta_j^{*\prime}X'W'PWX\beta^*_j$, and compute the $\alpha-$th bootstrap critical value w_α^{OLS*} as the solution of

$$\frac{1}{B}\sum_{j=1}^{B}1\left(\frac{\delta_j^{*1/2}}{\sigma^*_j}\frac{Y_j^{*\prime}W'P\hat{\epsilon}^*_j}{Y_j^{*\prime}W'PWY^*_j} \le w_\alpha^{OLS*}|H_0\right) \le \alpha.$$

The size of the test of (1.2.1) based on w_α^{OLS*} is defined as

$$Pr\left(\frac{\hat{\delta}^{1/2}}{\hat{\sigma}}\hat{\lambda} > w_\alpha^{OLS*}|H_0\right). \qquad (5.4.1)$$

Similarly, after obtaining Y^*_j $(j = 1, ...B)$ as described above, for every j we obtain $\beta^*_{IV,j}$ and $\lambda^*_{IV,j}$ by IV estimation and we construct

$$\sigma^{*2}_{IV,j} = \frac{(S(\lambda^*_{IV,j})Y^*_j - X\beta^*_{IV,j})'(S(\lambda^*_{IV,j})Y^*_j - X\beta^*_{IV,j})}{n}.$$

We obtain B IV null test statistics as

$$\frac{|\delta^*_{IV,1}|}{\sigma^*_{IV,j}\sqrt{Z_1'PZ_1}}\frac{Z_1'P\hat{\epsilon}^*_j}{Z_1'PWY^*_j},$$

where $\delta^*_{IV,1} = Z_1'PWX\beta^*_{IV,j}$ and compute the $\alpha-$th bootstrap quantile w_α^{IV*} as the solution of

$$\frac{1}{B}\sum_{j=1}^{B}1\left(\frac{|\delta^*_{IV,1}|}{\sigma^*_{IV,j}\sqrt{Z_1'PZ_1}}\frac{Z_1'P\hat{\epsilon}^*_j}{Z_1'PWY^*_j} \le w_\alpha^{IV*}|H_0\right) \le \alpha.$$

The size of the test of (1.2.1) based on w_{α}^{IV*} is then

$$Pr\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{Z_1'PZ_1}}\hat{\lambda}_{IV} > w_{\alpha}^{IV*}|H_0\right). \tag{5.4.2}$$

When dealing with testing (5.3.1), the bootstrap procedure based on restricted residuals depends on the choice of R. For instance, if $R = (1,0)'$, i.e. $H_0 : \beta_2 = 0$, we would be able to derive restricted residuals and pseudo-data by a straightforward modification of the procedures described above. For sake of generality, however, we rely on the general framework, without imposing H_0 to obtain residuals (e.g. DiCiccio and Efron (1996)). In particular, we obtain $\hat{\lambda}_{IV}$ and $\hat{\beta}_{IV}$ in (1.1.3) and hence the residuals $\hat{\epsilon}_{IV} = S(\hat{\lambda}_{IV})Y - X\hat{\beta}_{IV}$. We then generate B $n-$dimensional vector $\hat{\epsilon}^*_{IV,j}$, $j = 1, ...B$, from $N(0, \hat{\epsilon}_{IV}'\hat{\epsilon}_{IV}/n)$ and the corresponding pseudo-data

$$Y^*_{IV,j} = S^{-1}(\hat{\lambda}_{IV})(X\hat{\beta}_{IV} + \hat{\epsilon}^*_{IV,j}), \quad j = 1, ..., B.$$

In order to avoid notational confusion, we should notice that $Y^*_{IV,j}$ is obtained by a different procedure than Y^*_j defined above, since the latter was constructed by imposing (1.2.1).

From $Y^*_{IV,j}$, for every $j = 1, ...B$ we obtain $\beta^*_{IV,u,j}$ and $\lambda^*_{IV,u,j}$ by IV estimation and hence

$$\sigma^{*2}_{IV,u,j} = \frac{(S(\lambda^*_{IV,u,j})Y^*_j - X\beta^*_{IV,u,j})'(S(\lambda^*_{IV,u,j})Y^*_j - X\beta^*_{IV,u,j})}{n}.$$

The extra subscript "u" indicates that here the estimates are obtained from "unrestricted" (i.e. obtained without imposing H_0 in (5.3.1)) pseudo-samples. Similarly,

$$\delta^*_{IV,1,u} = Z_1'PWS^{-1}(\lambda^*_{IV,u,j})X\beta^*_{IV,u,j}$$

and

$$G^{*'} = \beta^{*'}_{IV,u,j}X'S^{-1}(\lambda^*_{IV,u,j})'W'(Q' - Q).$$

The size based on the bootstrap $\alpha-$quantile t_{α}^{IV*} is defined as

$$Pr\left(\frac{|\hat{\delta}_{IV,1}|}{\hat{\sigma}_{IV}\sqrt{\hat{G}'\hat{G}}}R'\hat{\beta}_{IV} > t_{\alpha}^{IV*}|H_0\right), \tag{5.4.3}$$

where t_{α}^{IV*} is obtained by solving

$$\frac{1}{B}\sum_{j=1}^{B}1\left(\frac{|\delta^*_{IV,1,u}|}{\sigma^*_{IV,u,j}\sqrt{G^{*'}G^*}}R'(\beta^*_{IV,u,j} - \hat{\beta}_{IV}) \leq t_{\alpha}^{IV*}|H_0\right) \leq \alpha.$$

Similarly to the notation used in Sections 2.5, 3.2 and 4.4, in the Tables we denote by "normal", "Edgeworth", "transformation" and "bootstrap" the simulated values of the size based on normal quantiles, Edgeworth corrected quantiles, Edgeworth-corrected null statistics and bootstrap quantiles.

In particular, in Table 5.1 we report the simulated values corresponding to (5.1.5), (5.1.6), (5.1.7) and (5.4.1). The sample has been increased consistently with the condition $\sqrt{n}/h = o(1)$. We set $\beta = (1, 0.5)'$, but the results are insensitive to the choice of β.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
normal	0	0	0.007	0.017
Edgeworth	0.129	0.078	0.063	0.054
transformation	0.095	0.063	0.055	0.053
bootstrap	0.028	0.040	0.040	0.058

Table 5.1: Empirical sizes of the tests of (1.2.1) against a one-sided alternative when λ and β in (1.1.3) are estimated by OLS. The reported values have to be compared with the nominal 0.05.

From Table 5.1 we observe that, as expected, the normal approximation does not work well since the nominal 5% is greatly underestimated for all sample sizes. On the other hand, the results obtained by both Edgeworth-corrected critical values and the Edgeworth-corrected statistic appear to overestimate 0.05 for very small sample sizes, but the convergence to the nominal value appear to be fast. By a comparison with the bootstrap based results, we notice that in this case the latter outperform the Edgeworth corrections. More precisely, on average across sample sizes the values obtained by Edgeworth-corrected critical values, Edgeworth-corrected statistic and bootstrap critical valeus are 36%, 66% and 72%, respectively, closer to 0.05 than the values obtained by standard inference. Moreover, as sample size increases the difference between actual value and 0.05 decreases, on average, by 62% and 57% when inference is based on Edgeworth-corrected critical values and Edgeworth-corrected statistics, respectively. These figures have to be compared with a decrease of only 12% when we rely on the standard statistic.

In Table 5.2 we report the simulated values corresponding to (5.2.8), (5.2.9), (5.2.10) and (5.4.2). Again, $\beta = (1, 0.5)'$.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
normal	0.023	0.028	0.034	0.033
Edgeworth	0.078	0.065	0.062	0.056
transformation	0.042	0.035	0.060	0.054
bootstrap	0.025	0.030	0.037	0.042

Table 5.2: Empirical sizes of the tests of (1.2.1) against a one-sided alternative when λ and β in (1.1.3) are estimated by IV. The reported values have to be compared with the nominal 0.05.

Similarly to Table 5.1, the results in Table 5.2 based on the normal approximation are below 5% for all sample size, although the discrepancy is less severe. On the other hand, Edgeworth-corrected quantiles and corrected IV statistic appear to improve upon the normal approximation by 29% and 54%, respectively, on average across sample sizes. The performance of Edgeworth corrections in this case appears to be better than the bootstrap one, since in the latter the obtained values are, on average across sample sizes, only 22% closer to 0.05 by than ones obtained by the standard normal approximation.

	$m = 8$		$m = 12$		$m = 18$		$m = 28$	
	$r = 5$		$r = 8$		$r = 11$		$r = 14$	
	λ		λ		λ		λ	
normal	0.1	0.032	0.1	0.041	0.1	0.042	0.1	0.044
	0.7	0.032	0.7	0.031	0.7	0.039	0.7	0.043
	λ		λ		λ		λ	
Edgeworth	0.1	0.071	0.1	0.051	0.1	0.051	0.1	0.047
	0.7	0.068	0.7	0.055	0.7	0.052	0.7	0.051
	λ		λ		λ		λ	
transformation	0.1	0.075	0.1	0.062	0.1	0.054	0.1	0.051
	0.7	0.088	0.7	0.063	0.7	0.048	0.7	0.049
	λ		λ		λ		λ	
bootstrap	0.1	0.055	0.1	0.040	0.1	0.045	0.1	0.046
	0.7	0.061	0.7	0.042	0.7	0.041	0.7	0.043

Table 5.3: Empirical sizes of the tests of (5.3.1) with $R = (0, 1)'$ against a one-sided alternative. λ and β in (1.1.3) are estimated by IV and $\lambda = 0.1, 0.7$. The reported values have to be compared with the nominal 0.05.

	$m=8$ $r=5$		$m=12$ $r=8$		$m=18$ $r=11$		$m=28$ $r=14$	
	λ		λ		λ		λ	
normal	0.1	0.033	0.1	0.031	0.1	0.035	0.1	0.041
	0.7	0.029	0.7	0.032	0.7	0.033	0.7	0.040
	λ		λ		λ		λ	
Edgeworth	0.1	0.070	0.1	0.066	0.1	0.055	0.1	0.047
	0.7	0.059	0.7	0.060	0.7	0.044	0.7	0.052
	λ		λ		λ		λ	
transformation	0.1	0.072	0.1	0.058	0.1	0.055	0.1	0.049
	0.7	0.069	0.7	0.059	0.7	0.053	0.7	0.052
	λ		λ		λ		λ	
bootstrap	0.1	0.064	0.1	0.041	0.1	0.044	0.1	0.046
	0.7	0.059	0.7	0.043	0.7	0.044	0.7	0.043

Table 5.4: Empirical sizes of the tests of (5.3.1) with $R = (-1, 1)'$ against a one-sided alternative. λ and β in (1.1.3) are estimated by IV and $\lambda = 0.1, 0.7$. The reported values have to be compared with the nominal 0.05.

Finally, in Tables 5.4 and 5.5 we display the simulated values of the size of test of (5.3.1) against a one-sided alternative when $R = (0, 1)'$ (i.e. $H_0 : \beta_2 = 0$) and $R = (-1, 1)$ (i.e. $H_0 : \beta_1 = \beta_2$), respectively. For each case, we choose two different values for λ, that is $\lambda = 0.1$ and $\lambda = 0.7$. In both cases the values of the size based on both Edgeworth corrected quantiles and corrected statistics are closer to 5% than the values obtained with the normal approximation. The pattern of the results is similar for the two choices of R and does not appear to change with the value of λ. More specifically, when $R = (0, 1)'$ and $\lambda = 0.1$, the values obtained by using Edgeworth-corrected critical values, Edgeworth-corrected statistic and bootstrap critical values are 52%, 15% and 32% closer, respectively, to the nominal 0.05 than the values obtained by the standard statistic.

A Appendix

Proof of Theorem 5.1

From (5.1.1),

$$\hat{\lambda} - \lambda \quad = \quad m^{11} u_1 + m^{12} u_2 = m^{11}(u_1 - m_{12} m_{22}^{-1} u_2).$$

By standard algebra, we can write the cdf of $\hat{\lambda}$ under H_0 as $\Pr(\hat{\lambda} \leq z) = \Pr(f \leq 0)$, where

$$f = u_1 - m_{12} m_{22}^{-1} u_2 - \frac{z}{m^{11}} \qquad (5.A.1)$$

and z being any real number.

Under H_0, by (5.1.2),

$$m_{11} = (X\beta + \epsilon)'W'W(X\beta + \epsilon)$$

and

$$m_{12} = m_{21}' = (X\beta + \epsilon)'W'X.$$

Therefore,

$$m_{12}m_{22}^{-1}m_{21} = (X\beta + \epsilon)'W'X(X'X)^{-1}X'W(X\beta + \epsilon)$$

and hence

$$m^{11} = ((X\beta + \epsilon)'W'PW(X\beta + \epsilon))^{-1}. \tag{5.A.2}$$

Similarly,

$$u_1 = (X\beta + \epsilon)'W'\epsilon$$

and hence

$$u_1 - m_{12}m_{22}^{-1}u_2 = (X\beta + \epsilon)'W'\epsilon - (X\beta + \epsilon)'W'X(X'X)^{-1}X'\epsilon = (X\beta + \epsilon)'W'P\epsilon. \tag{5.A.3}$$

Substituting (5.A.2) and (5.A.3), (5.A.1) becomes

$$\begin{aligned} f &= \epsilon'(W'P - zW'PW)\epsilon + (\beta'X'W'P - 2z\beta'X'W'P)\epsilon - z\beta'X'W'PWX\beta \\ &= \epsilon'W'P(I - zW)\epsilon + \beta'X'W'P(I - 2zW)\epsilon - z\beta'X'W'PWX\beta \\ &= \frac{1}{2}\epsilon'(C + C')\epsilon + c'\epsilon + d \end{aligned}$$

where

$$C = W'P(I - zW), \tag{5.A.4}$$

$$c' = \beta'X'W'P(I - 2zW), \tag{5.A.5}$$

and

$$d = -z\beta'X'W'PWX\beta. \tag{5.A.6}$$

We define $\delta = \beta'X'W'PWX\beta$, so that $d = -z\delta$.

Under Assumption 1, we can derive the characteristic function of f as

$$\begin{aligned} E(e^{itf}) =& E(e^{it(\frac{1}{2}\epsilon'(C+C')\epsilon + c'\epsilon + d)}) \\ =& \frac{e^{itd}}{(2\pi)^{n/2}\sigma^n} \int_{\Re^n} e^{it\frac{1}{2}\xi'(C+C')\xi + itc'\xi} e^{-\frac{\xi'\xi}{2\sigma^2}} d\xi \\ =& \frac{e^{itd}}{(2\pi)^{n/2}\sigma^n} \int_{\Re^n} e^{-\frac{1}{2}(\frac{\xi}{\sigma} - q)'(I - it\sigma^2(C+C'))(\frac{\xi}{\sigma} - q)} e^{\frac{1}{2}q'(I - it\sigma^2 C)q} d\xi, \end{aligned} \tag{5.A.7}$$

where q satisfies $itc' = q'(I - it\sigma^2(C + C'))/\sigma$. By standard algebra,

$$q = (I - it\sigma^2(C + C'))^{-1} itc\sigma,$$

and hence

$$q'(I - it\sigma^2(C + C'))q = -t^2 c'(I - it\sigma^2(C + C'))^{-1} c\sigma^2.$$

By Gaussian integration, (5.A.7) becomes

$$\begin{aligned} E(e^{itf}) =& e^{itd - \frac{\sigma^2}{2} t^2 c'(I - it\sigma^2(C+C'))^{-1} c} det(I - it\sigma^2(C + C'))^{-1/2} \\ =& e^{itd - \frac{\sigma^2}{2} t^2 c'(I - it\sigma^2(C+C'))^{-1} c} \prod_{j=1}^{n} (1 - it\sigma^2 \eta_j (C + C'))^{-1/2}. \end{aligned} \tag{5.A.8}$$

From (5.A.8), the cumulant generating function of f is

$$\begin{aligned} \psi(t) =& itd - \frac{\sigma^2}{2} t^2 c'(I - it\sigma^2(C + C'))^{-1} c - \frac{1}{2} \sum_{j=1}^{n} ln(1 - it\sigma^2 \eta_j (C + C')) \\ =& itd - \frac{\sigma^2}{2} t^2 c' \sum_{s=0}^{\infty} (it\sigma^2(C + C'))^s c + \frac{1}{2} \sum_{s=1}^{\infty} \frac{1}{s} (it\sigma^2)^s tr((C + C')^s), \end{aligned} \tag{5.A.9}$$

where the last displayed equality follows since

$$\begin{aligned} (I - it\sigma^2(C + C'))^{-1} &= \sum_{s=0}^{\infty} (it\sigma^2(C + C'))^s, \\ ln(1 - it\sigma^2 \eta_j (C + C')) &= -\sum_{s=1}^{\infty} \frac{(it\sigma^2 \eta_j (C + C'))^s}{s} \end{aligned}$$

and hence

$$\begin{aligned} -\frac{1}{2} \sum_{j=1}^{n} ln(1 - it\sigma^2 \eta_j) &= \frac{1}{2} \sum_{j=1}^{n} \sum_{s=1}^{\infty} \frac{1}{s} (it\sigma^2)^s \eta_j (C + C')^s \\ &= \frac{1}{2} \sum_{s=1}^{\infty} \frac{1}{s} (it\sigma^2)^s \sum_{j=1}^{n} \eta_j (C + C')^s \\ &= \frac{1}{2} \sum_{s=1}^{\infty} \frac{1}{s} (it\sigma^2)^s tr((C + C')^s). \end{aligned}$$

From (5.A.9) we can derive the *sth* cumulant of f as:

$$\kappa_1 = d + \sigma^2 trC,$$

$$\kappa_2 = \sigma^2 (c'c + \frac{\sigma^2}{2} tr((C + C')^2)),$$

$$\kappa_s = \frac{\sigma^{2s} s!}{2} \left(\frac{1}{\sigma^2} c'(C + C')^{s-2} c + \frac{tr((C + C')^s)}{s} \right), s > 2.$$

As in the proof of Theorem 2.1, let f^c be $f^c = (f - \kappa_1)/\kappa_2^{1/2}$ where $\kappa_s^c = \kappa_s/\kappa_2^{s/2}$. By the same argument outlined in detail in the proof of Theorem 2.1,

$$Pr(\hat{\lambda} \leq z) = Pr(f^c \leq -\kappa_1^c) = \Phi(-\kappa_1^c) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(-\kappa_1^c) + ... \quad (4.A.10)$$

Now,

$$\kappa_1^c = \frac{d + \sigma^2 tr(C)}{\sigma\left(c'c + \frac{\sigma^2}{2}tr((C+C')^2)\right)^{1/2}}.$$

Under Assumption 5,

$$d = -z\delta \sim -zn.$$

By standard linear algebra,

$$trC = tr(W'P(I-zW)) = tr(W'P) - ztr(W'PW) = -tr(X'W'X(X'X)^{-1}) - ztr(W'PW).$$

The first term of the RHS of the last displayed is $O(1)$ and non zero under Assumption 5, since each component of $(X'X)^{-1}$ has exact rate $1/n$ and at least one component of $X'W'X$ has exact rate n. By Lemma 2.1, $tr(W'PW) = O(n/h)$. Moreover,

$$\begin{aligned} c'c &= \beta'X'W'PWX\beta - 4z\beta'X'W'PWPWX\beta + 4z^2\beta'X'W'PWW'PWX\beta. \\ &= \delta - 4z\beta'X'W'PWPWX\beta + z^2O(n), \end{aligned}$$

since $\beta'X'W'PWW'PWX\beta = O(n)$ by Lemma 2.2. Finally,

$$tr((C+C')^2) = 2tr((W'P)^2) + 2tr(W'PW) - 8ztr((W'P)^2W) + 4z^2tr((W'PW)^2),$$

where $tr((W'P)^2 + W'PW)$, $tr((W'P)^2W)$, $tr((W'PW)^2)$ are $O(n/h)$ by Lemma 2.1.

We set

$$z = \frac{\sigma}{\delta^{1/2}}\zeta \sim \frac{1}{\sqrt{n}}.$$

Thus, when $\sqrt{n}/h = o(1)$,

$$\begin{aligned} \kappa_1^c =& \frac{-z\delta - \sigma^2 tr(X'WX(X'X)^{-1}) + O\left(\frac{\sqrt{n}}{h}\right)}{\sigma\delta^{1/2}\left(1 - \frac{4z}{\delta}\beta'X'W'PWPWX\beta + O\left(\frac{1}{h}\right)\right)^{1/2}} \\ =& \left(-\zeta - \frac{\sigma tr(X'WX(X'X)^{-1})}{\delta^{1/2}} + O\left(\frac{1}{h}\right)\right)\left(1 + \frac{2\sigma\zeta}{\delta^{3/2}}\beta'W'PWPWX\beta + O\left(\frac{1}{h}\right)\right) \\ =& -\zeta - \frac{\sigma tr(X'WX(X'X)^{-1})}{\delta^{1/2}} - \frac{2\sigma}{\delta^{3/2}}\beta'X'W'PWPWX\beta\zeta^2 + O\left(\frac{1}{h}\right). \quad (5.A.11) \end{aligned}$$

The second equality follows by a standard Taylor expansion around zero of the denominator. Under Assumption 5, both the second and the third term in (5.A.11) have exact rate $1/\sqrt{n}$. Similarly, when $h \sim \sqrt{n}$,

$$\begin{aligned}\kappa_1^c =& \frac{-z\delta - \sigma^2 tr(X'WX(X'X)^{-1}) - ztr(W'PW)}{\sigma\delta^{1/2}\left(1 - \frac{4z}{\delta}\beta'X'W'PWPWX\beta + \frac{\sigma^2}{\delta}tr((W'P)^2 + W'PW) + o\left(\frac{1}{\sqrt{n}}\right)\right)^{1/2}}\\ =& -\zeta - \frac{\sigma tr(X'WX(X'X)^{-1})}{\delta^{1/2}} - \frac{tr(W'PW)}{\delta}\zeta - \frac{2\sigma}{\delta^{3/2}}\beta'X'W'PWPWX\beta\zeta^2\\ &+\frac{\sigma^2}{2\delta}tr((W'P)^2 + W'PW)\zeta + o\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Also, by tedious but straightforward linear algebra,

$$\kappa_3^c = \frac{3\sigma^4\left(c'(C+C')c + \frac{\sigma^2}{3}tr((C+C')^3)\right)}{\sigma^3\left(c'c + \frac{\sigma^2}{2}tr((C+C')^2)\right)^{3/2}} \sim 6\sigma\frac{\beta'X'W'PWPWX\beta}{\delta^{3/2}} \sim \frac{1}{\sqrt{n}},$$

since, by Assumption 5,

$$c'(C+C')c \sim 2\beta'X'W'PWPWX\beta \sim n,$$

while by Lemma 2.1

$$tr(C+C')^3 \quad \sim \quad tr((W'P+PW)^3) = O\left(\frac{n}{h}\right).$$

For notational simplicity let

$$\vartheta_1(\zeta) = -\frac{\sigma tr(X'WX(X'X)^{-1})}{\delta^{1/2}} - \frac{2\sigma\zeta^2}{\delta^{3/2}}\beta'X'W'PWPWX\beta$$

and

$$\begin{aligned}\vartheta_2(\zeta) \quad = \quad & -\frac{\sigma tr(X'WX(X'X)^{-1})}{\delta^{1/2}} - \frac{tr(W'PW)}{\delta}\zeta - \frac{2\sigma\zeta^2}{\delta^{3/2}}\beta'X'W'PWPWX\beta\\ + \quad & \frac{\sigma^2}{2\delta}tr((W'P)^2 + W'PW)\zeta.\end{aligned}$$

Therefore, by Taylor expansion,

$$\Phi(-\kappa_1^c) = \Phi(\zeta) - \vartheta_1(\zeta)\phi(\zeta) + O\left(\frac{1}{h}\right)$$

when $\sqrt{n}/h = o(1)$, and

$$\Phi(-\kappa_1^c) = \Phi(\zeta) - \vartheta_2(\zeta)\phi(\zeta) + o\left(\frac{1}{\sqrt{n}}\right)$$

when $h \sim \sqrt{n}$.

Thus, when $\sqrt{n}/h = o(1)$, (4.A.10) becomes

$$Pr(\frac{\delta^{1/2}}{\sigma}\hat{\lambda} \leq \zeta) = \Phi(\zeta) - \vartheta_1(\zeta)\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + O\left(\frac{1}{h}\right)$$

and

$$Pr(\frac{\delta^{1/2}}{\sigma}\hat{\lambda} \leq \zeta) = \Phi(\zeta) - \vartheta_2(\zeta)\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + o\left(\frac{1}{\sqrt{n}}\right),$$

when $h \sim n$.

Remark

From the proof of Theorem 5.1 is clear that the remainder of rate $1/h$ would dominate the first Edgeworth correction if h was bounded or diverging at rate slower than $\sqrt{n}$. Similarly, it is clear that we cannot derive higher order terms, since the $O(1/h)$ remainder would dominate κ_4^c, which is expected to have rate $1/n$.

Some extra consideration about the impossibility of deriving a standard expansion when $\lambda \neq 0$ are worthwhile. When $\lambda \neq 0$ we would modify the proof of Theorem 5.1 as follows.

Since $Y = S^{-1}(\lambda)(X\beta + \epsilon)$, with $S(\lambda)$ defined according to (1.2.8), instead of (5.A.4), (5.A.5) and (5.A.6), we would have

$$C = S^{-1}(\lambda)'W'P(I - zWS^{-1}(\lambda)),$$

$$c' = \beta'X'S^{-1}(\lambda)'W'P(I - 2zWS^{-1}(\lambda)),$$

and

$$d = -z\beta'X'S^{-1}(\lambda)'W'PWS^{-1}(\lambda)X\beta = -z\delta,$$

where $\delta = \beta'X'S^{-1}(\lambda)'W'PWS^{-1}(\lambda)X\beta$.

Hence, setting $z = \sigma\zeta/\delta^{1/2}$, (5.A.11) would be

$$\begin{aligned}
\kappa_1^c &= \frac{-z\delta + \sigma^2 tr(WS^{-1}(\lambda)P) + O\left(\frac{\sqrt{n}}{h}\right)}{\sigma\delta^{1/2}\left(1 - \frac{4z}{\delta}\beta'X'S^{-1}(\lambda)'W'PWS^{-1}(\lambda)PWS^{-1}(\lambda)X\beta + O\left(\frac{1}{h}\right)\right)^{1/2}} \\
&= \left(-\zeta + \frac{\sigma tr(WS^{-1}(\lambda)P)}{\delta^{1/2}} + O\left(\frac{1}{h}\right)\right) \\
&\times \left(1 + \frac{2\sigma\zeta}{\delta^{3/2}}\beta'X'S^{-1}(\lambda)'W'PWS^{-1}(\lambda)PWS^{-1}(\lambda)X\beta + O\left(\frac{1}{h}\right)\right) \\
&= -\zeta + \frac{\sigma tr(WS^{-1}(\lambda)P)}{\delta^{1/2}} - \frac{2\sigma\zeta^2}{\delta^{3/2}}\beta'X'S^{-1}(\lambda)'W'PWS^{-1}(\lambda)PWS^{-1}(\lambda)X\beta + O\left(\frac{1}{h}\right)
\end{aligned}$$

Assuming $||S^{-1}(\lambda)||_r + ||S^{-1}(\lambda)||_c \leq K$, the second term in the last displayed expression is $O(\sqrt{n}/h)$, since $tr(WS^{-1}(\lambda)P) = O(n/h)$ by Lemma 2.1, while the third (under suitable assumptions) has rate $(1/\sqrt{n})$. It is straightforward to notice that the term of order $O(\sqrt{n}/h)$ term in κ_1^c dominates $\kappa_3^c \sim 1/\sqrt{n}$, unless h diverges at rate faster than or equal to n. However, by Assumption 2(iv), $h = O(n)$. Hence, we could derive a standard Edgeworth expansion for $\lambda \neq 0$ only when $h = n$. Clearly, this is a very limited case and does not deserve to be pursued any further.

Proof of Theorem 5.2

From (5.2.3),

$$\hat{\lambda}_{IV} - \lambda = m_{IV}^{11} Z_1'\epsilon + m_{IV}^{12} X'\epsilon = m_{IV}^{11}(Z_1'\epsilon - m_{IV,12}m_{IV,22}^{-1}X'\epsilon),$$

where

$$\begin{aligned} m_{IV}^{11} &= (m_{IV,11} - m_{IV,12}m_{IV,22}^{-1}m_{IV,21})^{-1} = (Z_1'WY - Z_1'X(X'X)^{-1}X'WY)^{-1} \\ &= (Z_1'(I - X(X'X)^{-1}X')WY)^{-1} = (Z_1'PWY)^{-1} \end{aligned}$$

and

$$m_{IV}^{12} = -m_{IV}^{11}m_{IV,12}m_{IV,22}^{-1} = -(Z_1'PWY)^{-1}Z_1'X(X'X)^{-1}.$$

Hence, for any real z, we can write

$$\begin{aligned} Pr(\hat{\lambda}_{IV} - \lambda \leq z) =& Pr(f_{IV} \leq 0), \quad \text{when} \quad \delta_{IV,1} > 0, \\ =& Pr(f_{IV} \geq 0), \quad \text{when} \quad \delta_{IV,1} < 0, \end{aligned} \tag{5.A.12}$$

where $\delta_{IV,1}$ is defined in (5.2.1) and

$$\begin{aligned} f_{IV} &= Z_1'\epsilon - m_{IV,12}m_{IV,22}^{-1}X'\epsilon - \frac{z}{m_{IV}^{11}} \\ &= Z_1'\epsilon - Z_1'X(X'X)^{-1}X'\epsilon - zZ_1'PWY \\ &= Z_1'P\epsilon - zZ_1'PWS^{-1}(\lambda)X\beta - zZ_1'PWS^{-1}(\lambda)\epsilon, \end{aligned}$$

since $Y = S^{-1}(\lambda)(X\beta + \epsilon)$.

An explanation of (5.A.12) is necessary at this stage. It is clear that such step depends on the sign of m_{IV}^{11}. Specifically,

$$Z_1'PWY = Z_1'PWS^{-1}(\lambda)X\beta + Z_1'PWS^{-1}(\lambda)\epsilon = \delta_{IV,1} + Z_1'PWS^{-1}(\lambda)\epsilon.$$

Under Assumption 7(ii), the first term has exact rate n. Regarding the second term, we have

$$E(Z_1'PWS^{-1}(\lambda)\epsilon)^2 = \sigma^2 Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1 = O(n),$$

by Lemma 2.2. Hence, by Markov inequality, $Z_1'PWS^{-1}(\lambda)\epsilon = O_p(\sqrt{n})$. We can therefore conclude that, at least for n large enough, the sign of m_{IV}^{11} is determined by $\delta_{IV,1}$.

Let

$$d_{IV} = -zZ_1'PWS^{-1}(\lambda)X\beta = -z\delta_{IV,1}$$

and

$$c_{IV}' = Z_1'P(I - zWS^{-1}(\lambda)). \tag{5.A.13}$$

Therefore,

$$f_{IV} = c_{IV}'\epsilon + d_{IV}.$$

Since f_{IV} is linear in ϵ, under Assumption 1 $f_{IV} \sim N(d_{IV}, \sigma^2 c_{IV}' c_{IV})$. Hence, the cumulants of f_{IV} are

$$\kappa_{IV,1} = d_{IV}, \quad \kappa_{IV,2} = \sigma^2 c_{IV}' c_{IV}, \quad \kappa_{IV,s} = 0, s > 2.$$

Accordingly, the centred and scaled cumulants of f_{IV} become

$$\kappa_{IV,1}^c = \frac{d_{IV}}{\sigma(c_{IV}'c_{IV})^{1/2}}, \quad \kappa_{IV,s}^c = 0, s > 2. \tag{5.A.14}$$

From (5.A.13),

$$\begin{aligned} c_{IV}'c_{IV} &= Z_1'PZ_1 - 2zZ_1'PWS^{-1}(\lambda)PZ_1 + z^2 Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1 \\ &= Z_1'PZ_1\left(1 - \frac{2zZ_1'PWS^{-1}(\lambda)PZ_1}{Z_1'PZ_1} + \frac{z^2 Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{Z_1'PZ_1}\right). \end{aligned}$$

Under Assumption 7(i),

$$Z_1'PZ_1 = Z_1'Z_1 - Z_1'X(X'X)^{-1}X'Z_1 \sim n.$$

We set

$$z = \sigma\frac{(Z_1'PZ_1)^{1/2}}{|\delta_{IV,1}|}\zeta \sim \frac{1}{\sqrt{n}},$$

since $|\delta_{IV,1}| \sim n$ under Assumption 7(ii). Hence (5.A.14) becomes

$$\begin{aligned} \kappa_{IV,1}^c &= \frac{-z\delta_{IV,1}}{\sigma(Z_1'PZ_1)^{1/2}\left(1 - 2z\frac{Z_1'PWS^{-1}(\lambda)PZ_1}{Z_1'PZ_1} + z^2\frac{Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{Z_1'PZ_1}\right)^{1/2}} \\ &= -\zeta\frac{\delta_{IV,1}}{|\delta_{IV,1}|}\{1 + z\frac{Z_1'PWS^{-1}(\lambda)PZ_1}{Z_1'PZ_1} - \frac{z^2}{2}\frac{Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{Z_1'PZ_1} \end{aligned}$$

$$+\frac{3z^2}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{(Z_1'PZ_1)^2}+O\left(\left(\frac{1}{n}\right)^{3/2}\right)\}$$
$$=-\zeta\frac{\delta_{IV,1}}{|\delta_{IV,1}|}-\frac{\sigma(Z_1'PWS^{-1}(\lambda)PZ_1)}{|\delta_{IV,1}|(Z_1'PZ_1)^{1/2}}\zeta^2\frac{\delta_{IV,1}}{|\delta_{IV,1}|}$$
$$+\frac{\sigma^2}{2}\frac{Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}^2}\zeta^3\frac{\delta_{IV,1}}{|\delta_{IV,1}|}$$
$$-\frac{3\sigma^2}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2Z_1'PZ_1}\zeta^3\frac{\delta_{IV,1}}{|\delta_{IV,1}|}+O\left(\left(\frac{1}{n}\right)^{3/2}\right).$$

Under Assumption 7(ii), the second term has exact rate $1/\sqrt{n}$, while the third and fourth ones have exact rate $1/n$.

Hence, when $\delta_{IV,1}>0$

$$Pr\left(\frac{\delta_{IV,1}}{\sigma\sqrt{Z_1'PZ_1}}(\hat{\lambda}_{IV}-\lambda)\leq\zeta\right)=Pr(f_{IV}\leq 0)$$
$$=Pr\left(\frac{f_{IV}-\kappa_{IV,1}}{\kappa_{IV,2}^{1/2}}\leq-\kappa_{IV,1}^c\right)=\Phi(-\kappa_{IV,1}^c)$$
$$=\Phi(\zeta)+\frac{\sigma(Z_1'PWS^{-1}(\lambda)PZ_1)}{\delta_{IV,1}(Z_1'PZ_1)^{1/2}}\zeta^2\phi(\zeta)$$
$$+\left(\frac{3}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2Z_1'PZ_1}-\frac{1}{2}\frac{Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}^2}\right)\sigma^2\zeta^3\phi(\zeta)$$
$$+\frac{\sigma^2}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2(Z_1'PZ_1)}\zeta^4\Phi^{(2)}(\zeta)+O\left(\left(\frac{1}{n}\right)^{3/2}\right),$$

where the fourth equality has been obtained by Taylor expansion of $\Phi(-\kappa_{IV,1}^c)$. Similarly, when $\delta_{IV,1}<0$,

$$Pr\left(\frac{-\delta_{IV,1}}{\sigma\sqrt{Z_1'PZ_1}}(\hat{\lambda}_{IV}-\lambda)\leq\zeta\right)=Pr(f_{IV}\geq 0)=Pr\left(\frac{f_{IV}-\kappa_{IV,1}}{\kappa_{IV,2}^{1/2}}\geq-\kappa_{IV,1}^c\right)$$
$$=1-\Phi(-\kappa_{IV,1}^c)$$
$$=1-\{\Phi(-\zeta)+\frac{\sigma(Z_1'PWS^{-1}(\lambda)PZ_1)}{\delta_{IV,1}(Z_1'PZ_1)^{1/2}}\zeta^2\phi(-\zeta)$$
$$-\left(\frac{3}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2Z_1'PZ_1}-\frac{1}{2}\frac{Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}^2}\right)\sigma^2\zeta^3\phi(-\zeta)$$
$$+\frac{\sigma^2}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2(Z_1'PZ_1)}\zeta^4\Phi^{(2)}(-\zeta)\}+O\left(\left(\frac{1}{n}\right)^{3/2}\right)$$
$$=\Phi(\zeta)-\frac{\sigma(Z_1'PWS^{-1}(\lambda)PZ_1)}{\delta_{IV,1}(Z_1'PZ_1)^{1/2}}\zeta^2\phi(\zeta)$$

$$+\left(\frac{3}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2 Z_1'PZ_1} - \frac{1}{2}\frac{Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}^2}\right)\sigma^2\zeta^3\phi(\zeta)$$
$$+\frac{\sigma^2}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2(Z_1'PZ_1)}\zeta^4\Phi^{(2)}(\zeta)\} + O\left(\left(\frac{1}{n}\right)^{3/2}\right),$$

since $\Phi^{(2)}(-\zeta) = -\Phi^{(2)}(\zeta)$. Collecting,

$$\begin{aligned} Pr\left(\frac{|\delta_{IV,1}|}{\sigma\sqrt{Z_1'PZ_1}}(\hat{\lambda}_{IV}-\lambda)\le\zeta\right) &= \Phi(\zeta) + \frac{\sigma(Z_1'PWS^{-1}(\lambda)PZ_1)}{|\delta_{IV,1}|(Z_1'PZ_1)^{1/2}}\zeta^2\phi(\zeta) \\ + & \left(\frac{3}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2 Z_1'PZ_1} - \frac{1}{2}\frac{Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}^2}\right)\sigma^2\zeta^3\phi(\zeta) \\ + & \frac{\sigma^2}{2}\frac{(Z_1'PWS^{-1}(\lambda)PZ_1)^2}{\delta_{IV,1}^2(Z_1'PZ_1)}\zeta^4\Phi^{(2)}(\zeta)\} + O\left(\left(\frac{1}{n}\right)^{3/2}\right). \end{aligned}$$

The claim in Theorem 5.2 follows by setting

$$\delta_{IV,2} = Z_1'PWS^{-1}(\lambda)PZ_1$$

and

$$\delta_{IV,3} = Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1.$$

Proof of Theorem 5.3

Similarly to (5.A.12), from (5.3.2) under H_0, we can write

$$\begin{aligned} Pr(R'\hat{\beta}_{IV}\le z) &= Pr(f_{IV,\beta}\le 0) \quad \text{if} \quad \delta_{IV,1}>0 \\ &= Pr(f_{IV,\beta}\ge 0) \quad \text{if} \quad \delta_{IV,1}<0, \end{aligned}$$

where

$$\begin{aligned} f_{IV,\beta} &= -R'(X'X)^{-1}X'WYZ_1'P\epsilon + R'(X'X)^{-1}X'\epsilon Z_1'PWY - zZ_1'PWY \\ &= -\epsilon'PZ_1R'(X'X)^{-1}X'WY + \epsilon'X(X'X)^{-1}RZ_1'PWY - zZ_1'PWY. \end{aligned}$$

Since $Y = S^{-1}(\lambda)(X\beta+\epsilon)$,

$$\begin{aligned} f_{IV,\beta} &= -\epsilon'PZ_1R'(X'X)^{-1}X'WS^{-1}(\lambda)X\beta - \epsilon'PZ_1R'(X'X)^{-1}X'WS^{-1}(\lambda)\epsilon \\ &+ \epsilon'X(X'X)^{-1}RZ_1'PWS^{-1}(\lambda)X\beta + \epsilon'X(X'X)^{-1}RZ_1'PWS^{-1}(\lambda)\epsilon \\ &- zZ_1'PWS^{-1}(\lambda)X\beta - zZ_1'PWS^{-1}(\lambda)\epsilon \end{aligned}$$

$$= \frac{1}{2}\epsilon'(C_{IV,\beta} + C'_{IV,\beta})\epsilon + c'_{IV,\beta}\epsilon + d_{IV,\beta},$$

where

$$C_{IV,\beta} = -PZ_1R'(X'X)^{-1}X'WS^{-1}(\lambda) + X(X'X)^{-1}RZ_1'PWS^{-1}(\lambda) = (Q-Q')WS^{-1}(\lambda), \tag{5.A.15}$$

$$\begin{aligned} c'_{IV,\beta} &= \beta'X'S^{-1}(\lambda)'W'(-X(X'X)^{-1}RZ_1P + PZ_1R'(X'X)^{-1}X') - zZ_1'PWS^{-1}(\lambda) \\ &= G' - zZ_1'PWS^{-1}(\lambda), \end{aligned}$$

and

$$d_{IV,\beta} = -zZ_1'PWS^{-1}(\lambda)X\beta = -z\delta_{IV,1}. \tag{5.A.16}$$

Following the same procedure used in the proof of Theorem 5.1, the cumulants of $f_{IV,\beta}$ are

$$\kappa_{IV,\beta,1} = d_{IV,\beta} + \sigma^2 trC_{IV,\beta},$$

$$\kappa_{IV,\beta,2} = \sigma^2(c'_{IV,\beta}c_{IV,\beta} + \frac{\sigma^2}{2}tr((C_{IV,\beta} + C'_{IV,\beta})^2))$$

and

$$\kappa_{IV,\beta,s} = \frac{\sigma^{2s}s!}{2}\{\frac{1}{\sigma^2}c'_{IV,\beta}(C_{IV,\beta} + C'_{IV,\beta})^{s-2}c_{IV,\beta} + \frac{tr((C_{IV,\beta} + C'_{IV,\beta})^s)}{s}\}, s > 2.$$

From (5.A.15) and (5.A.16),

$$\kappa_{IV,\beta,1} = -z\delta_{IV,1} + \sigma^2(tr(QWS^{-1}(\lambda) - Q'WS^{-1}(\lambda))).$$

Under Assumption 8(ii), $\delta_{IV,1} \sim n$. Before proceeding, we notice that

$$tr(QWS^{-1}(\lambda) - Q'WS^{-1}(\lambda)) = 0$$

when $(WS^{-1}(\lambda))' = WS^{-1}(\lambda)$. The latter may hold, for instance, when W is symmetric. Indeed under Assumption 2(i), (2.4.2) holds and hence symmetry of W implies symmetry of $S^{-1}(\lambda)$. Therefore,

$$(WS^{-1}(\lambda))' = S^{-1}(\lambda)'W' = S^{-1}(\lambda)W = WS^{-1}(\lambda),$$

since W and $S^{-1}(\lambda)$ commute when (2.4.2) holds true. In general, when $WS^{-1}(\lambda)$ is not symmetric,

$$|(QWS^{-1}(\lambda) - Q'WS^{-1}(\lambda))_{ii}| \le |(QWS^{-1}(\lambda))_{ii}| + |(Q'WS^{-1}(\lambda))_{ii}|.$$

We can show that under Assumptions 7(i) and 8 both terms at the RHS of the last displayed expression are $O(1/n)$, uniformly in i. We show the result in detail for

$|(QWS^{-1}(\lambda))_{ii}|$, a similar argument with very minor modification holds for the second term.

We have

$$\begin{aligned} |(QWS^{-1}(\lambda))_{ii}| &= |(X(X'X)^{-1}RZ_1'PWS^{-1}(\lambda))_{ii}| = |x_i'(X'X)^{-1}RZ_1'a_i| \\ &\leq |x_i'(X'X)^{-1}R||Z_1'a_i| = O\left(\frac{1}{n}\right), \end{aligned}$$

where x_i' and a_i indicate the $i-$th row of X and the $i-$th column of $A = PWS^{-1}(\lambda)$, respectively. Indeed, under Assumptions 7(i), and 8(i),

$$\frac{1}{n}|x_i'\left(\frac{X'X}{n}\right)^{-1}R| \leq \frac{1}{n}||x_i'||||\left(\frac{X'X}{n}\right)^{-1}||||R|| = O\left(\frac{1}{n}\right), \tag{5.A.17}$$

uniformly in i (the latter argument is very similar to one in the proof of Lemma 4.1). Also, $||A||_r + ||A||_c \leq K$, since P, W and $S^{-1}(\lambda)$ have the same property. We denote by z_{1i} the ith component of Z_1 and a_{ti} the $t-i$th element of A. Uniformly in i,

$$|Z_1'a_i| = |\sum_{t=1}^{n} z_{1t}a_{ti}| \leq \sum_{t=1}^{n}|z_{1t}||a_{ti}| \leq \max_t|z_{1t}|\sum_{t=1}^{n}|a_{ti}| = O(1). \tag{5.A.18}$$

Therefore,

$$|tr(QWS^{-1}(\lambda) - Q'WS^{-1}(\lambda))| \leq n\max_i|(QWS^{-1}(\lambda))_{ii}| + n\max_i|Q'WS^{-1}(\lambda))_{ii}| = O(1).$$

Moreover,

$$c_{IV,\beta}'c_{IV,\beta} = G'G - 2zG'S^{-1}(\lambda)'W'PZ_1 + z^2Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1.$$

Under Assumption 8(ii),

$$G'G \sim n,$$

$$G'S^{-1}(\lambda)'W'PZ_1 \sim n,$$

while

$$Z_1'PWS^{-1}(\lambda)S^{-1}(\lambda)'W'PZ_1 = O(n)$$

by Lemma 2.2.

Similarly,

$$\begin{aligned} & \frac{1}{2}\left(tr(C_{IV,\beta} + C_{IV,\beta}')^2\right) = tr(C_{IV,\beta}^2) + tr(C_{IV,\beta}C_{IV,\beta}') \\ = \ & tr((Q-Q')^2(WS^{-1}(\lambda))^2) + tr((Q-Q')WS^{-1}(\lambda)S^{-1}(\lambda)'W'(Q'-Q)) \\ = \ & tr(Q^2(WS^{-1}(\lambda))^2) + tr((Q')^2(WS^{-1}(\lambda))^2) - 2tr(QQ'(WS^{-1}(\lambda))^2) \\ + \ & tr((Q-Q')WS^{-1}(\lambda)S^{-1}(\lambda)'W'(Q'-Q)). \end{aligned}$$

It can be shown that each of the terms appearing in the last displayed expression is $O(1)$ under Assumptions 6, 7(i) and 8. We only show

$$tr((Q-Q')^2(WS^{-1}(\lambda))^2)=O(1),$$

but similar arguments apply also to the remaining terms. We have,

$$\begin{aligned}|tr(Q^2(WS^{-1}(\lambda))^2)| &= |tr(X(X'X)^{-1}RZ_1'PX(X'X)^{-1}RZ_1'P(WS^{-1}(\lambda))^2)| \\ &= |Z_1'P(WS^{-1}(\lambda))^2X(X'X)^{-1}RZ_1'PX(X'X)^{-1}R| \\ &\leq |Z_1'P(WS^{-1}(\lambda))^2X(X'X)^{-1}R||Z_1'PX(X'X)^{-1}R| \\ &= |Z_1'BX(X'X)^{-1}R||Z_1'PX(X'X)^{-1}R|\end{aligned}$$

where $B=P(WS^{-1}(\lambda))^2$ satisfies

$$||B||_r+||B||_c\leq K.$$

Now, denoting b_i the $i-$th column of B,

$$|Z_1'BX(X'X)^{-1}R|=|\sum_{i=1}^{n}Z_1'b_ix_i'(X'X)^{-1}R|\leq\sum_{i=1}^{n}|Z_1'b_i||x_i'(X'X)^{-1}R|=O(1),$$

since $|Z_1'b_i|=O(1)$, uniformly in i, by the same argument given in (5.A.18) and $|x_i'(X'X)^{-1}R|=O(1/n)$, uniformly in i, by (5.A.17). Similarly, $|Z_1'PX(X'X)^{-1}R|=O(1)$.

We choose

$$z=\frac{\sigma(G'G)^{1/2}}{|\delta_{IV,1}|}\zeta.$$

Hence, the first centred cumulant of $f_{IV,\beta}$, denoted $\kappa^c_{IV,\beta,1}$ becomes

$$\begin{aligned}\kappa^c_{IV,\beta,1} &= \frac{-z\delta_{IV,1}+\sigma^2tr((Q-Q')WS^{-1}(\lambda))}{\sigma(G'G)^{1/2}\left(1-\frac{2zG'S^{-1}(\lambda)'W'PZ_1}{G'G}+O\left(\frac{1}{n}\right)\right)^{1/2}} \\ &= \left(-\zeta\frac{\delta_{IV,1}}{|\delta_{IV,1}|}+\frac{\sigma tr((Q-Q')WS^{-1}(\lambda))}{(G'G)^{1/2}}\right)\left(1+\frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{|\delta_{IV,1}|(G'G)^{1/2}}\zeta+O\left(\frac{1}{n}\right)\right) \\ &= -\zeta\frac{\delta_{IV,1}}{|\delta_{IV,1}|}+\frac{\sigma tr((Q-Q')WS^{-1}(\lambda))}{(G'G)^{1/2}}-\frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{|\delta_{IV,1}|(G'G)^{1/2}}\zeta^2\frac{\delta_{IV,1}}{|\delta_{IV,1}|}+O\left(\frac{1}{n}\right) \\ &= -\zeta\frac{\delta_{IV,1}}{|\delta_{IV,1}|}+\frac{\sigma tr((Q-Q')WS^{-1}(\lambda))}{(G'G)^{1/2}}-\frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}(G'G)^{1/2}}\zeta^2+O\left(\frac{1}{n}\right).\end{aligned}$$

Under Assumption 8(ii), when $WS^{-1}(\lambda)$ is not symmetric (e.g. when (2.4.2) holds and W is not symmetric) both the second and third terms have exact rate $(1/\sqrt{n})$. As shown above, when $WS^{-1}(\lambda)$ is symmetric the second term vanishes.

Moreover,

$$\begin{aligned}\kappa^c_{IV,\beta,3} &= \frac{\kappa_{IV,3,\beta}}{\kappa^{3/2}_{IV,2,\beta}} \sim \frac{3\sigma^4 c'_{IV,\beta}(C_{IV,\beta}+C'_{IV,\beta})c_{IV,\beta}}{\sigma^2(G'G)^{3/2}} \\ &\sim \frac{6\sigma^2 G'(Q-Q')WS^{-1}(\lambda)G}{\sigma^2(G'G)^{3/2}} \sim \frac{1}{\sqrt{n}}\end{aligned}$$

under Assumption 8(ii), while each term of $tr((C_{IV,\beta}+C'_{IV,\beta})^3)$ is $O(1)$ (the latter property can be shown with a slight modification of the argument we used to show that each element of $tr((C_{IV,\beta}+C'_{IV,\beta})^2)$ is $O(1)$).

Finally, following the same procedure we described in the proof of Theorem 5.1, when $\delta_{IV,1}>0$,

$$\begin{aligned}Pr\left(\frac{\delta_{IV,1}}{\sigma(G'G)^{1/2}}R'\hat{\beta}\le\zeta\right) &= \Phi(\zeta) \\ &- \left(\frac{\sigma tr(Q-Q')WS^{-1}(\lambda)}{(G'G)^{1/2}} - \frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}(G'G)^{1/2}}\zeta^2\right)\phi(\zeta) \\ &- \frac{\kappa^c_{IV,\beta,3}}{3!}H_2(\zeta)\phi(\zeta)+O\left(\frac{1}{n}\right).\end{aligned}$$

When $\delta_{IV,1}<0$,

$$\begin{aligned}&Pr\left(\frac{-\delta_{IV,1}}{\sigma(G'G)^{1/2}}R'\hat{\beta}\le\zeta\right) = Pr(f_{IV,\beta}\ge 0) = Pr\left(\frac{f_{IV,\beta}-\kappa_{IV,1,\beta}}{\sqrt{\kappa_{IV,2,\beta}}}\ge -\kappa^c_{IV,1,\beta}\right) \\ =\; & 1-\{\Phi(-\zeta)-\left(\frac{\sigma tr(Q-Q')WS^{-1}(\lambda)}{(G'G)^{1/2}} - \frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}(G'G)^{1/2}}\zeta^2\right)\phi(-\zeta) \\ -\; & \frac{\kappa^c_{IV,\beta,3}}{3!}H_2(-\zeta)\phi(-\zeta)\}+O\left(\frac{1}{n}\right) \\ =\; & \Phi(\zeta)+\left(\frac{\sigma tr(Q-Q')WS^{-1}(\lambda)}{(G'G)^{1/2}} - \frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}(G'G)^{1/2}}\zeta^2\right)\phi(\zeta) \\ +\; & \frac{\kappa^c_{IV,\beta,3}}{3!}H_2(\zeta)\phi(\zeta)+O\left(\frac{1}{n}\right).\end{aligned}$$

Collecting terms,

$$\begin{aligned}Pr\left(\frac{|\delta_{IV,1}|}{\sigma(G'G)^{1/2}}R'\hat{\beta}\le\zeta\right) &= \Phi(\zeta) \\ &- \frac{\delta_{IV,1}}{|\delta_{IV,1}|}\left(\frac{\sigma tr(Q-Q')WS^{-1}(\lambda)}{(G'G)^{1/2}} - \frac{\sigma G'S^{-1}(\lambda)'W'PZ_1}{\delta_{IV,1}(G'G)^{1/2}}\zeta^2\right)\phi(\zeta) \\ &- \frac{\delta_{IV,1}}{|\delta_{IV,1}|}\frac{\kappa^c_{IV,\beta,3}}{3!}H_2(\zeta)\phi(\zeta)+O\left(\frac{1}{n}\right).\end{aligned}$$

The claim in Theorem 5.3 follows by setting $\kappa_{IV,\beta}=\kappa^c_{IV,\beta,3}$.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers.

Anselin, L. (2001). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*, **97**, 113-39.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, **89**, 3-25.

Ansley, C.F., R. Kohn and T.S. Shively (1992). Computing $p-$ values for the generalized Durbin-Watson and other invariant test statistics. *Journal of Econometrics*, **54**, 277-300.

Anselin, L. and R. Florax (1995). Small sample properties of tests for spatial dependence in regression models: some further results. In L. Anselin and R. Florax (Eds.) *New Directions in Spatial Econometrics.* Springer-Verlag, Berlin, 21-74.

Anselin, L. and S. Rey (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis*, **23**, 112-31.

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundation and Applications to Regional Analysis.* Springer-Verlag, Berlin.

Bao, Y and A. Ullah (2007). Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics*, **137**, 396-413.

Barndorff-Nielsen, O.E. and D.R. Cox (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

Bhattacharya, R.N. and J.K. Ghosh (1978). On the validity of the formal Edgeworth expansion. *Annals of Statistics*, **6**, 434-451.

Bhattacharya, R.N. and R.R. Rao (1976). *Normal Approximation and Asymptotic Expansions.* John Wiley & Sons.

Buonanno, P., D. Montolio and P. Vanin (2009). Does social capital reduce crime? *Journal of Law and Economics* **52**, 145-70.

Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society, Series B*, **42**, 107-8.

Case, A.C. (1991). Spatial Patterns in Household Demand. *Econometrica*, **59**, 953-65.

Cliff, A. and J.K. Ord, (1968). The problem of spatial autocorrelation. *Joint Discussion Paper, University of Bristol: Department of Economics*, **26**, *Department of Geography, Series A*, **15**.

Cliff, A. and J.K. Ord, (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, **4**, 267-84.

Cliff, A. and J.K. Ord, (1975). *Spatial Autocorrelation.* Pion, London.

Conley, T.G. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics*, **92**, 1-45.

Conley, T.G. and B. Dupor (2003). A spatial analysis of sectoral complementarity. *Journal of Political Economy*, **111**, 311-52.

Conley, T.G. and E. Ligon (2002). Economic distance and cross-country spillovers. *Journal of Economic Growth*, **7**, 157-87.

Conley, T.G. and F. Molinari (2007). Spatial correlation robust inference with errors in location or distances. *Journal of Econometrics*, **140**, 76-96.

Cornish, E.A. and R.A. Fisher (1937). Moments and cumulants in the specification of distributions. *International Statistical Reviews*, **5**, 307-22.

Cornish, E.A. and R.A. Fisher (1960). The percentile points of distributions having known cumulants. *Technometrics*, **2**, 209-26.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton, NJ.

Cressie, N. (1993). *Statistics for Spatial Data.* Wiley, New York.

Dahlhaus, R. and H. Künsch (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, **74**, 877-82.

Das, D., H.H. Kelejian, and I. Prucha (2003). Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science*, **82**, 1-26.

Davies, R.B. (1973). Numerical inversion of a characteristic function. *Biometrika*, **60**, 415-417.

Davies, R.B. (1980). The distribution of a linear combination of χ^2 random variables. *Applied Statistics*, **29**, 323-333.

DiCiccio, T.J. and J.P. Romano (1995). On bootstrap procedures for second-order accurate confidence limits in parametric models. *Statistica Sinica*, **5**, 141-60.

DiCiccio, T.J. and B. Efron (1996). Bootstrap confidence intervals. *Statistical Science*, **11**, 189-228.

Edgeworth, F.Y. (1896). The asymmetrical probability curve. *Philosophical Magazine, 5th Series*, **41**, 90-9.

Edgeworth, F.Y. (1905). The law of error. *Proceedings of the Cambridge Philosophical Society*, **20**, 36-65.

Efron, B. and R.J. Tibshirani (1993). *An Introduction to the Bootstrap.* London: Chapman and Hall.

Egger, P., M. Larch, M. Pfaffermayr and J. Walde (2009). Small sample properties of maximum likelihood versus generalized method of moments based tests for spatially autocorrelated errors. *Regional Science and Urban Economics*, **39**, 670-78.

Ghazal, G.A. (1996). Recurrence formula for expectation of products of quadratic forms. *Statistics and Probability Letters*, **27**, 101-9.

Guyon, X. (1982). Parametric estimation for a stationary process on a $d-$dimensional lattice. *Biometrika*,**69**, 95-106.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag.

Heijmans, R. (1999). When does the expectation of a ratio equal the ratio of expectations? *Statistical Papers*, **40**, 107-15.

Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika*, **72**, 653-659.

Horn, R.A. and C.R. Johnson (1985). *Matrix Analysis.* New York: Cambridge University Press.

Kakizawa, Y. (1999). Valid Edgeworth expansions of some estimators and bootstrap confidence intervals in first-order autoregression. *Journal of Time Series Analysis*, **20**, 343-59.

Kelejian, H.H. and I.R. Prucha (1998). A generalized spatial two-stages least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, **17**,99-121.

Kelejian, H.H. and I.R. Prucha (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, **104**, 219-57.

Kelejian, H.H, I.R. Prucha and Y. Yuzefovich (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances:

large and small sample results. In J.P. Lesage and R.K.Pace (Eds.) *Advances in Econometrics: Spatial and Spatiotemporal Econometrics* . Elsevier, New York, 163-98.

Kelejian, H.H. and D.P. Robinson (1992). Spatial autocorrelation - a new computationally simple test with an application to the per capita county police expenditures. *Regional Science and Urban Economics*, **22**, 317-31.

Kelejian, H.H. and D.P. Robinson (1998). A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results. *Regional Science and Urban Economics*, **28**, 389-417.

Konishi, S., N. Niki and A.K. Gupta (1988). Asymptotic expansions for the distribution of quadratic forms in normal variables. *Annals of the Institute of Statistical Mathematics*, **40**, 279-96.

Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 266-83.

Lee, L.F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory*, **18**, 252-277.

Lee, L.F. (2003). Best spatial two-stages least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, **22**, 307-335.

Lee, L.F. (2004). Asymptotic distribution of quasi-maximum likelihood estimates for spatial autoregressive models. *Econometrica*, **72**, 1899-1925.

Lieberman, O. (1994). A Laplace approximation to the moments of a ratio of quadratic forms. *Biometrika* **81**, 681-90.

Magee, L. (1989). An Edgeworth test size correction for the linar model with AR(1) errors. *Econometrica*, **57**, 661-74.

Moran, P.A.P. (1950). A test for the serial dependence of residuals. *Biometrika* **37**, 178-81.

Ochi, Y. (1983). Asymptotic expansions for the distribution of an estimator in the first order autoregressive process. *Journal of Time Series Analysis*, **4**, 57-67.

Pace, R.K. and R. Berry (1997). Quick computation of spatial autoregressive estimators. *Geographical Analysis*, **29**, 232-46.

Paparoditis, E. and D.N. Politis (2005). Bootstrap hypothesis testing in regression models. *Statistics & Probability Letters*, **74**, 356-365.

Phillips, P.C.B. and J.Y.Park (1988). On the formulation of the Wald tests of nonlinear restriction.*Econometrica*, **56**, 1065-1083.

Pinkse, J. (1999). Asymptotics of the Moran test and a test for spatial correlation in probit models. *UBC, mimeo.*

Pinkse, J. (2004). Moran - flavoured tests with nuisance parameters: examples. In Anselin, L., Florax, R.G.M., Rey, S. (eds.). *Advances in Spatial Econometrics: Methodology, Tools and Applications.* Springer-Verlag, Berlin, 66-77.

Pitman, E.G.J. (1937). Significant tests which may be applied to samples from any population. Supplement to *Journal of the Royal Statistical Society*, **4**, 119-30.

Robinson, P.M. (1977). Estimation of a time series model from unequally spaced data. *Stochastic Processes and their Applications*, **6**, 9-24.

Robinson, P.M. (2008a). Developments in the analysis of spatial data. *Journal of the Japan Statistical Society* (issue in honour of H. Akaike) **38**, 87-96.

Robinson, P.M. (2008b). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics*, **147**, 5-16.

Robinson, P.M. and J. Vidal Sanz (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis*, **97**, 1090-120.

Rothenberg, T.J. (1984). Approximating the distribution of econometric estimators and test statistics. In *Handbook of Econometrics*, **2**. Elsevier Science Publisher.

Sargan, J.D. (1976). Econometric estimators and the Edgeworth approximation. *Econometrica*, **44**, 421-48.

Sen, A. (1976). Large sample-size distribution of statistics used in testing for spatial correlation. *Geographical Analysis*, **9**, 175-84.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics*, **9**, 1187-95.

Stanca, L. (2009). The geography of economics and happiness: spatial patterns in the effects of economic conditions on well being. Forthcoming, *Social Indicators Research.*

Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Annals of Statistics* **11**, 157-69.

Taniguchi, M. (1986). Third order asymptotic properties of Maximum Likelihood Estimators for Gaussian ARMA processes. *Journal of Multivariate Analysis* **18**, 1-31.

Taniguchi, M. (1988). Asymptotic expansion of the distributions of some test statistics for Gaussian ARMA processes. *Journal of Multivariate Analysis* **27**, 494-511.

Taniguchi, M. (1991a). Third order asymptotic properties of a class of test statistics under a local alternative. *Journal of Multivariate Analysis*, **37**, 223-38.

Taniguchi, M. (1991b). *Higher Order Asymptotic Theory for Time Series Analysis*. Springer-Verlag, Berlin.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, **41**, 434-49.

Yanagihara, H. and K. Yuan (2005). Four improved statistics for contrasting means by correcting skewness and kurtosis. *British Journal of Mathematical and Statistical Psychology*, **58**, 209-37.