Alternative estimation methods and specification tests for moment condition models

Joaquim J. S. Ramalho

A thesis submitted to the University of Bristol in accordance with the requirements of the degree of Ph.D. in the Faculty of Social Sciences, Department of Economics.

February 2002

Word count: 37137

Abstract

The study of the generalized method of moments (GMM) and alternative estimation methods for models which are defined solely in terms of a set of moment conditions constitutes a recent and increasingly popular research topic in econometrics. In this thesis we focus on the analysis of GMM and generalized empirical likelihood (GEL) estimators and related statistics, providing an up-to-date survey of the existing literature and performing three major contributions to this subject.

Our first major investigation concerns the examination of the small sample bias of ten alternative estimators for moment condition models, which may be divided into two classes. The first includes the first-order asymptotically equivalent GMM, continuous-updating GMM, empirical likelihood and empirical information estimators. The second contains six GMM bootstrap estimators, three of which are developed in this thesis. Two extensive Monte Carlo studies reveal that one of the new bootstrap techniques produces the estimators with less bias in most cases.

Second, we derive several Pearson-type statistics suitable for testing overidentifying moment conditions and parametric restrictions. In a Monte Carlo study concerning the first class of tests, we find that, in small samples, the size behaviour of one of the new statistics is superior to that of both alternative tests based on their asymptotic distributions and bootstrap forms of the popular Hansen's (1982) J test.

The proposal of a number of new non-nested hypothesis tests that integrate and complement the work of other authors constitute our last major contribution. We derive generalized statistics that include most of the existing tests as particular cases and develop GEL parametric and moment encompassing tests that enlarge substantially the number of tests available to the practitioner to assess non-nested moment condition models. Simulation experiments indicate that GEL-based encompassing tests using a robust estimator for the variance matrix of the moment indicators are particularly efficacious.

Acknowledgements

In the preparation of this thesis I received help from numerous people. Above all, I would like to express my gratitude to my supervisors, Professor Richard J. Smith and Professor Andrew D. Chesher, for their useful advice, constructive criticism, patience and ability to motivate.

My thanks are also mainly due to Professor João Santos Silva, the principal instigator for starting my research in econometrics at the University of Bristol, to Montezuma Dumangane, for very helpful discussions, and to my wife and colleague Esmeralda Ramalho, for her useful suggestions and continuous support.

Financial support from the Universidade de Évora and the Fundação para a Ciência e a Tecnologia, for three-year funding through their PRAXIS XXI scholarship program, is also gratefully acknowledged.

Last, but not least, my most sincere thanks to my family, for their constant support and encouragement.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of this dissertation has been submitted for any other degree.

Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol.

The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

Joaquim J. S. Ramalho

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

For many years, standard econometric modelling practice was based on strong assumptions concerning the underlying data generation process. Such assumptions were known to be often highly unrealistic but nevertheless they were made since they allowed the construction of estimators with optimal theoretical properties. The most important example of this perspective was the maximum likelihood estimation method, whose utilization requires the complete specification of the model to be estimated, including the probability distribution of (at least) the variable of interest. However, since the seventies important changes have occurred in econometric practice. Nowadays, making strong distributional assumptions when *a priori* knowledge is insufficient to support them is no longer acceptable and econometricians in general consider that working under assumed likelihood functions is too restrictive. Thus, during the last two decades, estimation and hypothesis testing techniques that do not require the specification of a likelihood have been developed, a wide set of semiparametric and nonparametric tools being now available to the practitioner.

One of the new popular model formalizations requires only the specification of a set of moment conditions, or estimating equations, which the model to be estimated should satisfy. The analyst need only make mild assumptions on the existence of certain zero-valued moments, which are defined in terms of the observable variables and the parameters of interest. The utilization of moment-based estimation dates back over 100 years to Karl Pearson's method of moments but only in the last twenty years has it received renewed interest by econometricians, instigated by Hansen's (1982) seminal paper on the generalized method of moments (GMM). Indeed, a complete methodology for estimation and hypothesis testing applicable to a large number of models was provided in that paper. Therefore, it is no surprise that utilization of GMM in empirical work has become common in the econometric literature during the last decade or so.

Despite its popularity, GMM suffers from some important drawbacks, the principal of them being its finite sample behaviour. In fact, it has been recognized for several years that the first-order asymptotic distribution of the GMM estimator provides a poor approximation to its small sample distribution. There is increasing Monte Carlo evidence indicating that in finite samples GMM estimators may be badly biased and the associated tests may have actual sizes substantially different from the nominal ones. For this reason, several authors recently proposed some alternative estimation methods to deal with these moment condition models. Like GMM, these new estimation techniques possess the same asymptotic first-order properties. Unlike GMM, little is known about their small sample behaviour and some issues are still unexplored.

The main objective of this thesis is the study of the principal estimation methods for cross-sectional models defined solely in terms of moment conditions. Besides GMM, we focus on the analysis of Hansen, Heaton and Yaron's (1996) continuous-updating GMM (CU-GMM) and, especially, of Qin and Lawless (1994) and Imbens's (1997) empirical likelihood (EL) and Kitamura and Stutzer (1997) and Imbens, Spady and Johnson's (1998) exponential tilting or empirical information (EI) methods. The last two methods possess many common features, being particular cases of both the minimum discrepancy (MD) methods developed by Corcoran (1998) and the generalized empirical likelihood (GEL) method proposed by Smith (1997). Hence, most of the theoretical discussions concerning EL and EI estimators will be realized *via* the general framework provided by the GEL method.

The particular attention devoted to the study of GEL estimators throughout this thesis is justified by the limited knowledge about their finite sample properties and by their many attractive theoretical features relative to GMM, which more than compensate for the computational burden required in their estimation. To begin with, as likelihood-like methods, they allow the utilization of classical-type tests to evaluate various hypotheses concerning the specification of a particular model, including overidentifying moment conditions. Moreover, Newey and Smith (2000) demonstrated that GEL estimators have one less source of asymptotic bias than the GMM estimator. Finally, all moment conditions are imposed on the sample by appropriately reweighting the data, rather than only some linear combinations of them. This is achieved by employing a more efficient estimator of the distribution of the data than the empirical distribution implicitly used in the GMM case. This estimator, the so-called GEL distribution function, and corresponding GEL implied probabilities, has many different applications. In fact, all the major contributions of this thesis involve particular uses of the GEL implied probabilities. Namely, we show how to employ the GEL probabilities in the construction of three new bootstrap techniques applicable in the GMM framework, several Pearson-type statistics for assessing overidentifying moment conditions and parametric restrictions and a number of new non-nested hypothesis tests.

Throughout this thesis there is a continual comparison between the finite sample properties of GMM and GEL statistics. Our aim is the production of considerable evidence showing that, as expected, the latter method leads to more reliable estimation and inference. However, GEL estimation is not the only way to improve the small sample performance of GMM estimation. Indeed, econometricians can opt for utilizing techniques such as the bootstrap in order to obtain more accurate approximations to the finite-sample distribution of GMM estimators and related statistics. Thus, the merits of the employment of bootstrap methods in the GMM context are also discussed, namely when applied to obtain bias-corrected GMM estimators and to approximate the small sample distribution of Hansen's (1982) J test of overidentifying moment conditions.

In the next sub-section we outline the structure of this thesis.

1.2 Structure of the thesis

This dissertation is organized in six chapters. Below, we give a detailed description of chapters 2 to 6.

Chapter 2 provides an up-to-date survey of the principal extant literature on estimation methods for moment condition models. We start with a detailed description of the major characteristics of GMM estimation. Alternative GMM estimators are analyzed, a special emphasis being given to Hansen's (1982) two-step efficient GMM estimator, since this is the GMM estimator that is more often used both throughout this thesis and in applied work. The main specification tests for efficient GMM estimators are reviewed, namely tests of overidentifying moment conditions, tests for additional moment conditions and tests of parametric restrictions. In an independent section, by extending Smith's (1987) work for maximum likelihood estimators to the GMM framework, we discuss an original way of deriving most GMM tests of parametric restrictions and overidentifying moment conditions. In particular, we develop a test statistic generating equation from which, by evaluation at appropriate estimators, several tests may be obtained as special cases. This review of GMM concludes with a discussion of its main limitations in order to motivate the use of and the search for alternative procedures with better finite sample properties.

The second part of chapter 2 is thus dedicated to the study of alternative estimation methods for moment condition models that share the same first-order asymptotic properties as GMM. First, we briefly review the CU-GMM estimator. Then, we focus on the utilization of GEL methods, giving special attention, as previously mentioned, to EI and EL estimation techniques. Computational aspects concerning the calculation of these estimators are discussed, since this seems to be their most important drawback. Similarly to GMM, specification tests applicable in this context are presented.

In chapter 3 we compare the finite sample bias of GMM, CU-GMM, EI and EL estimators through two Monte Carlo simulation studies.[1] We examine also the ability of bootstrap methods to improve the small sample properties of the two-step GMM estimator. We consider three bootstrap techniques already applied in the moment condition framework and propose three new ones that employ the GEL implied probabilities to resample the data. Two different settings for which there is previous evidence of the poor performance of the efficient two-step GMM estimator are simulated using the package S-Plus, which is utilized in all Monte Carlo studies realized in this thesis.

In chapter 4, Pearson-type statistics suitable for testing overidentifying moment conditions and parametric restrictions are developed. These new test statistics are based on the comparison of two consistent estimators, under the corresponding null hypothesis, of the unknown distribution of the data. For the former class of tests those estimators are the empirical and the GEL distribution functions, while in the latter case two GEL distributions estimated under different assumptions are contrasted. Through a Monte Carlo simulation study based on two of the settings considered by Imbens, Spady and Johnson (1998), we examine the finite sample properties of several tests of overidentifying moment conditions, including bootstrap forms of Hansen's (1982) J test.

Chapter 5 deals with the issue of testing non-nested hypothesis in the moment condition framework.[2] To the best of our knowledge, there are relatively few papers addressing this subject. Indeed, only Singleton (1985), Ghysels and Hall (1990b) and Smith (1992), for the GMM case, and Smith (1997), for GEL estimators, have

[1] A paper containing the main findings of this chapter was presented at the 56th European Meeting of the Econometric Society, Lausanne, Switzerland, 25-29/08/2001.

[2] The main findings of this chapter were included in a joint paper with Richard J. Smith, which is forthcoming in the Journal of Econometrics and was presented by the author of this dissertation at the 8th World Congress of the Econometric Society, Seattle, U.S.A., 11-16/08/2000.

approached this question. In this chapter we derive a number of new tests that complement and integrate those works. With our proposals the number of tests available for the researcher assessing different non-nested theories is significantly increased. A Monte Carlo study involving instrumental variable models ends the chapter.

Chapter 6 concludes this thesis by summarizing our major findings and suggesting potential avenues for future research.

Chapter 2

Estimation methods for moment condition models

2.1 Introduction

The main goal of this chapter is to provide a general framework for dealing with moment condition models. The most relevant econometric literature concerning this kind of models is surveyed, the major characteristics of the main estimation methods and corresponding specification tests applicable in this context being presented. In the following we use the expression 'empirical-based' (EB) to designate the set of all techniques that are special cases of the MD and GEL estimation methods. When considering GMM specification tests, we discuss an original procedure for deriving most GMM tests of parametric restrictions and overidentifying moment conditions.

This chapter is organized as follows. Section 2.2 introduces some definitions and notation. Section 2.3 examines GMM estimation and inference. Section 2.4 discusses the CU-GMM estimator. Section 2.5 is dedicated to the study of EB estimation methods.

2.2 Definitions and notation

The notation and assumptions introduced in this section are used throughout this dissertation. Let y_i, $i = 1, ..., n$, be independent and identically distributed (i.i.d.) observations on a data vector y. Consider $g(y, \theta)$, an s-dimensioned vector of moment indicators known up to the k-vector of unknown parameters of interest θ, and assume that there are at least as many moment conditions as parameters to be estimated $(s \geq k)$. Define the true parameter vector θ_0 as the unique solution of the system of moment conditions

$$E_F [g(y, \theta)] = 0, \tag{2.1}$$

where $E_F[\cdot]$ denotes expectation taken with respect to the (unknown) distribution function F of y. We assume that θ_0 belongs to the interior of a compact k-dimensional set Θ, the moment function $g(y, \theta)$ is continuous in θ for all $\theta \in \Theta$, and the expectation $E_F[g(y, \theta)]$ exists and is finite for all $\theta \in \Theta$. Define also the $(s \times k)$ matrix $G(\theta) \equiv E_F\left[\frac{\partial g(y,\theta)}{\partial \theta'}\right]$ and the $(s \times s)$ positive definite matrix $V(\theta) \equiv E_F\left[g(y, \theta) g(y, \theta)'\right]$, where the moment indicators $g(y, \theta)$ are assumed to be continuously differentiable in θ for all $\theta \in \Theta$. When these matrices are evaluated at θ_0 we write $G \equiv G(\theta_0)$ and $V \equiv V(\theta_0)$, in which case the former matrix is assumed to be full column rank.

The sample counterparts of $g(y, \theta)$, $G(\theta)$ and $V(\theta)$ are denoted by $g_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} g(y_i, \theta)$, $G_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(y_i,\theta)}{\partial \theta}$ and $V_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} g(y_i, \theta) g(y_i, \theta)'$, respectively. We assume that $G_n(\theta)$ converges almost surely and uniformly in θ to $G(\theta)$ and, by applying a Central Limit Theorem, that $\sqrt{n} g_n(\theta_0) \xrightarrow{d} N(0, V)$, i.e. the random vector $\sqrt{n} g_n(\theta_0)$ has a limiting distribution $N(0, V)$. When these quantities are evaluated at $\hat{\theta}$, a consistent estimator for θ_0, we write $\hat{g}_n \equiv g_n\left(\hat{\theta}\right)$, $\hat{G}_n \equiv G_n\left(\hat{\theta}\right)$ and $\hat{V}_n \equiv V_n\left(\hat{\theta}\right)$, which are assumed to be consistent estimators for $g(y, \theta_0)$, G and V, respectively.[1] An analogous notational scheme is followed for other variables

[1] In time series models $\hat{V}_n$ calculated in this way would not be a consistent estimator for V. Robust estimators for V in this context are discussed *inter alia* by Newey and West (1987a, 1994), Andrews (1991) and Den Haan and Levin (1997). Note, however, that $\hat{V}_n$ is a heteroskedasticity-robust consistent estimator for V.

throughout this dissertation.

All estimation methods discussed in the next sections are semi-parametric in the sense they only require the specification of moment conditions like (2.1) rather than the full density of the variable of interest. Hence, they are more robust, although generally less efficient, than parametric methods such as maximum likelihood. As we will see later on, all these methods are asymptotically first-order equivalent. For this reason, we adopt the same terminology to denote all of these estimators.

2.3 Generalized method of moments

2.3.1 Introduction

GMM is the conventional way of estimating models when the only information available about the population of interest is in the form of moment conditions. Since its formalization by Hansen (1982), GMM has become an important research topic in the econometric literature, both from theoretical and applied points of view. In the theoretical literature, the popularity of GMM results from two main facts. On the one hand, it provides a unifying framework for the analysis of many familiar estimators. Indeed, GMM nests various popular estimation methods such as maximum likelihood, instrumental variables and least squares, providing a suitable setting for their comparison. On the other hand, it is a simple alternative to other estimation techniques, especially when it is difficult to write down the likelihood function. For a general discussion about GMM, see, for example, Davidson and MacKinnon (1993, chap. 17), Hall (1993), Newey and McFadden (1994) and the recent book edited by Mátyás (1999), which is entirely devoted to this method.

With regard to applied work, due to its flexibility and generality, GMM estimation has been used in certain models which, otherwise, would be computationally very burdensome to estimate. Since one of the attractions of GMM estimation is to allow easy handling of stationary dependent data, this method gained particular importance in the estimation of time series models, namely asset pricing models, nonlinear dynamic rational expectation models, business-cycle models, stochastic volatility models and covariance structure models. However, GMM has also been widely applied to cross-sectional and panel data [see Ogaki (1993, p. 461) for a list of references of empirical examples].

The basic idea behind GMM is very simple: the vector of parameters of interest, θ_0, is estimated in such a way that the sample moment indicators $g_n(\theta)$ that correspond to the population moment conditions given by (2.1) are as close as possible to zero. When the number of moment conditions and unknown parameters is identical ($s = k$), the system of equations $g_n\left(\hat{\theta}\right) = 0$ can be solved directly in order to obtain $\hat{\theta}$ as an estimator of θ_0. However, in the most common case of overidentifying moment conditions, where there are more estimator-defining equations than parameters to be estimated ($s > k$), solving that system would produce multiple solutions for θ. Thus, the number of estimating equations has to be reduced in some way to k. Hansen (1982) proposed using k linear combinations of the s initial equations as described next.

Let S_n be a $(s \times s)$ symmetric, positive definite weighting matrix that may depend on the observations and converges almost surely to a nonrandom, positive definite matrix S. Hansen's (1982) GMM estimator is obtained by minimizing with respect to θ the following quadratic form of the sample moment conditions:

$$Q_n(\theta) \equiv g_n(\theta)' S_n g_n(\theta). \tag{2.2}$$

Here, S_n is used to measure the proximity of the sample moment indicators to zero, *via* closeness of the quadratic form (2.2) to zero. Note that $Q_n(\theta) \geq 0$, $\forall\ \theta \in \Theta$, being equal to zero only if and only if $g_n(\theta) = 0$, that is, as in the just identified case. The resultant k first-order conditions for this minimization problem are

$$\hat{G}_n' S_n \hat{g}_n = 0. \tag{2.3}$$

Using standard asymptotic theory, it can be proved [see Hansen (1982)] that the

GMM estimator $\hat{\theta}$, under the assumptions made in section 2.2, is consistent and asymptotically normally distributed:

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0,\Sigma\right), \tag{2.4}$$

where

$$\Sigma \equiv \left(G'SG\right)^{-1} G'SVSG \left(G'SG\right)^{-1}. \tag{2.5}$$

Analyzing (2.3), (2.4) and (2.5), we see that for each chosen weighting matrix S_n, a different GMM estimator will be obtained, with different asymptotic properties. Usually, S_n is chosen according to either computational convenience, in which case the identity matrix is frequently employed, or efficiency criteria, the weighting matrix selected being the one that minimizes the matrix Σ defined in (2.5). In the next section we discuss the issue of efficient GMM estimation.

2.3.2 Efficient estimation

Although the precision of an estimator is always an important matter, in the GMM framework it gains particular significance as the assumptions made are very weak. According to (2.5), the asymptotic variance matrix of $\hat{\theta}$ depends on both the matrix S and the moment conditions $g\left(y,\theta_0\right)$. If we merely wish to obtain consistent, rather than efficient[2], estimates of θ_0, we can consider any weighting matrix S and any moment conditions that satisfy the assumptions discussed earlier. However, if our aim is to obtain efficient GMM estimators, both S and $g\left(y,\theta_0\right)$ must be chosen in conformity with specific rules, as long as there are more moment conditions available than parameters to estimate. The choice of the matrix S is obvious, as we will see below, but the latter is a much more complicated issue, with a few exceptions in some specific cases. Note that all results presented in this sub-section are asymptotic.

[2]By efficient estimator, we mean the most efficient estimator within the general class of GMM estimators, not efficiency in the absolute sense, that is, considering all possible estimators for a particular model.

In small samples, the calculation of GMM estimators according to the guidelines described below may not be the best option (see section 2.3.6 for some Monte Carlo evidence).

Choice of the weighting matrix

In this sub-section we assume that the moment conditions are given. When their number is equal to the number of parameters to be estimated, the GMM estimator does not depend on the choice of the weighting matrix. Indeed, in this situation the GMM estimator solves the system of equations $g_n\left(\hat{\theta}\right) = 0$ and the expression of the covariance matrix given in (2.5) can be simplified to

$$\begin{aligned} \Sigma &= G^{-1}VG'^{-1} \\ &= \left(G'V^{-1}G\right)^{-1}, \end{aligned} \tag{2.6}$$

as the matrix G is now invertible. This is the covariance matrix of the so-called efficient GMM estimator. This estimator is the most efficient one within the class of GMM estimators, in the sense that it always attains the smallest variance of all possible GMM estimators. To see that (2.6) indeed represents a lower bound for this class of estimators, that is, (2.5)$\geq$(2.6), $\forall S$, in a positive semi-definite sense, note that this is equivalent to proving that

$$(G'SG)^{-1} G'SV^{\frac{1}{2}} \left[I - V^{-\frac{1}{2}}G\left(G'V^{-1}G\right)^{-1} G'V^{-\frac{1}{2}}\right] V^{\frac{1}{2}}SG\,(G'SG)^{-1} \geq 0.$$

As the matrix in the square brackets is an orthogonal projection matrix and thus idempotent, the left-hand side of this equation is always positive semi-definite, which proves immediately the previous statement.

When there are more moment conditions than parameters to be estimated, the existence of overidentifying restrictions imply that different GMM estimators are obtained for different choices of S_n. Efficient GMM estimation, as defined above, is now achieved only when a specific weighting matrix is used in the estimation procedure, namely a matrix S_n that converges to $S^* = V^{-1}$, the inverse of the limiting covariance matrix of the empirical moments evaluated at the true value of the parameters. As it is easily seen, substituting S^* for S into (2.5) produces (2.6).

The optimal weighting matrix V^{-1} depends on θ_0, so an initial consistent estimator for this parameter vector is needed for feasible estimation. Given the availability of a consistent, although generally inefficient, estimator of θ_0 (see section 2.3.3 for details), as long as $\hat{V}_n$ is a consistent estimator for V, the utilization of the former in place of the latter does not affect the asymptotic properties of the resultant efficient GMM estimator. Hence, the GMM estimator obtained from the minimization of the objective function

$$Q_n(\theta) \equiv g_n(\theta)' \hat{V}_n^{-1} g_n(\theta) . \tag{2.7}$$

will have the same first-order asymptotic properties as that obtained from using the true V in (2.7). However, different $\hat{V}_n^{-1}$ matrices will give rise to different GMM estimators, which can exhibit different behaviour in small samples.

In the econometric literature, a estimator obtained by minimizing an expression such as (2.7) is often termed a minimum chi-square estimator or an optimal minimum distance estimator. This is the GMM estimator that we consider throughout this dissertation, so we call it simply the GMM estimator, usually omitting 'efficient'. Whenever we refer to a GMM estimator not based on $S^* = V^{-1}$, we will call it explicitly a non-efficient GMM estimator.

Choice of the moment conditions

When $S = V^{-1}$, the above GMM estimator is the most efficient one for the given set of moment conditions. However, econometric models are generally specified in terms of conditional moments. From one set of conditional moments, we can draw an extensive set of unconditional moment conditions. For instance, $E[u(y,\theta_0)| X] = 0 \Rightarrow E[H'(X) u(y,\theta_0)] = 0$, for any measurable function H. Thus, there is a large set of candidates to be used as orthogonality conditions. For each one of these potential sets, there exists an efficient GMM estimator in the sense discussed in the previous sub-section. Although throughout this dissertation we consider the moment conditions as given, we discuss next, briefly, the issue of how to select, for a given weighting matrix, the set of moment conditions which yields the most asymptotically efficient estimators.

One might think that the more moment conditions used, the more efficient the resulting estimator. Actually, although it is true that the inclusion of extra moment conditions allows an improvement in the efficiency of parameter estimates[3], there seems to be a trade-off between asymptotic efficiency and bias in finite samples. Using a large number of overidentifying moment conditions may lead to a smaller asymptotic covariance matrix, but the estimates may be seriously biased in small samples. This conjecture is confirmed by some Monte Carlo simulation studies undertaken [see, for example, Kocherlakota (1990) and Ferson and Foerster (1994)], which strongly suggest that one should be quite parsimonious in the selection of the orthogonality conditions to be used in estimation.

The question of how to optimally choose the moment conditions has been studied mainly for the case of instrumental variables (one of the main applications of GMM), where the moment conditions can be expressed as a set of orthogonality conditions between a matrix $(n \times s)$ of instrumental variables, $H(X)$, and an n-vector of disturbance terms, $u(y, \theta_0)$, which is assumed to be known up to the parameter vector θ_0. In this setting, the problem is then how to choose the instrument matrix $H(X)$ which yields the asymptotically most efficient GMM estimator given the function $u(\cdot)$. This issue has been extensively studied for some specific cases. Earlier examples are Amemiya (1974) and Jorgenson and Laffont (1974), who calculated covariance matrix bounds for instrumental variables estimators in the homoskedastic disturbances case. The general case (but still in the instrumental variables context) was discussed by

[3]Provided they are not redundant, contributing additional information to that contained in the existing moment conditions.

Hansen (1985), who suggested a procedure for calculating the greatest lower bound for the asymptotic covariance matrix of GMM estimators. In fact, although the higher the number of relevant conditions used, the more efficient the GMM estimator, a lower bound exists for its covariance matrix. This bound was calculated for some specific instrumental variable models by Hansen (1985), Hansen, Heaton and Ogaki (1988) and Heaton and Ogaki (1991). However, even in those cases, the authors did not calculate an expression for the optimal instruments that would attain that bound. Actually, although in theory there exists an optimal set of instrumental variables which allows the GMM bound to be achieved, in applied work it has been proved difficult to calculate. Exceptions occur in certain models, mainly when independence among observations can be assumed. For instance, Tauchen (1986a) showed that, for a model where the disturbance terms are serially uncorrelated, Hansen's (1985) bound implies that the optimal instruments matrix is

$$\begin{aligned} H^*(X) &= \left\{ E\left[u(y,\theta_0)\, u(y,\theta_0)' \,|X \right] \right\}^{-1} E\left[\left. \frac{\partial u(y,\theta_0)}{\partial \theta'} \right| X \right] C \\ &= [\Omega(X)]^{-1} U(X)\, C, \end{aligned} \tag{2.8}$$

where C is any nonsingular, nonrandom $(k \times k)$ matrix. This is the same expression found by Amemiya (1977), using another approach and also assuming homoskedasticity.

It can be proved that the asymptotic covariance matrix attained when we use $H^*(X)$ as instruments,

$$\Sigma^* = \left\langle E_X \left\{ U(X)' [\Omega(X)]^{-1} U(X) \right\} \right\rangle^{-1}, \tag{2.9}$$

constitutes a lower bound for all instrumental variables estimators under the validity of the assumptions made [see, for example, Newey (1993)]. Furthermore, Chamberlain (1987), assuming independent and identically distributed observations, demonstrated that the semi-parametric efficiency bound for conditional moment restriction models, derived in that paper, is attained by the optimal instrumental variables estimator defined in (2.8). Therefore, expression (2.9) represents a lower bound for the asymptotic covariance matrix of any consistent, asymptotically normal estimator for a model defined by $E\left[u\left(y,\theta_0\right)|X\right]=0$ and not just for the instrumental variable estimator. Note that $H^*\left(X\right)$ is an $(n\times k)$ matrix which means that the optimal instruments reduce the s original moment conditions to only k. With as many moment conditions as unknown parameters, GMM estimation becomes independent of the choice of the weighting matrix and, thus, estimators based on $H^*\left(X\right)$ are also efficient in the sense discussed in the previous section. However, the matrix $H^*\left(X\right)$ depends on unknown parameters and functions. The conditional expectations present in (2.8) are generally not available, so the relevant issue now is how to construct feasible optimal GMM estimators, which must use as instrumental variable matrix an estimate of $H^*\left(X\right)$ that does not affect the asymptotic distribution of the GMM estimator.

The estimation of $H^*\left(X\right)$ can be a hard task, so these optimal instruments may be difficult to implement in practice. However, when $U\left(X\right)$ is known up to some parameters, then estimating the optimal instruments is much easier. For instance, in linear models, $U\left(X\right)=-X$ because $\frac{\partial u(y,\theta_0)}{\partial\theta'}$ depends only on the set of conditioning variables, so, with $C=-I$,

$$H^*\left(X\right)=\left[\Omega\left(X\right)\right]^{-1}X, \tag{2.10}$$

which yields the Generalized Least Squares (GLS) estimator. Even so, unless homoskedasticity and serial independence $[\Omega\left(X\right)=I]$ may be assumed, we need an estimate of Ω, that is, we need to use feasible GLS. As it is known, feasible GLS produces estimators asymptotically equivalent to those from infeasible GLS estimation as long as Ω is a known function of X and depends on a vector of parameters that can be consistently estimated by an auxiliary procedure. Otherwise, GLS based on a nonparametric estimation of the variance can be implemented along the lines described in Carroll (1982) and Robinson (1987).

As mentioned above, a detailed study of this issue lies outside the scope of this dissertation. Therefore, from now on and in all circumstances, we assume that the moment conditions are given.

2.3.3 Alternative computing procedures

In this sub-section, we discuss how to obtain GMM estimators in practice. When a weighting matrix S_n not dependent on θ_0 is chosen, for example the identity matrix, then a one-step estimation procedure may be employed: the set of equations (2.3) is solved employing a standard numerical optimization routine. If an efficient GMM estimator is the aim, then an initial consistent estimate of the matrix V is needed, which in turn requires the availability of an initial consistent estimator for θ_0. As solution to this "circular" problem, a two-step procedure is generally utilized, which can be described as follows:

1. Find a consistent estimator for θ_0, for instance by using a non-efficient GMM estimator based on a matrix S_n that does not depend on θ_0 (although this preliminary estimate need not be obtained by GMM),

$$G_n\left(\hat{\theta}^1\right)' S_n g_n\left(\hat{\theta}^1\right) = 0, \tag{2.11}$$

   where $\hat{\theta}^1$ denotes the resultant one-step estimator of θ_0; $\hat{\theta}^1$ will generally be inefficient but consistent, so it can used to construct $\hat{V}_n \equiv V_n\left(\hat{\theta}^1\right)$, a consistent estimator of V;

2. Solve again the set of equations (2.11) but now using $S_n = \hat{V}_n^{-1}$,

$$G_n\left(\hat{\theta}\right)' \hat{V}_n^{-1} g_n\left(\hat{\theta}\right) = 0, \tag{2.12}$$

   in order to obtain $\hat{\theta}$, an efficient estimator of θ_0.

Usually, the resultant efficient GMM estimator is called a two-step GMM estimator. Although in theory one iteration in the second step is enough for achieving asymptotic efficiency, it is now much more common to iterate the two-step GMM estimator until full convergence is reached again, keeping $\hat{V}_n$ fixed in all iterations. All two-step GMM estimators referred to later in our Monte Carlo simulation studies were calculated using the latter technique.

Recently, Hansen, Heaton and Yaron (1996) proposed another computational procedure to estimate efficiently moment condition models by GMM. Their estimator, called the repeatedly-iterated GMM estimator, can be obtained by an analogous process to the one described above for the two-step GMM estimator but in the second step, instead of using the initial weighting matrix $\hat{V}_n \equiv V_n\left(\hat{\theta}^1\right)$ in all iterations, this matrix is re-estimated in each iteration in such a way that $V_n\left(\hat{\theta}^{j-1}\right)$ is used to estimate $\hat{\theta}^j$ in iteration j. Note that, although the weighting matrix continues to be treated as given in each iteration, this estimator can be characterized by the first-order conditions:

$$G_n\left(\hat{\theta}\right)'\left[V_n\left(\hat{\theta}\right)\right]^{-1} g_n\left(\hat{\theta}\right) = 0. \tag{2.13}$$

2.3.4 Specification tests

Several specification tests for models estimated by the efficient GMM have been proposed. In this section, we present the main existing tests for overidentifying moment conditions, additional moment conditions and parametric restrictions. Non-nested hypothesis will be discussed later, in chapter 5, where some alternative tests are developed.

Tests of overidentifying moment conditions

Hansen (1982) proposed the J test for assessing the specification of a model estimated by GMM, which is probably the most frequently applied test in the moment condition framework. The construction of this test and the idea behind it is very simple. If there are s estimator-defining equations and k parameters to be estimated, with $s > k$, only k moments are needed to identify the k parameters, so there are $s - k$ overidentifying moment conditions. One way to test whether all moment conditions are satisfied is to check if their sample versions are as close to zero as would be expected in that case. Hence, Hansen (1982) suggested using n times the minimized value of the GMM criterion function (2.7),

$$J \equiv n\hat{g}_n' \hat{V}_n^{-1} \hat{g}_n, \tag{2.14}$$

which he showed to possess a limiting chi-squared distribution with $s-k$ degrees of freedom under the hypothesis that all moment conditions are valid.

The J test may be adapted to test the correctness of only a subset of moment conditions. Following Eichenbaum, Hansen and Singleton (1988), partition the sample moments vector $g_n(\theta)'$ as $\left[\begin{array}{cc} g_{1n}(\theta)' & g_{2n}(\theta)' \end{array}\right]$, where the population moment conditions corresponding to the second sub-vector are presumed to hold only under the null hypothesis and the ones concerning the first are assumed also to be valid under the alternative hypothesis. Let s_1 and s_2 be the dimension of each sub-vector, respectively, with $s = s_1 + s_2$ and $s_1 \geq k$. Partition the matrix V as $\left[\begin{array}{cc} V_{11} & V_{12} \\ V_{21} & V_{22} \end{array}\right]$ conformably with the two subsets of moment conditions. Consider two optimizations, one based on the full set of moment conditions, the other using only the first s_1 moments. Under these assumptions, we can then assess the validity of the second subset of moment conditions through the test statistic

$$J_2 \equiv n\left[Q_n\left(\hat{\theta}\right) - Q_{1n}\left(\hat{\theta}_1\right)\right], \tag{2.15}$$

where $Q_{1n}\left(\hat{\theta}_1\right) = g_{1n}\left(\hat{\theta}_1\right)' \hat{V}_{11n}^{-1} g_{1n}\left(\hat{\theta}_1\right)$ and $\hat{\theta}_1$ is the vector of parameter estimates from the second minimization. Under the null hypothesis, the asymptotic distribution of J_1 is chi-square with s_2 degrees of freedom. Note that J_2 is equal to the difference between the J statistic, given by (2.14), and J_1, the statistic that would be used to test the validity of the moment conditions based on $g_1(y, \theta_0)$.

Tests for additional moment conditions

The J_2 test may also be applied to examine whether an additional s_2-dimensional vector of moments has mean zero and, thus, may be incorporated in the moment conditions (2.1) in order to improve inference on θ_0. For this purpose, assume that an initial GMM estimation based only on the vector $g_{1n}(\theta)$ defined above was performed, with the estimate $\hat{\theta}_1$ being obtained, and interpret $g_{2n}(\theta)$ as the sample counterpart of the set of additional moment conditions $E_F[g_2(y,\theta_0)] = 0$. It is easily seen that the statistic J_2 may be used to assess this hypothesis.

Alternatively, we may follow the approach due to Newey (1985b) and Tauchen (1985) and employ the test statistic

$$CM \equiv n\hat{g}_{2n}' \left(\hat{A}_n \hat{V}_n \hat{A}_n'\right)^{-1} \hat{g}_{2n}, \tag{2.16}$$

where $\hat{\cdot}$ means evaluation at $\hat{\theta}_1$, $\hat{A}_n$ is a consistent estimator for

$$A \equiv \left[\begin{array}{cc} -G_2\left(G_1' V_{11}^{-1} G_1\right)^{-1} G_1' V_{11}^{-1} & I_{s_2} \end{array}\right], \tag{2.17}$$

$G' \equiv \left[\begin{array}{cc} G_1' & G_2' \end{array}\right]$ and I_{s_2} is an s_2-dimensional identity matrix. This statistic has also a chi-squared distribution with s_2 degrees of freedom. In contrast to the J_2 test, only one model estimation is required.

Tests of parametric restrictions

As with any other estimation procedure, we may wish to test whether some specific restrictions concerning the vector of parameters of interest may be entertained in the model. To this end, in the GMM context it is possible to employ the usual Wald (W), Lagrange multiplier (LM), distance metric (DM), minimum chi-squared (MC), Hausman (H) and Neyman's (1959) $C(\alpha)$ statistics. The first four tests were proposed by Newey and West (1987b) and the last two in Newey and McFadden (1994) and Davidson and MacKinnon (1993), respectively. See those papers for the proofs of the results given below.

Consider the null hypothesis $H_0 : r(\theta_0) = 0$, where $r(\cdot)$ is a known, possibly non-linear, continuously differentiable q-vector representing certain parametric restrictions expressed in constraint equation form and q is the number of restrictions ($q \leq k$), and denote $\frac{\partial r(\theta)}{\partial \theta'}$ by $R(\theta)$, a full row rank $(q \times k)$ matrix. All statistics presented below have a limiting chi-squared distribution with q degrees of freedom under H_0.

The Wald statistic for testing H_0 is

$$W_n \equiv n\hat{r}' \left(\hat{R}\hat{\Sigma}_n\hat{R}' \right)^{-1} \hat{r}, \tag{2.18}$$

where $\hat{}$ denotes evaluation at the unconstrained GMM estimator $\hat{\theta}$ and $\hat{\Sigma}_n$ is a consistent estimator for the matrix Σ defined in (2.6). As usual for Wald statistics, this test has the disadvantage of not being invariant under reparametrization of the restrictions. On the other hand, it does not require optimization of the constrained model.

With regard to the LM or score test, it is only necessary to estimate the restricted model. Let $\tilde{\theta}$ be the constrained GMM estimator obtained by minimizing $Q_n(\theta)$ subject to the restrictions $r(\theta) = 0$. The test statistic for H_0 is:

$$LM_n \equiv n\tilde{g}_n' \tilde{V}_n^{-1} \tilde{G}_n \tilde{\Sigma}_n \tilde{G}_n' \tilde{V}_n^{-1} \tilde{g}_n. \tag{2.19}$$

If there are as many moment conditions as parameters, then $\tilde{G}_n$ is a square matrix and, thus, (2.19) can be written as $LM = nQ_n\left(\tilde{\theta}\right)$, that is, this test turns to be numerically identical to the J test based on $\tilde{\theta}$, provided that the same estimate of V is used in both statistics.

As for the distance metric or likelihood ratio-type test, the test statistic for H_0 is:

$$DM_n \equiv n\left[Q_n\left(\tilde{\theta}\right) - Q_n\left(\hat{\theta}\right)\right]. \tag{2.20}$$

Note that the same estimator of V must be used for both restricted and unrestricted estimation to guarantee that DM_n is non-negative [Newey and West (1987b)]. The main disadvantage of this test is that it requires two optimizations. If $s = k$, then $Q_n\left(\hat{\theta}\right) = 0$ and $DM = nQ_n\left(\tilde{\theta}\right)$, and, consequently, this test turns to be numerically equal to the J and LM tests.

The minimum chi-squared test statistic for H_0 is given by

$$MC_n \equiv n\left(\hat{\theta} - \tilde{\theta}\right)' \Sigma_n^{-1}\left(\hat{\theta} - \tilde{\theta}\right) \tag{2.21}$$

and the Hausman test by:

$$H_n \equiv n\left(\hat{\theta} - \tilde{\theta}\right)' R' \left(R\Sigma_n R'\right)^{-1} R\left(\hat{\theta} - \tilde{\theta}\right). \tag{2.22}$$

Like the distance metric test, two optimizations are needed to perform these tests. Estimation for the matrices R and Σ may be evaluated at either $\hat{\theta}$ or $\tilde{\theta}$.

Finally, Davidson and MacKinnon (1993) derived a $C(\alpha)$ test. Let $\dot{\theta}$ be any root-n consistent estimator of θ_0 that satisfies the null hypothesis, that is, $r\left(\dot{\theta}\right) = 0$. The test statistic for H_0 is:

$$C_n(\alpha) \equiv n\dot{g}_n' \dot{V}_n^{-1} \dot{G}_n \dot{\Sigma}_n \dot{R}' \left(\dot{R}\dot{\Sigma}_n \dot{R}'\right)^{-1} \dot{R}\dot{\Sigma}_n \dot{G}_n' \dot{V}_n^{-1} \dot{g}_n. \tag{2.23}$$

This statistic only requires $\dot{\theta}$ rather than the GMM estimator.

2.3.5 Generating GMM test statistics

Before continuing our survey on estimation methods for moment condition models, we present in this section an original way of deriving some of the tests discussed in the previous section. Furthermore, we consider a more general setting of implicit parametric restrictions, which allows us to generalize those tests for other kind of constraints. Essentially, we extend Smith's (1987) results for maximum likelihood estimators to the GMM framework.

A linearized GMM statistic for testing implicit parametric restrictions

Suppose that we aim to test whether some specific restrictions, involving not only the vector of parameters of interest θ but also a p-dimensional vector α of auxiliary parameters, may be entertained in a model estimated by GMM. We assume that the null hypothesis to be assessed can be expressed as

$$H_0 : r(\theta_0, \alpha_0) = 0, \tag{2.24}$$

where α_0 is the true value of α and, as before, $r(\cdot)$ is a q-vector of restrictions and q is the number of restrictions. The function $r(\cdot)$ satisfies the implicit function theorem, that is, $r(\cdot) = 0$ has a unique solution and the $(q \times k)$ matrix $R_\theta \equiv \frac{\partial r(\theta,\alpha)}{\partial \theta'}$ and the $(q \times p)$ matrix $R_\alpha \equiv \frac{\partial r(\theta,\alpha)}{\partial \alpha'}$ are of rank q and p, respectively, for values of (θ, α) near to the true values (θ_0, α_0), with $p \leq q \leq k$. The specification of the restrictions in the very general *implicit* form (2.24) includes the more common cases of *constraint* $[r(\theta_0) = 0]$ and *freedom* $[\theta_0 = r_\alpha(\alpha_0)]$ equation restrictions as special cases.

The standard approach for constructing Wald statistics to test the null hypothesis (2.24) against $H_1 : r(\theta_0, \alpha_0) \neq 0$ is not directly applicable here, since α is only identified under the null hypothesis. To circumvent this difficulty, Szroeter (1983) suggested using as an estimate of α_0 the solution $\hat{\alpha}$ to the program

$$\min_{\alpha} \Lambda(\alpha) \equiv r\left(\hat{\theta}, \alpha\right)' \Phi_n r\left(\hat{\theta}, \alpha\right), \tag{2.25}$$

where Φ_n is a $(q \times q)$ positive semi-definite matrix that converges almost surely to a nonrandom, positive definite matrix Φ. The estimator $\hat{\alpha}$ thus obtained is function of the unrestricted estimator $\hat{\theta}$ given the matrix Φ_n: $\hat{\alpha} = f\left(\hat{\theta}\middle| \Phi_n\right)$. The function $f(\cdot)$ can be defined for each value of θ, which allows the reformulation of the restrictions (2.24) into a constraint equation form,

$$r(\theta_0, \alpha_0) = r[\theta_0, f(\theta_0| \Phi)] = h_\Phi(\theta_0| \Phi) = 0. \tag{2.26}$$

Using the reformulated restrictions (2.26), which now omit α, we may construct a standard Wald statistic for testing H_0 against H_1,

$$W \equiv n\hat{h}'_\Phi \left(\hat{H}_\Phi \Sigma_n \hat{H}'_\Phi \right)^- \hat{h}_\Phi, \qquad (2.27)$$

where $\hat{h}_\Phi \equiv h_\Phi \left(\hat{\theta} \middle| \Phi_n \right) \equiv r\left(\hat{\theta}, \hat{\alpha} \right)$, $H_\Phi \equiv \frac{\partial h_\Phi}{\partial \theta'} = M_\Phi R_\theta$ is a $(q \times k)$ matrix, $M_\Phi \equiv I - R_\alpha \left(R'_\alpha \Phi_n R_\alpha \right)^{-1} R'_\alpha \Phi_n$ is a $(q \times q)$ idempotent matrix of rank $q - p$, $(\cdot)^-$ denotes a generalized inverse and $\hat{}$ denotes evaluation at $\left(\hat{\theta}, \hat{\alpha} \right)$. This statistic has a limiting chi-squared distribution with $q - p$ degrees of freedom under H_0. For a full discussion of the derivation of this test statistic see Szroeter (1983).

In the maximum likelihood context, Smith (1987) developed a linearized classical statistic, which shares the same first-order properties as W. In a similar manner, a linearized GMM ($LGMM$) statistic may also be constructed. To that end, the first point to note is that the unconstrained efficient GMM estimator $\hat{\theta}$ can be obtained from the first-step of, for instance, the Newton-Raphson algorithm based on an consistent and asymptotically normal unconstrained estimator $\ddot{\theta}$ for θ_0,

$$\hat{\theta} = \ddot{\theta} - \Sigma_n \ddot{G}'_n V_n^{-1} \ddot{g}_n; \qquad (2.28)$$

a single iteration is enough to assure the asymptotic efficiency of $\hat{\theta}$. Then, linearizing the constraints $\hat{h}_\Phi$ around $\ddot{\theta}$ produces:

$$\begin{aligned} \hat{h}_\Phi &= \ddot{h}_\Phi + \ddot{H}_\Phi \left(\hat{\theta} - \ddot{\theta} \right) + o_p\left(n^{-\frac{1}{2}} \right) \\ &= \ddot{h}_\Phi - \ddot{H}_\Phi \Sigma_n \ddot{G}'_n V_n^{-1} \ddot{g}_n + o_p\left(n^{-\frac{1}{2}} \right). \end{aligned} \qquad (2.29)$$

Finally, substituting (2.29) into (2.27) yields the $LGMM$ statistic

$$LGMM = n\left(\ddot{h}_\Phi - \ddot{H}_\Phi \Sigma_n \ddot{G}'_n V_n^{-1} \ddot{g}_n \right)' \left(\ddot{H}_\Phi \Sigma_n \ddot{H}'_\Phi \right)^- \left(\ddot{h}_\Phi - \ddot{H}_\Phi \Sigma_n \ddot{G}'_n V_n^{-1} \ddot{g}_n \right), \quad (2.30)$$

which has a limiting chi-square distribution with $q - p$ degrees of freedom under H_0

and can be used to test the hypothesis (2.24) in the GMM framework. Note that $\ddot{}$ denotes evaluation at $\left(\ddot{\theta},\ddot{\alpha}\right)$, where $\ddot{\alpha}=f\left(\ddot{\theta}\middle|\Phi_n\right)$, that is $\ddot{\alpha}$ solves (2.25) with $\hat{\theta}$ replaced by $\ddot{\theta}$.

For the statistic $LGMM$ to be operational, the matrix Φ_n in (2.25) and the generalized inverse present in (2.30) must be chosen. The asymptotic power of the test based on W, and thus that of the $LGMM$ statistic, is independent of those choices, as shown by Szroeter (1983). However, in finite samples, the statistical properties of both W and $LGMM$ tests will depend on the matrices selected. For any given choice of the matrix Φ_n, a generalized inverse for $\hat{H}_\Phi\Sigma_n\hat{H}_\Phi'$ is $M_\Omega'\Omega M_\Omega$, with $\Omega\equiv\left(R_\theta\Sigma R_\theta'\right)^{-1}$ and $M_\Omega\equiv I-R_\alpha\left(R_\alpha'\Omega R_\alpha\right)^{-1}R_\alpha'\Omega$.[4] On the other hand, a choice for Φ_n that allows certain simplifications to these and other statistics presented below is a consistent estimator Ω_n for Ω, in which case α_0 is efficiently estimated.[5] In this case, the matrix Ω is a generalized inverse for $\hat{H}_\Phi\Sigma_n\hat{H}_\Phi'=\hat{M}_\Omega\Omega_n^{-1}\hat{M}_\Omega'$ and, since $r\left(\hat{\theta},\hat{\alpha}\right)\equiv\hat{h}_\Omega$, the W statistic is numerically equal to $n\Lambda\left(\hat{\alpha}\right)$.

We can define similar statistics to (2.30) appropriate for the more familiar constraint and freedom equation restrictions. In the former case, $H_0:r\left(\theta_0\right)=0$, so $R_\alpha=0$, which implies $M_\Phi=I_q$ and $H_\Phi=R_\theta$. Hence, it follows that

$$LGMM_c=n\left(\ddot{r}-\ddot{R}_\theta\Sigma_n\ddot{G}_n'V_n^{-1}\ddot{g}_n\right)'\left(\ddot{R}_\theta\Sigma_n\ddot{R}_\theta'\right)^{-1}\left(\ddot{r}-\ddot{R}_\theta\Sigma_n\ddot{G}_n'V_n^{-1}\ddot{g}_n\right), \qquad (2.31)$$

which has an asymptotic chi-squared distribution with q degrees of freedom.

For freedom equation restrictions, as $H_0:\theta_0=r_\alpha\left(\alpha_0\right)$, we have $q=k$ and $R_\theta=I_k$, which gives $H_\Phi=M_\Phi$ and $\Omega=\Sigma^{-1}$. Let $\breve{\theta}=r_\alpha\left(\ddot{\alpha}\right)$, with $\ddot{\alpha}$ solving:

$$\min_\alpha\Lambda\left(\alpha\right)=\left[\ddot{\theta}-r_\alpha\left(\alpha\right)\right]'\Phi_n\left[\ddot{\theta}-r_\alpha\left(\alpha\right)\right]; \qquad (2.32)$$

[4]Note that $H_\Phi\Sigma H_\Phi'M_\Omega'\Omega M_\Omega H_\Phi\Sigma H_\Phi'=M_\Phi\Omega^{-1}M_\Phi'M_\Omega'\Omega M_\Omega M_\Phi\Omega^{-1}M_\Phi'=M_\Phi\Omega^{-1}M_\Phi'$, for $M_\Omega M_\Phi=M_\Omega$ and $\Omega^{-1}M_\Omega'\Omega M_\Omega\Omega^{-1}=M_\Omega'\Omega^{-1}M_\Omega$.

[5]Note that in this case an initial consistent estimator of α is needed to evaluate $\Phi_n=\Omega_n$ in (2.25). It can be obtained solving also (2.25) but considering a matrix Φ_n, such as the identity matrix, not dependent on unknown parameter values.

cf. (2.25). The first-order conditions corresponding to (2.32) are given by $\ddot{R}'_{\alpha}\Phi_n\left(\ddot{\theta}-\breve{\theta}\right)$ $=0$, so $\ddot{M}_{\Phi}\left(\ddot{\theta}-\breve{\theta}\right)=\left(\ddot{\theta}-\breve{\theta}\right)$. Using $M'_{\Omega}\Omega M_{\Omega}=M'_{\Sigma^{-1}}\Sigma^{-1}M_{\Sigma^{-1}}$ as generalized inverse and expression (2.28) for $\hat{\theta}$, we obtain

$$\begin{aligned} LGMM_f &= n\left(\ddot{\theta}-\breve{\theta}-\Sigma_n\ddot{G}'_nV_n^{-1}\ddot{g}_n\right)'\ddot{M}'_{\Sigma^{-1}}\Sigma_n^{-1}\ddot{M}_{\Sigma^{-1}}\left(\ddot{\theta}-\breve{\theta}-\Sigma_n\ddot{G}'_nV_n^{-1}\ddot{g}_n\right) \\ &= n\left(\hat{\theta}-\breve{\theta}\right)'M'_{\Sigma^{-1}}\Sigma_n^{-1}M_{\Sigma^{-1}}\left(\hat{\theta}-\breve{\theta}\right). \end{aligned} \tag{2.33}$$

The degrees of freedom are now $k-p$.

Besides being suitable for testing parametric restrictions in models estimated by GMM, the three $LGMM$ statistics derived, when evaluated at certain estimators, give rise to various well known tests, as is emphasized next.

A test statistic generating equation

Consider again the $LGMM$ statistic given in (2.30) but permitting any consistent asymptotically normal estimator of θ_0 (and, hence, α_0) to be used:

$$TSGE\left(\theta,\alpha\right)=n\left(h_{\Phi}-H_{\Phi}\Sigma_nG'_nV_n^{-1}g_n\right)'\left(H_{\Phi}\Sigma_nH'_{\Phi}\right)^{-}\left(h_{\Phi}-H_{\Phi}\Sigma_nG'_nV_n^{-1}g_n\right). \tag{2.34}$$

This equation may be viewed as a test statistic generating equation ($TSGE$) for implicit parametric restrictions in the GMM framework. Accordingly to the estimators (θ,α) at which $TSGE$ is evaluated, several different test statistics may be obtained. Note that all results presented below are exact.

First of all, evaluating $TSGE$ at $\left(\ddot{\theta},\ddot{\alpha}\right)$, where $\ddot{\alpha}=f\left(\ddot{\theta}\middle|\Phi_n\right)$, gives, obviously, the $LGMM$ statistic.

Second, using $\left(\hat{\theta},\hat{\alpha}\right)$, where $\hat{\theta}$ is the efficient GMM estimator and $\hat{\alpha}=f\left(\hat{\theta}\middle|\Phi_n\right)$, implies that $\hat{G}'_nV_n^{-1}\hat{g}_n=0$, so

$$TSGE\left(\hat{\theta},\hat{\alpha}\right)=n\hat{h}'_{\Phi}\left(\hat{H}_{\Phi}\Sigma_n\hat{H}'_{\Phi}\right)^{-}\hat{h}_{\Phi}, \tag{2.35}$$

the generalized Wald statistic (2.27).

Third, let $\left(\dot{\theta},\dot{\alpha}\right)$ be any consistent estimator of θ_0 and α_0 which satisfy $r\left(\dot{\theta},\dot{\alpha}\right)=0$. Then, since $\dot{H}_\Phi'\left(\dot{H}_\Phi\Sigma_n\dot{H}_\Phi'\right)^-\dot{H}_\Phi=\dot{H}_\Omega'\Omega_n\dot{H}_\Omega$, it follows that

$$TSGE\left(\dot{\theta},\dot{\alpha}\right)=n\dot{g}_n'V_n^{-1}\dot{G}_n\Sigma_n\dot{H}_\Omega'\Omega_n\dot{H}_\Omega\Sigma_n\dot{G}_n'V_n^{-1}\dot{g}_n, \tag{2.36}$$

Neyman's $C\left(\alpha\right)$ statistic.[6]

Finally, consider the efficient constrained GMM estimator $\tilde{\theta}$ that results from

$$\min_{\theta,\alpha} Q_n\left(\theta\right) \text{ subject to } r\left(\theta,\alpha\right)=0. \tag{2.37}$$

Solving this optimization problem using a Lagrangian function yields as first-order conditions $\tilde{h}_\Phi=r\left(\tilde{\theta},\tilde{\alpha}\right)=0$, $\tilde{G}_n'V_n^{-1}\tilde{g}_n+\tilde{R}_\theta'\tilde{\psi}=0$ and $\tilde{R}_\alpha'\tilde{\psi}=0$, where $\tilde{\psi}$ is the Lagrange multiplier associated with the restrictions. Substituting $\tilde{G}_n'V_n^{-1}\tilde{g}_n$ for $\tilde{R}_\theta'\tilde{\psi}$ into (2.36) and noting that $\tilde{\psi}'\tilde{M}_\Omega=\tilde{\psi}$, we see that evaluation of $TSGE$ at $\left(\tilde{\theta},\tilde{\alpha}\right)$ produces

$$\begin{aligned} TSGE\left(\tilde{\theta},\tilde{\alpha}\right) &= n\tilde{\psi}'\tilde{R}_\theta\Sigma_n\tilde{R}_\theta'\tilde{\psi} &(2.38)\\ &= n\tilde{g}_n'V_n^{-1}\tilde{G}_n\Sigma_n\tilde{G}_n'V_n^{-1}\tilde{g}_n, &(2.39)\end{aligned}$$

the two usual forms of a Score test based on GMM estimation.

Test statistics for constraint and freedom equation restrictions may be obtained defining a similar test generating equation to (2.34) but from (2.31) and (2.33), respectively. Alternatively, we can derive them directly from (2.35), (2.36) or (2.39), in the same way as (2.31) and (2.33) were deduced from (2.30). For the case of constraint equation restrictions, versions of the above Wald and Score statistics were presented by Newey and West (1987b) and of the $C\left(\alpha\right)$ statistic by Davidson and MacKinnon

[6] Remember that α is not identified under the alternative hypothesis. Thus, as this statistic is evaluated under H_0, $\dot{\alpha}\neq f\left(\dot{\theta}\middle| S_n\right)$, as we do not need to use program (2.25) to estimate α. The same happens with the Score test presented next.

(1993), as discussed in section 2.3.4.

We now demonstrate that the $TSGE$ given in (2.34) may be adapted to another type of hypothesis. Indeed, it is straightforward to show that the statistic J given in (2.14) may also be obtained from the test generating equation described above. Now, defining $r(\theta, \alpha)$ as $g_n(\theta)$ and, hence, R_θ as G_n, equation (2.34) can be simplified to

$$\begin{aligned} TSGE(\theta) &= n\left(g_n - G_n\Sigma_n G_n' V_n^{-1} g_n\right)' \left(G_n\Sigma_n G_n'\right)^{-} \left(g_n - G_n\Sigma_n G_n' V_n^{-1} g_n\right) \\ &= n g_n' M_{V^{-1}}' V_n^{-1} M_{V^{-1}} g_n, \end{aligned} \tag{2.40}$$

where we have used V_n^{-1} as generalized inverse for $G_n\Sigma_n G_n'$. As $rank\left(M_{V^{-1}}\right) = s - k$, the two tests of overidentifying moment conditions we next generate from equation (2.40) have, under H_0, a limiting chi-squared distribution with $s-k$ degrees of freedom. First, evaluating (2.40) at $\hat{\theta}$ implies $\hat{M}_{V^{-1}}\hat{g}_n = \hat{g}_n$, which yields

$$TSGE\left(\hat{\theta}\right) = n\hat{g}_n' V_n^{-1}\hat{g}_n, \tag{2.41}$$

the J test. Second, evaluation of (2.40) at $\ddot{\theta}$ produces

$$TSGE\left(\ddot{\theta}\right) = n\ddot{g}_n' \ddot{M}_{V^{-1}}' V_n^{-1} \ddot{M}_{V^{-1}} \ddot{g}_n, \tag{2.42}$$

which we designate as a linearized J statistic because it could be deduced by linearizing $\hat{g}_n$ around $\ddot{\theta}$ of (2.28) and substituting the consequent result into the expression for the J test. While (2.41) can only be evaluated at efficient GMM estimators, (2.42) may be used to assess any model estimated by GMM, irrespective of the weighting matrix employed in GMM estimation.

2.3.6 Limitations

To complete our survey of the GMM estimator and motivate the following discussion of alternative estimation methods for moment condition models, we now discuss some important drawbacks of this estimator, the principal of them being their finite sample behaviour. In fact, it has long been recognized that the first-order asymptotic distribution of the GMM estimator provides a poor approximation to its small sample distribution. There is increasing Monte Carlo evidence indicating that in finite samples GMM estimators may be badly biased and the associated tests may have actual sizes substantially different from the nominal ones. One of the first studies of this issue was Tauchen (1986b), using data generated from artificial nonlinear consumption-based asset pricing models. He concluded that the two-step GMM estimator performed reasonably well but was very sensitive to the choice of instruments: the more lags (instruments), the lower the variance of the estimators, but, at the same time, the higher their bias. Kocherlakota (1990) conducted a similar study confirming the results obtained in Tauchen (1986b) and also finding evidence on the tendency of the J test to reject the true hypothesis too often. Ferson and Foerster (1994) considered another example, estimating a seemgly unrelated regression model with cross equation restrictions for asset returns using real financial data. They concluded that, in simple models, with few assets, the biases of the estimators were relatively small but their standard errors were underestimated, mainly for smaller sample sizes and where more instruments were utilized. In more complex models, with more assets, they found that both the coefficient estimates and the estimated standard errors could be severely biased.

The level of the concern about the small-sample properties of GMM estimators has been sufficient great that, recently (in 1996), the *Journal of Business & Economic Statistics* dedicated a special issue to studies investigating this question. Among other articles published there, Andersen and Sorensen (1996), considering stochastic volatility models, confirmed previous results about the deleterious effect of the number of instruments on the performance of GMM estimators in small samples and Altonji and Segal (1996), simulating covariance structure models, reported that, at least in some circumstances, the two-step GMM estimator displays worse behaviour than a non-efficient GMM estimator obtained using the identity matrix as weighting matrix.

Hansen, Heaton and Yaron (1996), also in that issue, undertook an extensive

Monte Carlo study comparing the performance of, among others, the two-step and the repeatedly-iterated GMM estimators. In their experiments, both methods performed very poorly in many cases, producing heavy biases in the parameter estimates. Moreover, the J test led to over-rejection of the null hypothesis and Wald tests of parametric restrictions were unreliable. Surprisingly, the repeatedly-iterated estimator in some cases behaved the worst, whereas originally it had been proposed as alternative to the two-step estimator with the objective of improving its finite-sample properties. Hansen, Heaton and Yaron (1996) also noticed that the poorest performances occurred when more moment conditions were used.

To summarize, most of these investigations suggest that the finite sample performance of GMM is sensitive to both the number of moment conditions and the sample size. When the former increases or the latter decreases, the finite sample properties of the GMM estimator and related statistics deviate more from the nominal asymptotic properties, no matter which kind of GMM estimator is used. In this dissertation, we provide additional evidence on this subject and investigate the behaviour of alternative methods.

With regard to the J test, there is an additional problem. Newey (1985a) argued that the J test may fail to detect a misspecified model, showing that there exist local alternatives for which the assumed population moment conditions are invalid but the statistic J will still converge to a χ^2_{s-k} random variable. Therefore, the J test may be inconsistent in some circumstances, failing to reject the null hypothesis of no misspecification when the model is incorrect. An example of such a situation occurs when the sample is characterized by structural instability [see Ghysels and Hall (1990a)].

A final problem associated with GMM is that the use of a consistent estimate of the optimal weighting matrix, instead of estimating it jointly with the model parameters, can lead to the sensitivity of GMM estimators to the choice of the initial weighting matrix. Indeed, GMM estimation, by holding that matrix fixed, fails to account for the dependence of the weighting matrix on the parameter vector. Hence, unless we have as many parameters as moment conditions, GMM estimators are not invariant to linear transformations of the original moment functions.

2.4 The continuous-updating GMM estimator

As some of the problems of the standard GMM estimator seem to arise from the necessity of utilizing a consistent estimate of V in an initial step, alternative one-step methods have recently been suggested. One of those alternatives is the CU-GMM estimator, proposed by Hansen, Heaton and Yaron (1996). This estimator is obtained by minimizing the function

$$Q_n(\theta) \equiv g_n(\theta)' [V_n(\theta)]^{-1} g_n(\theta), \tag{2.43}$$

whose first-order conditions are given by

$$\hat{G}_n' \hat{V}_n^{-1} \hat{g}_n + \frac{1}{2} \sum_{j=1}^{s} \hat{g}_{jn} \left(\frac{\partial \hat{V}_n^j}{\partial \theta} \right)' \hat{g}_n = 0, \tag{2.44}$$

where $\hat{g}_{jn}$ corresponds to the j^{th} sample moment condition, $\hat{V}_n^j$ is the j^{th} s-dimensional row of the matrix $\hat{V}_n^{-1}$ and $\frac{\partial \hat{V}_n^j}{\partial \theta}$ is a $(s \times k)$ matrix. As the weighting matrix and the parameters are now estimated simultaneously, the CU-GMM estimator is invariant to parameter-dependent transformations of the moment indicators. The weighting matrix used in (2.43) can no longer be viewed as a nonrandom matrix but, nevertheless, according to Hansen, Heaton and Yaron (1996), the extra term present in the first-order conditions (2.44) does not distort the limiting distribution of the CU-GMM estimator relative to the standard GMM estimator. Therefore, the two estimators are asymptotically first-order equivalent and, hence, all specification tests discussed in the previous section could also be evaluated at the CU-GMM estimator.

In fact, the presence of the additional term in (2.44) implies that the CU-GMM estimator should have smaller bias in finite samples than the standard GMM estimator, as both Donald and Newey (2000) and Newey and Smith (2000) argue. The former authors gave a jackknife interpretation of the CU-GMM estimator, demonstrating that, in (2.44), own observation terms are automatically deleted, which eliminates one known important source of bias for GMM estimators. On the other hand, Newey and Smith (2000) derived stochastic expansions for both estimators, providing asymptotic expressions for their biases. They find that the asymptotic bias of the CU-GMM estimator is given by

$$b_{cu} = -\frac{1}{n} Ha + \frac{1}{n} E_F \left(HG_i H g_i \right) + \frac{1}{n} H E_F \left(g_i g_i' P g_i \right), \tag{2.45}$$

where $g_i \equiv g(y_i, \theta)$, $G_i \equiv \frac{\partial g_i}{\partial \theta'}$, $H \equiv \Sigma G' V^{-1}$, $P \equiv V^{-1} - V^{-1} G \Sigma G' V^{-1}$ and a is an s-vector such that $a_j \equiv \frac{1}{2} tr \left\{ \Sigma E_F \left[\frac{\partial^2 g_{ij}(\theta_o)}{\partial\theta \partial\theta'} \right] \right\}$, g_{ij} denotes the jth element of g_i, $j = 1, ..., s$. The bias (2.45) is composed by three terms, each of which has its own interpretation. Following Newey and Smith (2000), the sum of the two first terms is the bias for the (infeasible) optimal GMM estimator based on the moment vector $G'V^{-1} g(y, \theta)$, where the optimal linear combination matrix $G'V^{-1}$ does not need to be estimated. The first term arises from nonlinearity of the moments, while the second is generally nonzero whenever there is endogeneity but it should tend to be not very large. The third term is due to estimation of V in the optimal linear combination of moments. For the two-step GMM estimator Newey and Smith (2000) demonstrated that

$$b_{2s} = b_{cu} - \frac{1}{n} \Sigma E_F \left(G_i' P g_i \right) + \frac{1}{n} H \sum_{j=1}^{s} \bar{V}_{\theta_j} \left(H_S - H \right)' e_j, \tag{2.46}$$

where $\bar{V}_{\theta_j} \equiv E \left(\frac{\partial g_i g_i'}{\partial \theta_j} \right)$, $H_S \equiv \left(G' S^{-1} G \right) G' S^{-1}$ and e_j is an s-vector whose j-element is one and the others are zero. Hence, the two-step GMM estimator has two further and important sources of bias. The first arises from the necessity of estimating G in the linear combination matrix $G'V^{-1}$ for the infeasible optimal GMM estimator. The second arises from the choice of the first-step estimator, being zero if S is a scalar multiple of V. Note that for the repeatedly-iterated GMM estimator the last source of bias is not present.

There is relatively little Monte Carlo evidence on the small sample properties of the CU-GMM estimator. To the best of our knowledge, only Hansen, Heaton and Yaron (1996) and Stock and Wright (2000) have undertaken simulation studies of this estimator, obtaining similar conclusions, which indicate that the CU-GMM estimator is effectively approximately median unbiased but has a finite sample distribution with very fat tails, exhibiting sometimes extreme outlier behaviour. We investigate this question further in the next chapter.

2.5 Empirical-based estimation methods

2.5.1 Introduction

The CU-GMM is just one of several methods that can be used as an alternative to GMM for the estimation of moment condition models. Indeed, a number of other alternative estimation procedures have been recently suggested. Like CU-GMM, these new techniques produce estimators that are insensitive to how the moment conditions are scaled but, in addition, possess the advantages of not requiring a weighting matrix and of ensuring that all moment conditions, rather than only k linear combinations, are satisfied in the sample. Furthermore, as likelihood-like methods, they allow the use of classical tests to evaluate various hypotheses concerning the specification of a particular model, including overidentifying moment conditions. Conversely, their main disadvantage is computational: the system of equations requiring solution is at least twice as large as that of GMM.

In this dissertation, we concentrate on the study of two of these new EB estimation procedures, namely the EL and EI methods. Both are particular cases of the MD methods discussed by Corcoran (1998) and of the GEL method proposed by Smith (1997). Here, we follow the former approach to motivate the utilization of EB methods but employ the analytical framework provided by the latter author to present the main results concerning them.

2.5.2 Minimum discrepancy estimators

Consider again the moment conditions given in (2.1), $E_F[g(y,\theta_0)] = 0$, where the distribution $F \equiv F(y)$ is unknown. Implicitly, by giving the same weight $(\frac{1}{n})$ to each observation, GMM uses the empirical distribution function $F_n(y) \equiv \frac{1}{n}\sum_{i=1}^{n} 1(y_i \leq y)$ as an estimate for $F(y)$, where the indicator function $1(y_i \leq y)$ is equal to 1 if $y_i \leq y$ and 0 otherwise. The distribution $F_n(y)$ is the nonparametric maximum likelihood estimator of $F(y)$, being the best estimator when no information about the population of interest is available. However, because the moment conditions (2.1) are assumed to be satisfied in the population, this information can be exploited in order to obtain a more efficient estimator of $F(y)$. Thus, the idea behind MD estimation is the following: to estimate θ_0, we may select, as suggested firstly by Back and Brown (1993), the estimator $\hat{\theta}$ that minimizes the distance, relatively to some metric, between $F_n(y)$ and a distribution function $F_{md}(y)$ satisfying the moment conditions (2.1). The distribution $F_{md}(y)$ is, hence, the member of the class $\mathcal{F}(\theta)$ of all distribution functions that satisfy (2.1),

$$\mathcal{F}(\theta) \equiv \{F_{md} : E_{F_{md}}[g(y,\theta_0)] = 0\}, \tag{2.47}$$

that is closest to $F_n(y)$. As we will see below, the MD estimators $\hat{\theta}$ and $\hat{F}_{md}(y)$ are calculated simultaneously.

In the selection of a particular probability measure in $\mathcal{F}(\theta)$, different metrics for the closeness between $F_{md} \equiv F_{md}(y)$ and $F_n \equiv F_n(y)$ may be used, which gives rise to different estimation methods. Let $\mathcal{M}(F_n, F_{md})$ be the distance metric used in each method. Then, the MD estimator $\hat{\theta}$ can be described as the solution to the program

$$\min_{\theta} \mathcal{M}(F_n, F_{md}) \text{ subject to } p_i^{md} \geq 0, \ \sum_{i=1}^{n} p_i^{md} = 1 \text{ and } \sum_{i=1}^{n} p_i^{md} g(y_i,\theta) = 0, \tag{2.48}$$

where $p_i^{md} \equiv dF_{md}(y)$ denotes the probability assigned to (functions of) the i-th sample outcome, $i = 1, ..., n$. Note that the last restriction is an empirical measure counterpart to the moment conditions (2.1). Thus, all moment restrictions assumed to hold in the population are imposed numerically by appropriately reweighting the data, unlike the GMM context, where only k linear combinations of the sample moment conditions are set equal to zero.

Several estimation methods based on the program (2.48), differing only in the choice of metric $\mathcal{M}(\cdot)$, have been proposed. The most common choices for $\mathcal{M}(\cdot)$ are particular cases of the Cressie-Read power-divergence statistic [Cressie and Read (1984)].[7] In the moment condition framework, the employment of this statistic as discrepancy metric in (2.48) was suggested firstly by Imbens, Spady and Johnson (1998). The Cressie-Read statistic measures the proximity between two distribution functions F and G by

$$\mathcal{M}_{\lambda}(F,G) = \frac{1}{\lambda(1+\lambda)} \sum_{i=1}^{n} dF(y_i) \left\{ \left[\frac{dF(y_i)}{dG(y_i)} \right]^{\lambda} - 1 \right\}, \tag{2.49}$$

where λ is a fixed scalar. The most well known special cases of this measure are the empirical likelihood ($\lambda \to 0$), Kullback-Leibler ($\lambda \to -1$), Euclidean ($\lambda = -2$) and Hellinger ($\lambda = -\frac{1}{2}$) discrepancies. We present next only the first two, which are the subject of analysis throughout this dissertation and are the most widely applied in the moment condition context. In fact, the study of the estimation methods that are based on those two discrepancy metrics is now being introduced in textbooks, such as that of Mittelhammer, Judge and Miller (2000), who dedicate two autonomous chapters (12 and 13) to their analysis.

Empirical likelihood

The utilization of EL as a general statistical tool was first suggested by Owen (1988, 1990, 1991), who demonstrated that EL criterion-based statistics parallel many properties of parametric likelihood ratios; see also his recent survey textbook [Owen

[7] For a more general specification of $\mathcal{M}(\cdot)$, which includes the Cressie-Read family as a particular case, see Corcoran (1998).

(2001)]. Namely, he proved that an asymptotic chi-square distribution holds for the EL ratio, which implies that this empirical version of the likelihood ratio can be used in the same fashion as the parametric form, both for constructing confidence regions and performing hypothesis tests. Later, Qin and Lawless (1994) and Imbens (1997) extended these results to the case in which there is information available about the parameters of interest in the form of a set of moment conditions. They showed how to combine this information with the EL ratio in order to obtain consistent, asymptotically normal and efficient estimators for both the parameters and the underlying distribution of moment condition models.

The EL estimator is obtained by solving the problem (2.48) using as distance metric the EL ratio

$$ELR = -2\sum_{i=1}^{n} \ln \frac{p_i^{md}}{dF_n} \qquad (2.50)$$

or, equivalently, the Cressie-Read statistic $\mathcal{M}_0\left(F_n, F_{md}\right)$

$$\mathcal{M}_0\left(F_n, F_{md}\right) = \lim_{\lambda \to 0} \mathcal{M}_\lambda\left(F_n, F_{md}\right) = \sum_{i=1}^{n} dF_n \ln \frac{dF_n}{p_i^{md}} = \sum_{i=1}^{n} dF_n \ln \frac{dF_n}{p_i^{md}}. \qquad (2.51)$$

In both cases, since $dF_n = \frac{1}{n}$, the EL estimator may be defined as the solution to:

$$\max_{\theta} \sum_{i=1}^{n} \ln p_i^{md} \text{ subject to } p_i^{md} \geq 0,\ \sum_{i=1}^{n} p_i^{md} = 1 \text{ and } \sum_{i=1}^{n} p_i^{md} g\left(y_i, \theta\right) = 0. \qquad (2.52)$$

The problem (2.52) can be solved by optimizing the Lagrangian function[8]

$$\mathcal{L}\left(p_i^{md}, \gamma, \phi, \theta\right) = \sum_{i=1}^{n} \ln p_i^{md} - \gamma\left(\sum_{i=1}^{n} p_i^{md} - 1\right) - n\phi' \sum_{i=1}^{n} p_i^{md} g\left(y_i, \theta\right), \qquad (2.53)$$

where γ and the normalized s-vector ϕ are Lagrange multipliers. Apparently, there are $(n + k + s + 1)$ elements to be estimated but, as explained next, this difficulty can be circumvented. Indeed, solving the first-order conditions from (2.53), it follows

[8] In order to simplify the computational estimation procedures, we do not impose explicitly the restrictions $p_i^{eb} > 0$, $i = 1, ..., n$. See section 2.5.5 for an explanation of the method adopted.

that $\hat{\gamma} = n$ and

$$\hat{p}_i^{md} = \frac{\left[1 + \hat{\phi}' g\left(y_i, \hat{\theta}\right)\right]^{-1}}{\sum_{i=1}^{n}\left[1 + \hat{\phi}' g\left(y_i, \hat{\theta}\right)\right]^{-1}} = \frac{1}{n\left[1 + \hat{\phi}' g\left(y_i, \hat{\theta}\right)\right]}, \tag{2.54}$$

$i = 1, ..., n$. Using the latter result to concentrate out p_i^{md} from $\sum_{i=1}^{n} \ln p_i^{md}$ and dropping irrelevant terms, we obtain the so-called EL criterion function,

$$Q_{el}(\theta, \phi) = -\sum_{i=1}^{n} \ln\left[1 + \phi' g(y_i, \theta)\right], \tag{2.55}$$

whose optimization produces the same estimates as in (2.53) but where only $(k + s)$ parameters need to be estimated, namely the k parameters of interest θ and the s-vector of Lagrange multipliers ϕ. Note that each one of the elements of ϕ is associated with a moment indicator. Hence, a value for ϕ statistically close to zero implies that the moment conditions hold in the population.

Empirical information

Another distance metric widely used in the moment condition framework is the Kullback-Leibler information criterion (KLIC). The KLIC measures the proximity between the distribution functions $F_n(y)$ and $F_{md}(y)$ by:

$$K\left(F_{md} \left| F_n\right.\right) = E_{F_{md}}\left[\ln \frac{p_i^{md}}{dF_n}\right] = \sum_{i=1}^{n} p_i^{md} \ln \frac{p_i^{md}}{dF_n} = \lim_{\lambda \to -1} \mathcal{M}_\lambda(F_n, F_{md}) = \mathcal{M}_{-1}(F_n, F_{md}). \tag{2.56}$$

Note that this measure is not a distance in the usual sense because it is not symmetric, that is, $K(F|G) \neq K(G|F)$. However, the KLIC can be used as a discrepancy measure between two distributions because it is always nonnegative, being equal to zero if and only if $F = G$.

The estimator obtained by using (2.56) as metric in (2.48),

$$\min_{\theta} \sum_{i=1}^{n} p_i^{md} \ln p_i^{md} \text{ subject to } p_i^{md} \geq 0, \ \sum_{i=1}^{n} p_i^{md} = 1 \text{ and } \sum_{i=1}^{n} p_i^{md} g\left(y_i, \theta\right) = 0, \quad (2.57)$$

excluding irrelevant terms, is usually called the exponential tilting or EI estimator. It was firstly proposed by Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998). Note that switching the roles of F_{md} and F_n in (2.56) yields $K\left(F_n \left| F_{md}\right.\right)$, which is equal to $\mathcal{M}_0\left(F_n, F_{md}\right)$. Hence, in this sense, the EL estimator may also be interpreted as minimizing a KLIC. As observed by Imbens, Spady and Johnson (1998), the principal difference between EL and EI estimators is that the discrepancy between $F_n\left(y\right)$ and $F_{md}\left(y\right)$ is weighted by $dF_n\left(y\right)$ in the former case and by $dF_{md}\left(y\right)$ in the latter. For this reason, these authors advocate the use of the EI estimator, as $F_{md}\left(y\right)$ is a more efficient estimator of the true distribution of the data, since it takes into account of the information provided by the moment conditions. However, as we will see later on, these estimators are asymptotically first-order equivalent and, furthermore, the EL estimator seems to have more desirable higher-order properties.

The Lagrangian function for the constrained minimization of (2.57) is:

$$\mathcal{L}\left(p_i^{md}, \gamma, \phi, \theta\right) = \sum_{i=1}^{n} p_i^{md} \ln\left(p_i^{md}\right) - \gamma\left(\sum_{i=1}^{n} p_i^{md} - 1\right) - \phi' \sum_{i=1}^{n} p_i^{md} g\left(y_i, \theta\right). \quad (2.58)$$

Again, the dimension of this optimization problem can be reduced. Solving the first-order conditions from (2.58), it results that $\gamma = 1 - \ln \sum_{i=1}^{n} e^{\hat{\phi}' g\left(y_i, \hat{\theta}\right)}$ and

$$\hat{p}_i^{md} = \frac{e^{\hat{\phi}' g\left(y_i, \hat{\theta}\right)}}{\sum_{i=1}^{n} e^{\hat{\phi}' g\left(y_i, \hat{\theta}\right)}}, \quad (2.59)$$

$i = 1, ..., n$. Concentrating out p_i^{md} from $\sum_{i=1}^{n} p_i^{md} \ln\left(p_i^{md}\right)$ and omitting irrelevant terms yields the EI objective function

$$Q_{ei}\left(\theta, \phi\right) = \sum_{i=1}^{n} e^{\phi' g\left(y_i, \theta\right)}, \quad (2.60)$$

which is maximized and minimized with respect to θ and ϕ, respectively.

Before discussing the asymptotic properties of EL and EI estimators, we introduce in the next sub-section the GEL method that enables us to deal with both estimators simultaneously.

2.5.3 Generalized empirical likelihood estimation

Smith (1997) proposed alternative criteria for the estimation of moment condition models which, among others, includes as particular cases both the EL and EI methods. His approach is based on Chesher and Smith's (1997) paper, which is concerned with generating likelihood ratio test statistics for implied moment conditions in a fully parametric likelihood context by augmenting the null parametric density for the observations, $dF(y;\theta)$, by a multiplicative factor that carries a weighted sample version of the information contained in the implied moment conditions, $h[\phi' g(y,\theta)]$, where ϕ is an s-vector of auxiliary parameters. Apart normalizing constants, the augmented density is

$$r(\theta,\phi) = dF(y;\theta)\, h[\phi' g(y,\theta)], \tag{2.61}$$

where the carrier function $h(\cdot)$ is chosen such that, when $\phi = 0$, $r(\theta,\phi) = dF(y;\theta)$.

In the GMM context, however, there is no explicit knowledge of the underlying density function for the data, the only parametric information being contained in the moment conditions (2.1). Hence, Smith (1997) suggested using the empirical distribution function $F_n(y)$ and the consequent augmented function

$$r(\theta,\phi) = dF_n(y)\, h[\phi' g(y,\theta)], \tag{2.62}$$

from where we can form the semi-parametric quasi-likelihood function $Q(\theta,\phi) = -n \ln n + \sum_{i=1}^{n} \ln h[\phi' g(y_i,\theta)]$. Excluding irrelevant terms and dropping the operator 'ln' in order to simplify some of the expressions to be presented later, we obtain the equivalent criterion function

$$Q_{gel}(\theta, \phi) = \sum_{i=1}^{n} h\left[\phi' g(y_i, \theta)\right], \tag{2.63}$$

optimization of which yields the so-called GEL estimators.[9] As we can see immediately, this formulation includes as special cases the criterion functions (2.55) and (2.60): $h(\cdot)$ is equal to $-\ln\left[1 + \phi' g(y_i, \theta)\right]$ for EL estimation and to $e^{\phi' g(y_i, \theta)}$ for the EI method. Thus, from now on, we adopt the analytical framework provided by the optimization of (2.63) to present in an integrated way the main results concerning EL and EI estimators. Therefore, these results will be expressed in a very general form, being valid not only for those two estimators but for any GEL estimator, unless we explicitly mention that they were specialized for the EL and EI cases.

Before proceeding our discussion, we emphasize that GEL estimators are not always identical to MD estimators. For the chosen $h(\cdot)$ functions above, GEL estimators are indeed equal to MD estimators based on the minimization of the EL and Kullback-Leibler discrepancies. However, as discussed by Newey and Smith (2000), this equivalence occurs only when γ, the Lagrange multiplier associated with the last restriction of (2.48), can be factored out of the first-order conditions corresponding to that problem. This is possible when $h(\cdot)$ is a member of the Cressie-Read family but, for other cases, it appears impossible to do so and, hence, MD and GEL estimators are different in general. Notice that, in those cases, the MD problem will have a much larger dimension, with a $(n + k + s + 1)$-vector of parameters to estimate.

[9] All derivations and results presented throughout this thesis assume this specification for the GEL criterion function. Note that Smith (1997, 1998) uses the equivalent quasi-likelihood function above defined and Newey and Smith (2000) the normalized function $Q(\theta, \phi) = \sum_{i=1}^{n} h\left[\frac{\nabla h(0)}{\nabla^2 h(0)} \phi' g(y_i, \theta)\right]$, where $\nabla h(0)$ and $\nabla^2 h(0)$ are defined in sub-section 2.5.4. These different specifications must be taken into account when comparing the expressions derived in this dissertation with those presented in those papers.

2.5.4 Asymptotic properties of GEL estimators

In this sub-section, we discuss the asymptotic properties of GEL estimators. Taking the first derivatives of (2.63) with respect to θ and ϕ, we find that GEL estimators satisfy the conditions

$$\sum_{i=1}^{n} \pi_i\left(\hat{\theta},\hat{\phi}\right)\begin{bmatrix} g\left(y_i,\hat{\theta}\right) \\ G_i\left(\hat{\theta}\right)'\hat{\phi} \end{bmatrix} = 0, \tag{2.64}$$

with $G_i\left(\hat{\theta}\right) \equiv \frac{\partial g\left(y_i,\hat{\theta}\right)}{\partial \theta'}$ and

$$\pi_i\left(\hat{\theta},\hat{\phi}\right) \equiv \nabla h\left[\hat{\phi}' g\left(y_i,\hat{\theta}\right)\right], \tag{2.65}$$

where $\nabla h(v) \equiv \frac{\partial h(v)}{\partial v}$. In EL estimation $\pi_i\left(\hat{\theta},\hat{\phi}\right) = -\left[1+\hat{\phi}' g\left(y_i,\hat{\theta}\right)\right]^{-1}$ and for the EI method $\pi_i\left(\hat{\theta},\hat{\phi}\right) = e^{\hat{\phi}' g\left(y_i,\hat{\theta}\right)}$. Below we denote the second and third derivatives of $h(v)$ by $\nabla^2 h(v)$ and $\nabla^3 h(v)$, respectively.

The estimating equations (2.64) form a just-identified system of $(k+s)$ equations. Thus, while the efficient GMM estimator needs a two-step procedure due to the estimation of the optimal weighting matrix, thereby having its finite sample properties depending on the first step, the GEL method does not require such an initial step. As emphasized by Bera and Bilias (2000), this feature is expected to improve the small sample properties of the estimation, since the GEL approach "offers an operational way of optimally combining estimating equations".[10]

Under suitable regularity conditions (similar to those necessary for the consistency of the GMM estimator but excluding those concerning the weighting matrix), it can be proved that the estimator $\hat{\theta}$ that satisfies the system of equations (2.64) is a consistent estimator of θ_0; see *inter alia* Newey and Smith (2000) for a rigorous asymptotic analysis of the properties of GEL estimators. Moreover, expanding linearly (2.64) around $(\theta,\phi) = (\theta_0, 0)$ and using standard asymptotic theory, it can also be

[10] Note that, unlike GMM, the GEL method cannot be directly applied in the estimation of time series models. Both Kitamura and Stutzer (1997) and Smith (1997) suggest the smoothing of the observations before the optimization. See those papers for details.

demonstrated that

$$\sqrt{n}\begin{bmatrix}\hat{\phi}\\ \hat{\theta}-\theta_0\end{bmatrix} = -\begin{bmatrix}\frac{\nabla h(0)}{\nabla^2 h(0)}V^{-1}M\\ \Sigma G'V^{-1}\end{bmatrix}\sqrt{n}g_n(\theta_0)+O_p\left(n^{-\frac{1}{2}}\right), \tag{2.66}$$

where Σ was defined in (2.6),

$$M \equiv I - G\Sigma G'V^{-1} \tag{2.67}$$

is an $(s \times s)$ idempotent matrix, $\nabla h(0) = -1$ and $\nabla^2 h(0) = 1$ for EL estimators and $\nabla h(0) = \nabla^2 h(0) = 1$ in case of EI estimation. Thus, it follows that GEL estimators are asymptotically normal distributed,

$$\sqrt{n}\begin{bmatrix}\hat{\phi}\\ \hat{\theta}-\theta_0\end{bmatrix} \xrightarrow{d} N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\left[\frac{\nabla h(0)}{\nabla^2 h(0)}\right]^2 M'V^{-1}M & 0\\ 0 & \Sigma\end{pmatrix}\right], \tag{2.68}$$

and, hence, the GEL estimator of the parameters of interest is first-order equivalent to the efficient GMM estimator.

After obtaining $\hat{\theta}$ and $\hat{\phi}$, using the computational procedures discussed in the next sub-section, the implied probabilities referred to in Back and Brown (1993), previously denoted by p_i^{md} and from now on by p_i^{gel}, $i = 1, ..., n$, may be estimated using expressions (2.54) for EL and (2.59) for EI estimators. This gives rise to the same estimates as calculating

$$\hat{p}_i^{gel} \equiv p_i^{gel}\left(\hat{\theta},\hat{\phi}\right) = \frac{\pi_i\left(\hat{\theta},\hat{\phi}\right)}{\sum_{i=1}^n \pi_i\left(\hat{\theta},\hat{\phi}\right)}. \tag{2.69}$$

We can also estimate the distribution $F(y)$ in (2.1) by

$$\hat{F}_{gel}(y) \equiv F_{gel}\left(y,\hat{\theta},\hat{\phi}\right) = \sum_{i=1}^n \hat{p}_i^{gel} 1(y_i \leq y), \tag{2.70}$$

which is a more efficient estimator than $F_n(y)$ as it takes into account the information provided by the moment conditions (2.1). Indeed, assuming that

$$\sqrt{n}\left[F_n(y) - F(y)\right] \xrightarrow{d} N\left(0, \sigma^2\right), \tag{2.71}$$

it can be proved that [see Qin and Lawless (1994) and Smith (2000)]

$$\sqrt{n}\left[\hat{F}_{gel}(y) - F(y)\right] \xrightarrow{d} N\left(0, \omega^2\right), \tag{2.72}$$

where $\omega^2 = \sigma^2 - B'M'V^{-1}MB$ and $B = E_F\left[1\left(y_i \leq y\right) g\left(y_i, \theta_0\right)\right]$. As $\omega^2 < \sigma^2$ in all cases, clearly the GEL distribution $\hat{F}_{gel}(y)$ is more efficient than $F_n(y)$. Thus, the GEL distribution $\hat{F}_{gel}(y)$ can be used as alternative to $F_n(y)$ to obtain consistent estimators of statistics such as V and G, by weighting each observation i by the estimated probability $\hat{p}_i^{gel}$, $i = 1, ..., n$. Note that when the number of moment conditions is identical to the number of parameters to be estimated, the value of θ that optimizes (2.63) is the same that solves $\sum_{i=1}^{n} g\left(y_i, \hat{\theta}\right) = 0$. In this case $\hat{\phi} = 0$ and, hence, $\hat{p}_i^{gel} = \frac{1}{n}$, $\forall i = 1, ..., n$, and $\hat{F}_{gel}(y) = F_n(y)$.

Similarly to the CU-GMM estimator, Newey and Smith (2000) derived asymptotic expressions for the bias of GEL estimators:

$$b_{gel} = -\frac{1}{n} Ha + \frac{1}{n} E\left[HG_i Hg_i\right] + \frac{1}{n}(1 - \eta) HE\left[g_i g_i' P g_i\right], \tag{2.73}$$

where $\eta = \frac{\nabla h(0) \cdot \nabla^3 h(0)}{2\left[\nabla^2 h(0)\right]^2}$ is a scalar. This expression is very similar to that presented for the CU-GMM estimator in (2.45), apart from the weight $(1 - \eta)$. Hence, like the CU-GMM estimator, GEL estimators have one less source of bias than the two-step GMM estimator. Furthermore, for the EL estimator the last term of (2.73) disappears, as $\eta = 1$; for EI estimation $\eta = \frac{1}{2}$. Thus, following the interpretation of the higher-order bias terms in section 2.4, the EL estimator removes the bias due to estimation of the weighting matrix in the optimal linear combination of moments. Its bias is then the same as for the (infeasible) GMM estimator based on the optimal linear combination of moment conditions. On the other hand, under zero expectation of third powers of the moment indicators, CU-GMM and GEL estimators are higher-order asymptotically equivalent.

Unlike GMM, there are very few papers investigating the finite sample properties of GEL estimators. Although those studies show promising results, much more research is still needed which is, therefore, one of the main aims of this thesis. As we will see in the next chapter, where we assess in two different settings the small sample bias of GEL estimators, the higher-order bias presented above is very useful for explaining the results obtained.

2.5.5 Computational issues

We have seen throughout this section various advantages of GEL estimation over GMM. We now discuss what appears to be the only disadvantage of GEL estimators: their practical computation. Indeed, two main problems arise when we try to estimate moment condition models employing GEL methods. Firstly, the number of parameters to be estimated is at least twice larger: $(s+k)$ versus only k. Hence, GEL estimation is more time consuming. Secondly, and this is the main issue, the GEL criterion (2.63) is a saddle function. Therefore, either optimizing it directly or solving the system of equations (2.64) is unattractive from a computational standpoint.

One possibility is first minimize (2.63) with respect to ϕ for given θ,

$$\hat{\phi}(\theta) \equiv \arg\min_{\phi} Q_{gel}(\theta, \phi), \tag{2.74}$$

which yields the first set of conditions in (2.64), and then maximize $Q_{gel}\left[\theta, \hat{\phi}(\theta)\right]$ with respect to θ,

$$\hat{\theta} \equiv \arg\max_{\theta}\min_{\phi} Q_{gel}(\theta, \phi), \tag{2.75}$$

which produces the second set of first-order conditions in (2.64). Finally, ϕ is estimated by $\hat{\phi} = \hat{\phi}\left(\hat{\theta}\right)$. However, in our Monte Carlo experiments this procedure did not prove successful. The minimization with respect to ϕ was very quick and easy but the second step failed to converge to a solution almost all the time.

Following Imbens, Spady and Johnson (1998), in our simulation we employed their penalty approach, which worked very well. Thus, instead of directly optimizing (2.63), we opted for solving the program

$$\max_{\theta,\phi} Q_{gel}(\theta,\phi) - 0.5 \cdot A \cdot \nabla_{\phi} Q_{gel}(\theta,\phi)' \cdot W^{-1} \cdot \nabla_{\phi} Q_{gel}(\theta,\phi), \tag{2.76}$$

where A is a large scalar and W an arbitrary positive definite matrix of dimension s. For any positive definite matrix W and for finite but large enough A, the solutions to (2.63) and (2.76) are numerically identical. As in Imbens, Spady and Johnson, we choose

$$W = \nabla_{\phi\phi} Q_{gel}\left(\bar{\theta},\bar{\phi}\right) - \nabla_{\phi} Q_{gel}\left(\bar{\theta},\bar{\phi}\right) \nabla_{\phi} Q_{gel}\left(\bar{\theta},\bar{\phi}\right)', \tag{2.77}$$

where $\left(\bar{\theta},\bar{\phi}\right)$ are some initial estimates of (θ,ϕ). In all Monte Carlo simulation studies undertaken throughout this dissertation $\bar{\theta}$ is the two-step GMM estimator $\hat{\theta}$ and $\bar{\phi}$ the estimates resulting from the optimization (2.74), with θ replaced by $\hat{\theta}$. However, these choices were inessential for the results obtained, as the solution to (2.76) is insensitive to the estimates $\left(\bar{\theta},\bar{\phi}\right)$ utilized in the evaluation of W.

After calculating the GEL estimators $\left(\hat{\theta},\hat{\phi}\right)$ as described above, it is then necessary to check whether the resulting implied probabilities $\hat{p}_i^{gel} \equiv p_i^{gel}\left(\hat{\theta},\hat{\phi}\right)$, $i = 1,...,n$, calculated as in (2.69), are non-negative, since we are not imposing this restriction during the optimization procedure. If $\hat{p}_i^{gel} > 0$, $\forall i = 1,...,n$, we accept $\left(\hat{\theta},\hat{\phi}\right)$ as GEL estimators; however, in no cases did we find these constraints to be a problem.

2.5.6 Specification Tests

In this section, we discuss tests of overidentifying moment conditions, tests for additional moment conditions and tests of parametric restrictions for models estimated by GEL methods. For a detailed derivation of most of those tests see Smith (2000). Non-nested tests will only be discussed, and some alternatives proposed, in chapter 5.

Tests of overidentifying moment conditions

In the GEL framework there are several ways to assess the validity of the moment conditions assumed to hold in the population. Indeed, as a sample version of each moment condition is associated with a Lagrange multiplier, the validity of those restrictions can be analyzed by testing the hypothesis $H_0 : \phi = 0$. Hence, the three classical tests may be employed.

Qin and Lawless (1994), for EL estimators, and Kitamura and Stutzer (1997), for EI estimators, proposed distance metric statistics[11] for testing overidentifying restrictions,

$$DM_n \equiv 2\frac{\nabla^2 h(0)}{[\nabla h(0)]^2}\left[Q_{gel}\left(\tilde{\theta}, 0\right) - Q_{gel}\left(\hat{\theta}, \hat{\phi}\right)\right], \tag{2.78}$$

where $\left(\hat{\theta}, \hat{\phi}\right)$ are the GEL estimators resulting from the optimization of (2.63) and $\left(\tilde{\theta}, 0\right)$ are the GEL estimators under the null hypothesis. Note that $\tilde{\theta}$ is not identified because imposing $\phi = 0$ prevents the use of the information contained in the moment conditions. However, the non-identification of $\tilde{\theta}$ is not problematic since we know from (2.63) that, when $\phi = 0$, $Q_{gel}\left(\tilde{\theta}, 0\right)$ is equal to $\sum_{i=1}^{n} h(0)$, which is 0 and n for the EL and EI methods, respectively. Therefore, to calculate the statistic (2.78), only the estimation of the unconstrained model is required. Under the null hypothesis, the statistic DM has an asymptotic chi-square distribution with $s-k$ degrees of freedom.

In the GEL context, Smith (1997) proposed testing H_0 employing Wald or score tests. The Wald test statistic for H_0 is

$$W_n \equiv n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2 \hat{\phi}'\hat{V}_n\hat{\phi}, \tag{2.79}$$

and the score test statistic is

$$LM_n \equiv n\hat{g}_n'\hat{V}_n^{-1}\hat{g}_n. \tag{2.80}$$

[11] Note that only in the case of EL does this statistic correspond to a likelihood ratio. Therefore, we adopted the distance metric term, as it is valid for both EL and EI estimation.

Note that the latter statistic has exactly the same expression as the J test presented in (2.14) and that it is not evaluated under the null hypothesis. Again, the problem is the non-identification of $\tilde{\theta}$ under H_0. Thus, Smith (1997) suggested the replacement of $\tilde{\theta}$ by $\hat{\theta}$, which may be regarded as a least favorable choice of estimator for θ_0. Under the null hypothesis, both statistics have a limiting chi-squared distribution with $s-k$ degrees of freedom.

In chapter 4, we propose a new class of test statistics for overidentifying moment conditions appropriate for models estimated by GEL methods.

Tests for additional moment conditions

Both the J_2 and CM tests described for GMM (see expressions 2.15 and 2.16) may also be utilized to test the validity of further moment conditions in the GEL framework. Additionally, following Smith (1997), we may employ classical tests as well. To this end, we need to incorporate a sample version of the new moment conditions in the GEL criterion (2.63),

$$Q_{gel}^{*}(\theta, \phi, \psi) \equiv \sum_{i=1}^{n} h\left[\phi' g_1(y_i, \theta) + \psi' g_2(y_i, \theta)\right], \tag{2.81}$$

where $g_1(\cdot)$ represents the original s_1 moment conditions, $g_2(\cdot)$ is the s_2-vector of additional moment restrictions and ψ is the corresponding s_2-vector of Lagrange multipliers.

The parameters contained in (2.81) may be estimated in a similar manner to that described above for standard GEL estimators. Denote such estimators by $\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right)$. It can be shown that those estimators satisfy the set of first-order conditions

$$\sum_{i=1}^{n} \pi_i^{*}\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right) \begin{bmatrix} g_1\left(y_i, \tilde{\theta}\right) \\ g_2\left(y_i, \tilde{\theta}\right) \\ G_{1i}\left(\tilde{\theta}\right)' \tilde{\phi} + G_{2i}\left(\tilde{\theta}\right)' \tilde{\psi} \end{bmatrix} = 0, \tag{2.82}$$

where $\pi_i^*\left(\tilde{\theta},\tilde{\phi},\tilde{\psi}\right) \equiv \nabla h\left[\tilde{\phi}' g_1\left(y_i,\tilde{\theta}\right) + \tilde{\psi}' g_2\left(y_i,\tilde{\theta}\right)\right]$ and $G_{ji}\left(\tilde{\theta}\right) \equiv \frac{\partial g_j\left(y_i,\tilde{\theta}\right)}{\partial \theta'}$, $j = 1, 2$, $i = 1, ..., n$.

In this setting, we may test $E_F\left[g_2\left(y,\theta\right)\right] = 0$ by assessing the parametric hypothesis $H_0 : \psi = 0$. Noting that $Q_{gel}^*\left(\hat{\theta},\hat{\phi},0\right) = Q_{gel}\left(\hat{\theta},\hat{\phi}\right)$, a distance metric statistic for this hypothesis is

$$DM_n \equiv 2\frac{\nabla^2 h\left(0\right)}{\left[\nabla h\left(0\right)\right]^2}\left[Q_{gel}\left(\hat{\theta},\hat{\phi}\right) - Q_{gel}^*\left(\tilde{\theta},\tilde{\phi},\tilde{\psi}\right)\right], \tag{2.83}$$

which has a limiting chi-squared distribution with s_2 degrees of freedom under the null hypothesis. This statistic has a similar interpretation to that of the J_2 statistic, as both statistics correspond to the difference between statistics for testing all the moment conditions, J and $DM_n \equiv 2\frac{\nabla^2 h(0)}{[\nabla h(0)]^2}\left[Q_{gel}^*\left(\breve{\theta},0,0\right) - Q_{gel}^*\left(\tilde{\theta},\tilde{\phi},\tilde{\psi}\right)\right]$, and for assessing only the first s_1 conditions, J_1 and $DM_n \equiv 2\frac{\nabla^2 h(0)}{[\nabla h(0)]^2}\left[Q_{gel}\left(\breve{\theta},0\right) - Q_{gel}\left(\hat{\theta},\hat{\phi}\right)\right]$.

A Wald test statistic for $H_0 : \psi = 0$ may be defined as

$$W_n \equiv n\left[\frac{\nabla^2 h\left(0\right)}{\nabla h\left(0\right)}\right]^2 \tilde{\psi}'\left(S_\psi' \tilde{M}_n' \tilde{V}_n^{-1} \tilde{M}_n S_\psi\right)^{-1} \tilde{\psi}, \tag{2.84}$$

where $S_\psi' = \begin{bmatrix} 0 & I_{s_2} \end{bmatrix}$ is a $(s_2 \times s)$ selection matrix and $\tilde{M}_n$ is a consistent estimator of the matrix M defined in (2.67). This statistic has also an asymptotic chi-squared distribution with s_2 degrees of freedom under the null hypothesis.

Finally, an asymptotic equivalent score statistic for testing H_0 is

$$LM_n \equiv \frac{1}{n}\frac{1}{\left[\nabla h\left(0\right)\right]^2}\sum_{i=1}^{n}\pi_i\left(\hat{\theta},\hat{\phi}\right) g_2'\left(y_i,\hat{\theta}\right) S_\psi' \hat{M}_n' \hat{V}_n^{-1} \hat{M}_n S_\psi \sum_{i=1}^{n}\pi_i\left(\hat{\theta},\hat{\phi}\right) g_2'\left(y_i,\hat{\theta}\right). \tag{2.85}$$

Tests of parametric restrictions

Test statistics to assess parametric restrictions in models estimated by GEL methods were presented in Qin and Lawless (1995), Kitamura and Stutzer (1997) and Smith (1997, 2000). As before (see section 2.3.4, which concerned this kind of tests in the

GMM framework), consider the null hypothesis $H_0 : r(\theta_0) = 0$, where $r(\cdot)$ is a known continuously differentiable q-vector, q being the number of restrictions, and denote $\frac{\partial r(\theta)}{\partial \theta'}$ by $R(\theta)$, a $(q \times k)$ matrix of rank q. Following Smith (1997), the constrained model incorporating H_0 may be estimated by optimizing the modified GEL function

$$Q^*_{gel}(\theta, \phi, \psi) = h[\phi' g(y_i, \theta) + \psi' r(\theta)]. \tag{2.86}$$

The resultant estimators, $\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right)$, satisfy the first-order conditions

$$\sum_{i=1}^{n} \pi_i^* \left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right) \begin{bmatrix} g\left(y_i, \tilde{\theta}\right) \\ r\left(\tilde{\theta}\right) \\ G_i\left(\tilde{\theta}\right)' \tilde{\phi} + R\left(\tilde{\theta}\right)' \tilde{\psi} \end{bmatrix} = 0, \tag{2.87}$$

where $\pi_i^* \left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right) \equiv \nabla h \left[\tilde{\phi}' g\left(y_i, \tilde{\theta}\right) + \tilde{\psi}' r\left(\tilde{\theta}\right)\right]$.

Using standard asymptotic theory, it is easy to derive the limiting distribution of $\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right)$ and, then, deduce expressions for the classical tests to assess H_0. As demonstrated by Smith (1997),

$$\sqrt{n} \begin{bmatrix} \tilde{\phi} \\ \tilde{\psi} \\ \tilde{\theta} - \theta_0 \end{bmatrix} = - \begin{bmatrix} \frac{\nabla h(0)}{\nabla^2 h(0)} \left(V^{-1} - V^{-1} G \Sigma P G' V^{-1}\right) \\ -\frac{\nabla h(0)}{\nabla^2 h(0)} \left(R \Sigma R'\right)^{-1} R \Sigma G' V^{-1} \\ \Sigma P G' V^{-1} \end{bmatrix} \sqrt{n} g_n(\theta_0) + O_p\left(n^{-\frac{1}{2}}\right), \tag{2.88}$$

and, hence,

$$\sqrt{n} \begin{bmatrix} \tilde{\phi} \\ \tilde{\psi} \\ \tilde{\theta} - \theta_0 \end{bmatrix} \xrightarrow{d} N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \left[\frac{\nabla h(0)}{\nabla^2 h(0)}\right]^2 \left(V^{-1} - V^{-1} G \Sigma P G' V^{-1}\right) & -\left[\frac{\nabla h(0)}{\nabla^2 h(0)}\right]^2 V^{-1} G \Sigma R' \left(R \Sigma R'\right)^{-1} & 0 \\ -\left[\frac{\nabla h(0)}{\nabla^2 h(0)}\right]^2 \left(R \Sigma R'\right)^{-1} R \Sigma G' V^{-1} & \left[\frac{\nabla h(0)}{\nabla^2 h(0)}\right]^2 \left(R \Sigma R'\right)^{-1} & 0 \\ 0 & 0 & \Sigma P \end{pmatrix} \right], \tag{2.89}$$

where $R \equiv R(\theta_0)$ and $P = I - R'(R\Sigma R')^{-1} R\Sigma$ is an $(k \times k)$ idempotent matrix of rank q. From (2.89), the following score statistic for testing H_0 can be derived:

$$LM_n \equiv n \left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2 \tilde{\psi}' \tilde{R} \tilde{\Sigma}_n \tilde{R}' \tilde{\psi}. \tag{2.90}$$

This expression has not been discussed previously in this chapter but it is just another form of (2.19), the score statistic for testing parametric restrictions in the GMM framework. Likewise, it is straightforward to see that the Wald statistic (2.18) presented previously for GMM estimators is also appropriate for GEL estimators. As for the DM statistic for H_0, its expression is now

$$DM_n \equiv 2 \frac{\nabla^2 h(0)}{[\nabla h(0)]^2} \left[Q_{gel}\left(\hat{\theta}, \hat{\phi}\right) - Q_{gel}\left(\tilde{\theta}, \tilde{\phi}\right)\right], \tag{2.91}$$

since $Q^*_{gel}\left(\hat{\theta}, \hat{\phi}, 0\right) = Q_{gel}\left(\hat{\theta}, \hat{\phi}\right)$ and $Q^*_{gel}\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right) = Q_{gel}\left(\tilde{\theta}, \tilde{\phi}\right)$. All test statistics have an asymptotic chi-squared distribution with q degrees of freedom.

Smith (2000) derived also Hausman and minimum chi-squared tests, which can be based on the contrasts $\sqrt{n}\left(\hat{\theta} - \tilde{\theta}\right)$ or $\sqrt{n}\left(\hat{\phi} - \tilde{\phi}\right)$. Using the first contrast, as one could expect from the comments above, identical statistics to those found for GMM estimators are obtained. Indeed, from (2.66) and (2.88), it follows that

$$\sqrt{n}\left(\hat{\theta} - \tilde{\theta}\right) = -\Sigma(I - P) G' V^{-1} \sqrt{n} g_n(\theta_0) + O_p\left(n^{-\frac{1}{2}}\right) \tag{2.92}$$

and, hence,

$$\sqrt{n}\left(\hat{\theta} - \tilde{\theta}\right) \xrightarrow{d} N\left[0, \Sigma(I - P)\right], \tag{2.93}$$

since $\Sigma(I - P)\Sigma^{-1}(I - P)'\Sigma = \Sigma(I - P)$. Noting that $\Sigma(I - P) = var\left(\sqrt{n}\hat{\theta}\right) - var\left(\sqrt{n}\tilde{\theta}\right)$ and that $\hat{\theta}$ is a consistent estimator under both the alternatives, while $\tilde{\theta}$ is consistent only under H_0, in which case it is more efficient than $\hat{\theta}$, then $n\left(\hat{\theta} - \tilde{\theta}\right)'$ $\left[\Sigma(I - P)\right]^{-}\left(\hat{\theta} - \tilde{\theta}\right)$ is a Hausman test statistic. A generalized inverse for $\Sigma(I - P)$ is $R'(R\Sigma R')^{-1} R$, so the statistic H_n (2.22) is obtained. Another generalized inverse for

$\Sigma(I-P)$ is Σ^{-1}. As this is the inverse of the variance of $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$, its utilization gives rise to the minimum chi-squared statistic MC_n (2.21).

In the case of the second contrast, it follows from (2.66) and (2.88) that

$$\sqrt{n}\left(\hat{\phi}-\tilde{\phi}\right)=\frac{\nabla h(0)}{\nabla^2 h(0)}V^{-1}G\Sigma(I-P)G'V^{-1}\sqrt{n}g_n(\theta_0)+O_p\left(n^{-\frac{1}{2}}\right) \tag{2.94}$$

and, hence,

$$\sqrt{n}\left(\hat{\phi}-\tilde{\phi}\right)\xrightarrow{d}N\left\{0,\left[\frac{\nabla h(0)}{\nabla^2 h(0)}\right]^2 V^{-1}G\Sigma(I-P)G'V^{-1}\right\}. \tag{2.95}$$

Again, $\left[\frac{\nabla h(0)}{\nabla^2 h(0)}\right]^2 V^{-1}G\Sigma(I-P)G'V^{-1}=var\left(\sqrt{n}\tilde{\phi}\right)-var\left(\sqrt{n}\hat{\phi}\right)$. Therefore, using $G\Sigma G'$ as a generalized inverse for $V^{-1}G\Sigma(I-P)G'V^{-1}$, we obtain the Hausman test statistic

$$H_n^{\phi}\equiv n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2\left(\hat{\phi}-\tilde{\phi}\right)'G\Sigma G'\left(\hat{\phi}-\tilde{\phi}\right). \tag{2.96}$$

Using the alternative generalized inverse V, we obtain the minimum chi-squared test statistic

$$MC_n^{\phi}\equiv n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2\left(\hat{\phi}-\tilde{\phi}\right)'V\left(\hat{\phi}-\tilde{\phi}\right), \tag{2.97}$$

as V is also a generalized inverse for the variance of $\sqrt{n}\hat{\phi}$. Both (2.96) and (2.97) have a limiting chi-square distribution with $rk(P)=q$ degrees of freedom. Estimators for the matrices G and V may be evaluated at either $\hat{\theta}$ or $\tilde{\theta}$.

In chapter 4, we propose a Pearson-type statistic appropriate for testing parametric restrictions in models estimated by GEL methods.

Chapter 3

GMM, CU-GMM, EI, EL and bootstrap GMM estimators: small sample evidence

3.1 Introduction

The finite sample properties of the efficient two-step GMM estimator have been the subject of intensive investigation recently. As emphasized in the previous chapter, a number of Monte Carlo simulation studies have concluded that the performance of this estimator and related statistics in small samples differs significantly from that predicted by large sample theory (see, for example, the July 1996 special issue of the *Journal of Business & Economic Statistics* or the other references cited in section 2.3.6). This problem has motivated the search for alternative efficient estimators with better small sample properties, which may be divided into two main classes. The first class contains alternative procedures which are asymptotically first-order equivalent to efficient two-step GMM estimation such as CU-GMM, EI and EL. On the other hand, the possibility of improving the finite sample properties of the two-step GMM estimator using bootstrap techniques was addressed by both Hall and Horowitz (1996) and Brown, Newey and May (1997), whose proposals form the second set of alternative methods.

While there is substantial evidence of the poor small sample properties of the two-step GMM estimator, there are very few studies examining the properties of the other methods applicable in the moment condition framework. Concentrating on studies investigating the bias of parameter estimators for moment condition models, the main focus of this chapter, Hansen, Heaton and Yaron (1996) analyzed the finite sample performance of their CU-GMM estimator, Imbens (1997) examined the behaviour of EL estimators, and Horowitz (1998) considered Hall and Horowitz's (1996) bootstrap GMM estimators. To the best of our knowledge, no other papers have examined this issue. Thus, although all these studies reported promising results, further investigation is still needed in order to assess the ability of those and other alternative methods to produce improved estimators for the parameters of moment condition models.

In this chapter we undertake two simulation studies examining the finite sample properties of three methods that are asymptotically first-order equivalent to GMM and six alternative bootstrap techniques in two different settings for which there is previous evidence of the poor performance of efficient GMM estimators. With regard to the former methods, we investigate the small sample bias of CU-GMM, EL and EI estimators. All these methods have already been described in the previous chapter (see sections 2.4 and 2.5). As will be seen, the theoretical findings by Newey and Smith (2000), who analysed the higher-order properties of these estimators, will be crucial in the justification of the results obtained in the Monte Carlo experiments.

In our consideration of the bootstrap methods, we consider three techniques already applied in the moment condition framework and suggest three new ones. The most commonly applied bootstrap, the so-called nonparametric (NP) bootstrap, is expected to fail in producing substantial reductions in the bias of GMM estimators. Indeed, it attempts to approximate the distribution of the data making use of the fact that the empirical distribution function is similar to the true data generating process. However, in the overidentified moment condition framework, the moment restrictions are not satisfied in the sample, so the NP bootstrap does not take them into account and, hence, does not mimic correctly the underlying distribution of the data. Hall and Horowitz (1996) and Brown, Newey and May (1997) proposed two alternative bootstrap methods that deal with this issue. The former authors suggested the recentered nonparametric (RNP) bootstrap, which still employs the empirical distribution function to resample the data but recenters the moment indicators at their sample values. Alternatively, Brown, Newey and May (1997) proposed what we call here the first-stage GEL (FSGEL) bootstrap, which generates the bootstrap samples using a distribution that imposes the moment conditions on the original sample.

All bootstrap methods that we propose in this chapter are based on the GEL distribution, the main motivation for this choice being the fact that this is a more efficient estimator of the distribution of the data than the two used by the existing methods (see section 2.5.4). We first consider direct application of the GEL bootstrap. However, it suffers from the same problems as the NP bootstrap because, while it does impose the moment conditions on the sample when considering GEL estimators, when applied to correct the bias of GMM estimators only asymptotically are those restrictions satisfied. Thus, we suggest two modified versions of the GEL bootstrap: the recentered GEL (RGEL) bootstrap, which recenters the moment indicators in an analogous manner to the RNP bootstrap; and the post-hoc GEL (PHGEL) bootstrap, which introduces a post-sample adjustment in the calculation of the bias of the GMM estimator.

This chapter is organized as follows. Section 3.2 discusses the general principles of bootstrap methods, showing how to use them to eliminate the bias of parameter estimators. Section 3.3 describes the major characteristics of the various bootstrap methods applicable in the GMM framework. A first Monte Carlo study, for covariance structure models, is presented in section 3.4. Section 3.5 considers another Monte Carlo study, for instrumental variable models. Section 3.6 concludes.

3.2 Bias-corrected GMM estimators

The progress in computer technology in the last two decades stimulated the development of computer-intensive statistical methods. One of the methods that benefited from the increasing availability of inexpensive, powerful and fast computing was the bootstrap, introduced by Efron (1979). The main appeal of this technique is its simplicity, the theoretical derivations required in traditional methods (such as obtaining derivatives, the form of the asymptotic variance, calculating explicit expressions for the bias of an estimator, etc.) being replaced by repeatedly resampling the data and making inference from the resamples. As example of its increasing popularity several books dedicated to bootstrap techniques have been published in the last ten years, for example Hall (1992), Efron and Tibshirani (1993), Shao and Tu (1995), Davison and Hinkley (1997) and Chernick (1999).

Basically, the bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling the original data set, which is treated as though it was the population. If the data were, in fact, the population, then the bias of an estimator or test statistic could be computed with arbitrary accuracy by repeatedly resampling the data. Since the data are not the population, the bootstrap provides only an approximation to the distribution of statistics that, however, turns out to be often more accurate than the approximation obtained from first-order asymptotic distribution theory. See, for instance, Hall and Horowitz (1996), Horowitz (1998) and Ziliak (1997), for examples where the bootstrap is shown to improve significantly inference from models estimated by GMM.

Assume that a random sample S of size n is collected from a population whose (unknown) cumulative distribution function is $F(y)$. Bootstrap samples are generated by randomly sampling the original data with replacement. This resampling is based on a certain cumulative distribution function, $F^*(y)$, which assigns each observation a given probability of being sampled. For each bootstrap sample S_j^*, $j = 1, ..., B$, we calculate a statistic of interest (e.g. a parameter estimator, a standard deviation, a test statistic), obtaining thus B observations of those statistics from which measures of interest can be computed (e.g. bias, variability of estimators, improved critical values for tests). Here, we focus exclusively on the utilization of bootstrap techniques to obtain bias-corrected GMM estimators.

Consider again the moment conditions

$$E_F\left[g\left(y,\theta_0\right)\right]=0. \tag{3.1}$$

As discussed in the previous chapter (see section 2.3), the efficient GMM estimator $\hat{\theta}$ is obtained from the minimization of the optimal quadratic form of the sample moment indicators,

$$\hat{\theta}\equiv\arg\min_{\theta} g_n\left(\theta\right)\hat{V}_n^{-1}g_n\left(\theta\right), \tag{3.2}$$

its bias being given by:

$$b\left(\theta_0\right)=E_F\left(\hat{\theta}-\theta_0\right). \tag{3.3}$$

If we are able to estimate $b\equiv b\left(\theta_0\right)$, we can estimate a bias-corrected GMM estimator $\tilde{\theta}$ by calculating

$$\tilde{\theta}=\hat{\theta}-\hat{b}, \tag{3.4}$$

where $\hat{b}$ denotes the estimated bias. Instead of deriving an analytic expression for the bias function[1] and then evaluating it at the GMM estimator [or using it to correct the first-order conditions defining GMM estimators - see Firth (1993)], we can simply use the bootstrap and estimate the bias (3.3) as follows:

1. Compute $\hat{\theta}$ accordingly to (3.2) using the original data;

2. Generate B bootstrap samples S_j^*, $j=1,...,B$, of size n by sampling the original data randomly with replacement accordingly with the chosen distribution function $F^*\left(y\right)$:

$$S_j^*=\left\{y_{j1}^*,...,y_{jn}^*\right\},$$

[1]Newey and Smith (2000) provide such bias functions for both GMM and GEL estimators, as referred to in the previous chapter.

where y_{ji}^*, $i = 1, ..., n$, denotes the observations included in the bootstrap sample S_j^*;

3. For each bootstrap sample calculate the GMM estimator $\hat{\theta}_j^*$:

$$\hat{\theta}_j^* \equiv \arg\min_\theta g_{jn}^*(\theta) \hat{V}_{jn}^{*-1} g_{jn}^*(\theta),$$

$j = 1, ...B$, where $g_{jn}^*(\theta) = \frac{1}{n}\sum_{i=1}^n g\left(y_{ji}^*, \theta\right)$ and $\hat{V}_{jn}^{*-1}$ is obtained using estimators from the bootstrap sample S_j^*;

4. Average the B GMM estimators calculated in the preceding step:

$$\bar{\theta}^* = \frac{1}{B}\sum_{j=1}^B \hat{\theta}_j^*;$$

5. Estimate the bias of the GMM estimator $\hat{\theta}$ by calculating:

$$\hat{b} = \bar{\theta}^* - \hat{\theta}. \tag{3.5}$$

Subtracting the bias (3.3) from the GMM estimator $\hat{\theta}$, it is then possible to obtain the bias-corrected GMM estimator defined in (3.4):

$$\tilde{\theta} = 2\hat{\theta} - \bar{\theta}^*. \tag{3.6}$$

This general procedure to obtain bootstrap estimators may be implemented in several distinct forms, as discussed in the next section.

3.3 Alternative bootstrap GMM estimators

In this section we discuss six alternative procedures for obtaining bootstrap GMM estimators, two of which are expected to fail in reducing significantly the bias of the

GMM estimator. We discuss first the three existing methods and then present our three proposals.

3.3.1 Nonparametric bootstrap

Until now nothing was said about the choice of the distribution $F^*(y)$ from which bootstrap samples are generated. In the most commonly applied bootstrap, the so-called NP bootstrap, the resampling is based on the empirical distribution function $F_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1(y_i \leq y)$, so each observation has equal probability $\frac{1}{n}$ of being drawn. Denote by $p^* \equiv dF^*(y) \equiv (p_1^*, ..., p_n^*)$ the n-dimensional resampling vector that assigns each observation a given probability of being sampled:

$$p^* = \left(\frac{1}{n}, ..., \frac{1}{n}\right). \tag{3.7}$$

Direct application of the NP bootstrap in the GMM framework seems to be unsatisfactory in many cases, though. When the model is overidentified, while the population moment conditions $E_F[g(y,\theta)] = 0$ are satisfied at $\theta = \theta_0$, the estimated sample moments are typically not zero, that is, there is no θ such that $E_{F_n}[g(y,\theta)] = 0$ is met, except in very special cases. Therefore, the empirical distribution function may be a poor approximation to the true underlying distribution of the data and, hence, the NP bootstrap may not yield a substantial improvement over first-order asymptotic theory in standard applications of GMM.

This problem is particularly serious for the J test of overidentifying moment conditions. As Brown, Newey and May (1997) argued, bootstrapping from the empirical distribution will produce, even asymptotically, a wrong size for that test. This failure results from the fact that, instead of imposing the null hypothesis (3.1), the empirical distribution corresponds to an alternative hypothesis where the moment conditions (3.1) do not hold. An empirical example using NP bootstrap GMM estimators by Ziliak (1997) confirmed that, in fact, inference based on the NP bootstrap J test is severely distorted. In chapter 4 we present additional empirical evidence on the poor performance of this test in finite samples.

On the other hand, Hahn (1996) demonstrated theoretically that the NP bootstrap distribution of any GMM estimator converges weakly to the limit distribution of the estimator. According to this author, the arguments against the use of the NP bootstrap in the moment condition context apply to the J test, not to the GMM estimator. Hence, we decided to include the analysis of the bias of the NP bootstrap GMM estimator in the two Monte Carlo experiments that we conduct in sections 3.4 and 3.5, investigating whether or not it behaves better than simple GMM estimators in finite samples and how it performs comparatively with the more refined bootstrap methods discussed below.

3.3.2 Recentered nonparametric bootstrap

As discussed above, the doubts concerning the efficacy of applying the NP bootstrap in the GMM framework arise from the fact that there is no θ such that $E_{F_n}\left[g\left(y,\theta\right)\right]=0$ is met. Thus, the key factor to successful application of bootstrap techniques in the GMM context seems to require the satisfaction of a bootstrap version of the population moment conditions. There are two alternative ways to deal with this question. One implies looking for a different resampling distribution, say $F_1\left(y\right)$, such that $E_{F_1}\left[g\left(y,\theta\right)\right]=0$ for $\theta=\hat{\theta}_{2s}$, the two-step GMM estimator. This hypothesis will be discussed in the next sub-section. The other alternative was proposed by Hall and Horowitz (1996), who suggested keeping $F_n\left(y\right)$ as the resampling distribution and, instead, recentering the moment indicators as follows:

$$E_{F_n}\left[g^c\left(y_j^*,\theta\right)\right]=0, \tag{3.8}$$

where

$$g^c\left(y_j^*,\theta\right) \quad = \quad g\left(y_j^*,\theta\right)-E_{F_n}\left[g\left(y,\hat{\theta}_{2s}\right)\right]$$

$$= g\left(y_j^*, \theta\right) - \frac{1}{n}\sum_{i=1}^{n} g\left(y_i, \hat{\theta}_{2s}\right), \tag{3.9}$$

$j = 1, ..., B$. Now, clearly, the expectation of the modified moment indicators $g^c(\cdot)$ with respect to the empirical distribution is zero.

The implementation of this RNP bootstrap method follows very closely the general procedures described in sub-section 3.2. The only difference is in the way bootstrap GMM estimators are calculated in step 3, their estimation now being based on the recentered moment indicators (3.9). If we evaluate the required weighting matrix at a non-efficient GMM estimator, corrected moment indicators constructed in an analogous way to (3.9) must be used in its calculation.

Monte Carlo evidence reported by Horowitz (1998) indicates that the RNP bootstrap attenuates the bias of GMM estimators, although it has not been completely eliminated in all the cases considered by him. In section 3.4, in our first Monte Carlo study, we consider a similar experimental design in order to assess the performance of this method relative to the other bootstrap techniques discussed in this chapter.

3.3.3 First-stage GEL bootstrap

Another modification to the NP bootstrap was suggested by Brown, Newey and May (1997). Instead of using the NP distribution to resample the original data, they propose the employment of a distribution that, by assigning to each observation a different weight, imposes the moment conditions, evaluated at the two-step GMM estimator $\hat{\theta}_{2s}$, on the sample. Such a distribution is obtained from the first step of one of the estimation procedures described in section 2.5.5 appropriate for the calculation of GEL estimators, so we call it here the first-stage GEL (FSGEL) cumulative distribution.

The procedures necessary to estimate this distribution are the following. First, the GEL objective function $Q_{gel}(\theta, \phi)$, given in (2.63), is minimized only with respect to the Lagrange multipliers ϕ, keeping $\theta = \hat{\theta}_{2s}$. Then, the resulting estimators, $\hat{\phi}_{fsgel}$, are used to obtain the FSGEL distribution $F_{fsgel}(y) = \sum_{i=1}^{n} \hat{p}_i^{fsgel} 1(y_i \leq y)$, where the probabilities $\hat{p}_i^{fsgel} \equiv p_i^{gel}\left(\hat{\theta}_{2s}, \hat{\phi}_{fsgel}\right)$ are calculated as in (2.69). Since $\sum_{i=1}^{n} \hat{p}_i^{fsgel}$ $g\left(y_i, \hat{\theta}_{2s}\right) = 0$ is the first-order condition of the FSGEL optimization problem, it follows that the moment conditions are in fact imposed on the sample.

The FSGEL bootstrap, based on the resampling vector

$$p^* = \left(\hat{p}_1^{fsgel}, ..., \hat{p}_n^{fsgel}\right), \tag{3.10}$$

is asymptotically efficient relative to any bootstrap based on the empirical distribution function, as shown by Brown, Newey and May (1997). They reported promising Monte Carlo results concerning the J test of overidentifying moment conditions, which showed that the FSGEL bootstrap provides a good improvement over both asymptotic first-order theory and the NP bootstrap in this case. Similar empirical evidence is presented by Ziliak (1997). The capacity of the FSGEL bootstrap to produce bias-corrected GMM estimators is investigated later in this chapter.

3.3.4 GEL bootstrap

In the previous methods, the bootstrap samples are drawn accordingly to the empirical distribution $F_n(y)$ or the FSGEL distribution $F_{fsgel}(y)$. However, if a more efficient estimator of $F(y)$ is available, in principle bootstrap inference can be improved. For example, if the true distribution of the data was known up to the parameter θ, say $F_\theta(y)$, the so-called parametric bootstrap, where resampling is based on $F_{\hat{\theta}}(y)$, could be applied. This is not possible in the GMM framework without making additional assumptions. Nevertheless, in the construction of the resampling vector p^*, the special nature of the data can be taken into account, namely the information provided by the moment conditions, which is what the FSGEL bootstrap partially achieves. Thus, all bootstrap methods that we propose in this and the next two sub-sections are based on the GEL distribution $F_{gel}(y) = \sum_{i=1}^{n} \hat{p}_i^{gel} 1(y_i \leq y)$, where $\hat{p}_i^{gel} \equiv p_i^{gel}\left(\hat{\theta}_{gel}, \hat{\phi}_{gel}\right)$ denotes the estimated GEL implied probabilities and $\left(\hat{\theta}_{gel}, \hat{\phi}_{gel}\right)$ are GEL estimators;

see section 2.5.4.

This GEL bootstrap, based on the resampling vector

$$p^* = \left(\hat{p}_1^{gel}, ..., \hat{p}_n^{gel}\right), \tag{3.11}$$

could be directly applied to improve the finite sample properties of GEL estimators, without any modifications. In fact, in this case, the moment indicators would not need to be recentered because the moment conditions (3.1) are imposed on the data by giving different weights to different data points: $\sum_{i=1}^{n} \hat{p}_i^{gel} g\left(y_i, \hat{\theta}_{gel}\right) = 0$. However, our objective in this chapter is the analysis of the ability of bootstrap methods to reduce the bias of the two-step GMM estimator. In this case, some correction is still necessary, since in finite samples $\sum_{i=1}^{n} \hat{p}_i^{gel} g\left(y_i, \hat{\theta}_{2s}\right) \neq 0$. Thus, the next two subsections discuss two alternative procedures that adapt this bootstrap method to the GMM case.

3.3.5 Recentered GEL bootstrap

The first modified GEL bootstrap that we suggest is very simple. Analogously to Hall and Horowitz (1996), the moment indicators can be recentered as follows:

$$E_{F_{gel}}\left[g^c\left(y_j^*, \theta\right)\right] = 0, \tag{3.12}$$

where

$$\begin{aligned} g^c\left(y_j^*, \theta\right) &= g\left(y_j^*, \theta\right) - E_{F_{gel}}\left[g\left(y, \hat{\theta}_{2s}\right)\right] \\ &= g\left(y_j^*, \theta\right) - \sum_{i=1}^{n} \hat{p}_i^{gel} g\left(y_i, \hat{\theta}_{2s}\right), \end{aligned} \tag{3.13}$$

$j = 1, ..., B$. The expectation of the corrected moment indicators $g^c(\cdot)$ taken with respect to the distribution $F_{gel}(y)$ is zero.

This RGEL bootstrap can be implemented applying similar procedures to those described for the RNP method, with two alterations: $F_{gel}(y)$ is used instead of $F_n(y)$ to generate the bootstrap samples and the calculation of bootstrap GMM estimators is based on the recentered moment indicators (3.13) instead of (3.9).

3.3.6 Post-hoc GEL bootstrap

Another explanation for the expected failure of the GEL bootstrap to provide less biased GMM estimators is the following. By using the resampling vector (3.11) and estimating the bias utilizing the standard formula given in (3.5), $\hat{b} = \bar{\theta}^* - \hat{\theta}_{2s}$, we are not adequately estimating the bias of the GMM estimator $\hat{\theta}_{2s}$ that we intended to correct. Actually, in the calculation of the bias, we are comparing GMM estimators that can be based on quite distinct samples: while $\hat{\theta}_{2s}$ results from the minimization of the quadratic form (3.2), $\bar{\theta}^*$ is the average of the standard GMM estimators $\hat{\theta}_j$, $j = 1, ..., B$, each of which, due to the way the bootstrap samples are constructed, can be interpreted as minimizing also (3.2) but with $g_n(\theta)$ replaced by $g_p(\theta) \equiv \sum_{i=1}^{n} \hat{p}_i^{gel} g(y_i, \theta)$. In small samples, $g_n(\theta)$ and $g_p(\theta)$ can be rather different. Therefore, in this subsection, we propose a slight modification to the GEL bootstrap method in order to improve the approximations to bias. We suggest the utilization of a post-sampling adjustment to GEL bootstrap GMM estimators in a similar way to that considered by Efron (1990) in another context and with different objectives (he proposed a post-hoc bootstrap with the aim of reducing the number of bootstrap samples needed to obtain reliable statistics and improved estimates of the bias, keeping the usual bootstrap sampling).

Define

$$p_j^a \equiv \left(p_{j1}^a, ..., p_{jn}^a\right), \tag{3.14}$$

$j = 1, ..., B$, as the actual or post-resampling vector calculated from the bootstrap sample S_j^*, that is,

$$p_{ji}^a = \frac{\#\left\{y_{ji}^* = y_i\right\}}{n}, \tag{3.15}$$

$j = 1, ..., B$, $i = 1, ..., n$, is the proportion of times that the i-th original data point appeared in the bootstrap sample S_j^*. Define also the average post-resampling vector:

$$\bar{p}^a \equiv (\bar{p}_1^a, ..., \bar{p}_n^a) = \frac{1}{B}\sum_{j=1}^{B} p_j^a. \tag{3.16}$$

In this framework, the j-th bootstrap estimator $\bar{\theta}_j^*$ can be expressed as a function of the j-th post-resampling vector: $\bar{\theta}_j^* = \theta\left(p_j^a\right)$. Similarly, we have for the original GMM estimator $\hat{\theta}_{2s} = \theta\left(p^0\right)$, where $p^0 = \left(\frac{1}{n}, ..., \frac{1}{n}\right)$. Define also $\hat{\theta}^a = \theta\left(\bar{p}^a\right)$ as the estimator resultant from the application of the average post-sampling probabilities $\bar{p}^a$.

Instead of using $\hat{b} = \bar{\theta}^* - \theta\left(p^0\right)$, we propose the calculation of the bias of the GMM estimator as:

$$\bar{b} = \bar{\theta}^* - \theta\left(\bar{p}^a\right). \tag{3.17}$$

The intuition behind this is the following. Although the theoretical expectation of the resampling vector is p^0, its actual average is $\bar{p}^a$. Thus, using $\theta\left(\bar{p}^a\right)$ instead of $\theta\left(p^0\right)$ in the estimation of the bias, we might be able to correct this discrepancy. In fact, in (3.17), we are effectively comparing GMM estimators based on similar samples, as opposed to previously. The bias-corrected GMM estimator is then found by calculating:

$$\tilde{\theta}_{2s} = \hat{\theta}_{2s} - \bar{\theta}^* + \hat{\theta}^a. \tag{3.18}$$

When both n and B go to infinity, $\hat{\theta}^a$ will converge to $\hat{\theta}_{2s}$, so asymptotically this method will produce the same results as the other bootstrap techniques discussed in the previous sections. Note that we could have also opted for estimating the bias by $\bar{b} = \bar{\theta}^* - \theta\left(\hat{p}^{gel}\right)$, where $\hat{p}^{gel} \equiv \left(\hat{p}_1^{gel}, ..., \hat{p}_n^{gel}\right)$, with similar results being obtained, as $\bar{p}^a \simeq \hat{p}^{gel}$. The utilization of the post-resampling probabilities are only expected to provide a slight further improvement.[2]

[2]For this reason, the employment of post-resampling probabilities in the NP and RNP bootstrap methods would not produce significant improvements, as in these two cases $\bar{p}^a \simeq p^0$ and, hence, $\theta\left(\bar{p}_a\right) \simeq \hat{\theta}_{2s}$. Also, the same would happen for the FSEB bootstrap, since $\bar{p}^a \simeq \hat{p}^{fsgel}$ and the utilization of the latter weighting scheme in the GMM criterion function yields the same estimator

In terms of procedures, the algorithm presented in section 3.2 must be modified as follows. In step 3, for each bootstrap sample, in addition to the GMM estimator $\bar{\theta}^*_j$, we calculate also the post-resampling vector p^a_j using (3.14) and (3.15). In step 4, the average post-resampling vector $\bar{p}^a$ is also determined as in (3.16). In the final step, $\hat{\theta}^a$ is calculated by using $\bar{p}^a$ to weight each moment condition, i.e. instead of utilizing sample means to estimate the population moment conditions in expression (3.2), we use the post-resampling probabilities:

$$\bar{g}^a(\theta) = \sum_{i=1}^{n} \bar{p}^a_i g(y_i, \theta). \tag{3.19}$$

The estimation of the covariance matrix of the moment indicators needs also to be adapted to conform with this new weighting scheme as well as non-efficient GMM estimation for consistent initial estimates for θ_0. Finally, the bias is estimated according to expression (3.17).

3.4 Monte Carlo simulation study I: covariance structure models

Our first Monte Carlo investigation concerns models of covariance structures, which are important in the analysis of a variety of economic processes. Basically, they are employed to model the serial correlation structure of one economic variable in longitudinal data or the relation between movements in different economic variables (such as earnings and hours changes) over time. For applications involving these models see, for example, Abowd and Card (1987, 1989), Behrman, Rozenzweig and Taubman (1994), Griliches (1979) and Hall and Mishkin (1982).

Altonji and Segal (1996) carried out an extensive Monte Carlo analysis of the finite sample properties of GMM estimators for covariance structure models. They

$(\hat{\theta}_{2s})$ as that obtained when the empirical distribution is employed.

found that the efficient two-step GMM estimator is severely downward biased in small samples for most distributions and in relatively large samples for badly behaved distributions. They explain this poor performance as due to the correlation between the estimated second moments used to construct the moment indicators and the sampling optimal weighting matrix. Indeed, as they argue, moment conditions consisting of second moments are likely to be highly correlated with their covariance matrix "because individual observations that increase the sample estimate of a variance will also tend to increase the sample estimate of the variance of the variance". Thus, it is not surprising that both the equally weighted GMM, which uses the identity matrix as weighting matrix, and efficient GMM estimation based on split-sample estimators for the covariance matrix of the moment conditions produce parameter estimators with significantly improved properties in finite samples, as showed by Altonji and Segal (1996) in their investigation and also by Horowitz (1998) in a similar study. The latter author also considered the RNP bootstrap GMM estimator which, although also biased in some cases, performed much better than the standard two-step GMM estimator. In this section we examine the performance of the other estimation methods applicable in this context.

3.4.1 Experimental design

In order to investigate the behaviour of both asymptotically first-order equivalent methods to efficient GMM and bootstrap techniques for GMM estimators, we use the simplest experimental design analyzed by Altonji and Segal (1996). We consider a setting where the objective is the estimation of a population variance for a scalar random variable X from observations on a panel of individuals covering 10 time periods. Let each observation be denoted by X_{ti}, where $t = 1, ..., 10$ indexes the time period and $i = 1, ..., n$, indexes the individuals. For each period, the mean and the variance of the observations can be computed using the standard unbiased estimators

$$\bar{X}_t = \frac{1}{n}\sum_{i=1}^{n} X_{ti} \tag{3.20}$$

and

$$m_t = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ti} - \bar{X}_t\right)^2, \tag{3.21}$$

respectively. The estimates of the second moments are stacked into a 10-dimensional vector, m, and are related to the population variance, denoted by the parameter θ_0 (a scalar), through the 10-vector of moment conditions

$$E\left[g\left(\theta_0\right)\right] = E\left(m - \iota\theta_0\right) = 0, \tag{3.22}$$

where ι is a 10-vector of ones. With this formulation, we are assuming the equality between the variances of the 10 components of m and the nullity of the covariance of X across time periods.

Hence, in this Monte Carlo study, all samples are generated in a way that ensures that the data are independent across both t and i. The observations for all time periods were independently generated from the same distribution, with equal number of observations in each period, so the model defined by (3.22) is also homoskedastic. Although the elements of m are independent, both the diagonal and off-diagonal components of the estimated covariance matrix $\hat{V}_n$ of the moment indicators use sample estimates. Eight different distributions for X, scaled to have mean 0 and variance 1 (so $\theta_0 = 1$), and two sample sizes, 100 and 500, were considered. In each experiment, 1000 Monte Carlo replications were performed.

In this framework, the two-step GMM estimator represents a weighted mean of the ten sample variances,

$$\hat{\theta}_{2s} = w'm = \sum_{t=1}^{10} w_t m_t, \tag{3.23}$$

where $w = \left(\iota'\hat{V}_n^{-1}\iota\right)^{-1}\iota'\hat{V}_n^{-1}$ is a 10-dimensional vector of weights and $\hat{V}_n$ is a consistent estimator of V evaluated at equally weighted GMM estimators [which results from considering $w = \frac{1}{10}$ in (3.23)]. In turn, the GEL estimators are obtained using the procedures described in section 2.5.5 and satisfy the first-order equations (2.64). In this particular case, as $\frac{\partial g(\theta)}{\partial \theta'} = -\iota$, it can be proved that the second of those conditions may be reduced to (see the Appendix):

$$\sum_{t=1}^{10} \hat{\phi}_t = 0. \tag{3.24}$$

Furthermore, as $\hat{\phi}'\iota = 0$ and, hence, $\hat{\phi}'g\left(\hat{\theta}_{gel}\right) = \hat{\phi}'m$, the other first-order condition implies

$$\hat{\theta}_{gel} = \frac{1}{10}\sum_{t=1}^{10} m_t^*, \tag{3.25}$$

where

$$m_t^* = \frac{n}{n-1}\sum_{i=1}^{n} \hat{p}_i^{gel}\left(X_{ti} - \bar{X}_t\right)^2. \tag{3.26}$$

Compared to the two-step GMM estimator given in (3.23), we detect two important and interesting differences between these estimators. First, noting that (3.23) can be written as

$$\hat{\theta}_{2s} = \sum_{t=1}^{10} w_t \frac{n}{n-1}\sum_{i=1}^{n} \frac{1}{n}\left(X_{ti} - \bar{X}_t\right)^2, \tag{3.27}$$

we see that in each time period the two-step GMM estimator gives the same weight to each observation ($\frac{1}{n}$) while the GEL methods use the GEL implied probabilities as weights. Second, the former method assigns distinct weights, given by the vector w in (3.23), to each time period, while for the latter each time period receives an equal weight. Thus, notice that, if $\hat{p}_i^{gel} = \frac{1}{n}$, $i = 1, ..., n$, GEL estimators would be identical to the equally weighted GMM estimator.

3.4.2 Results

Table 3.1 reports the estimated mean bias (as a percentage)[3], standard error (SE), root mean squared error (RMSE) and median absolute error (MAE) of four asymptotically first-order equivalent methods for estimating moment condition models. The results obtained for the two-step GMM estimator are very similar to those presented by Altonji and Segal (1996). As in their study, this estimator is clearly downward biased, this distortion particularly marked for "badly-behaved" distributions, namely thicker-tailed symmetric (student-t with 5 degrees of freedom) and long-tailed skewed (lognormal and exponential) distributions. Increasing the sample size significantly improves inference but, for the aforementioned distributions, GMM estimators still display substantial bias. The worst case is given by the lognormal distribution, where the bias (MAE) is 41.5% (43%) and 22.5% (22.7%) for $n = 100$ and 500, respectively.

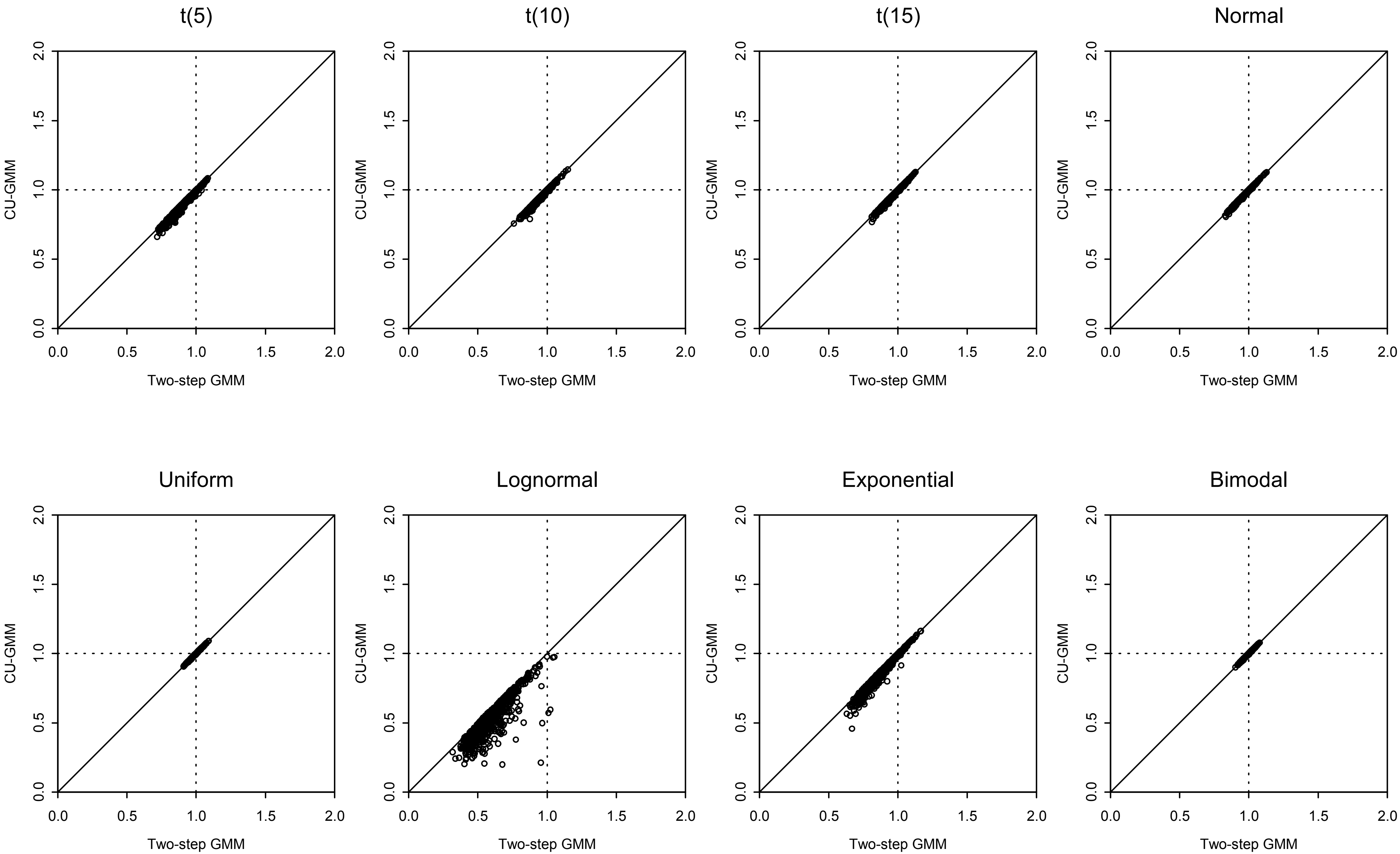
This poor performance of the two-step GMM estimator is due to the correlation between the moment indicators and the weighting matrix $\hat{V}_n^{-1}$, as discussed above. As this correlation is not eliminated by the iterative or continuous updating of the weighting matrix, it is perfectly natural that the CU-GMM estimator [which is numerically equal to Hansen, Heaton and Yaron's (1996) repeatedly-iterated GMM estimator in this framework] does not provide any improvement over two-step GMM estimation. Actually, as observed in Table 3.1, the results are even worse, as confirmed by the analysis of Figure 3.1, where some scatter plots comparing CU and two-step GMM estimates for the $n = 100$ case are shown. While for "well-behaved" distributions these methods produced very similar estimates (for $n = 500$ the Monte Carlo results are virtually identical), for t(5), exponential and, mainly, lognormal cases the CU-GMM clearly amplified the underestimation of the parameter of interest in most of the replications. Note also that in the last case both methods produced estimates less than 1, the true value of θ_0, in almost all replications.

[3]We do not report the median bias because it was very similar, with a single exception, referred to later on.

Table 3.1: Monte Carlo results for Two-Step GMM, CU-GMM, EI and EL estimators of a variance parameter using ten sample variances from one distribution (1000 replications)

Estimator	n=100				n=500			
	Mean bias	SE	RMSE	MAE	Mean bias	SE	RMSE	MAE
				t(5)				
2S-GMM	-.111	.065	.129	.116	-.041	.034	.053	.042
CU-GMM	-.125	.069	.143	.128	-.042	.034	.054	.043
EI	-.094	.067	.115	.099	-.029	.033	.044	.031
EL	-.065	.067	.094	.073	-.016	.034	.038	.026
				t(10)				
2S-GMM	-.059	.053	.079	.062	-.016	.025	.029	.021
CU-GMM	-.066	.055	.086	.068	-.016	.025	.030	.021
EI	-.046	.054	.071	.053	-.010	.024	.026	.018
EL	-.028	.055	.062	.043	-.004	.025	.025	.017
				t(15)				
2S-GMM	-.045	.056	.072	.052	-.012	.023	.026	.018
CU-GMM	-.051	.058	.077	.057	-.012	.023	.026	.018
EI	-.034	.056	.066	.045	-.007	.023	.024	.016
EL	-.018	.056	.059	.040	-.002	.023	.023	.015
				Normal				
2S-GMM	-.036	.047	.059	.041	-.008	.021	.022	.015
CU-GMM	-.040	.049	.063	.044	-.008	.021	.022	.015
EI	-.026	.048	.055	.038	-.005	.020	.021	.014
EL	-.015	.048	.050	.035	-.001	.021	.021	.014
				Uniform				
2S-GMM	-.007	.029	.030	.021	-.002	.013	.013	.009
CU-GMM	-.008	.030	.031	.021	-.002	.013	.013	.009
EI	-.005	.030	.030	.020	-.001	.013	.013	.009
EL	-.003	.030	.030	.020	-.001	.013	.013	.009
				Lognormal				
2S-GMM	-.415	.111	.429	.430	-.225	.082	.239	.227
CU-GMM	-.481	.125	.497	.490	-.231	.085	.246	.233
EI	-.396	.120	.414	.408	-.178	.079	.194	.182
EL	-.303	.131	.331	.317	-.118	.081	.143	.125
				Exponential				
2S-GMM	-.141	.087	.166	.147	-.041	.040	.057	.044
CU-GMM	-.162	.097	.189	.166	-.042	.040	.058	.045
EI	-.108	.088	.140	.113	-.024	.039	.046	.032
EL	-.058	.087	.105	.076	-.006	.039	.040	.029
				Bimodal				
2S-GMM	-.009	.028	.029	.020	-.002	.012	.013	.009
CU-GMM	-.010	.028	.030	.021	-.002	.012	.013	.009
EI	-.006	.028	.029	.020	-.001	.012	.012	.008
EL	-.002	.028	.028	.019	-.000	.012	.012	.008

Figure 3.1: Scatter plots for two-step and CU-GMM estimators of a variance parameter using ten sample variances from one distribution (n=100; 1000 replications)



A theoretical explanation for the small sample behaviour of these estimators arises from the results derived in Newey and Smith (2000), which were presented in the previous chapter. Indeed, comparing expressions (2.45) and (2.46), we see that the additional terms present in the bias function of the two-step GMM estimator disappears because, in this example, $G = -\iota$ and $\bar{V}_{\theta_j} = 0$, $j = 1, ..., s$. Therefore, the asymptotic biases of the CU and two-step GMM estimators are identical, which explains why these two estimators behave in such a similar way in this Monte Carlo experiment.[4]

Thus, it appears that estimation methods using estimators of the optimal weighting matrix based on simple sample means do not work well in this context. In order to obtain asymptotically efficient estimators with better finite sample properties, one solution consists in keeping two-step or CU-GMM estimation but utilizing split-sample estimators for V as those suggested by Altonji and Segal (1996) and Horowitz (1998), which reduce the correlation between the moment indicators and the covariance matrix and thus work relatively well. Another possible solution, which is now investigated, is the employment of asymptotically first-order equivalent methods not requiring the utilization of any weighting matrix such as GEL techniques.

The results obtained for EI and EL estimators are also reported in Table 3.1. In all cases both methods produce estimators with better finite sample properties relative to GMM. While all methods have very similar standard errors, the improvement in terms of bias, RMSE and MAE is clear, mainly in the case of EL estimation, although the bias is not completely eliminated in some cases. Relative to the two-step GMM estimator, for $n = 100$, the bias of the EL estimator is less between 27% (lognormal) and 79% (bimodal), the MAE between 4% (uniform) and 48% (exponential) and the RMSE between 2% (uniform) and 37% (exponential). For the EI estimator, the improvements are much more modest, ranging from 4% (lognormal) to 35% (bimodal) for the bias, from 0% (bimodal) to 23% (exponential) for the MAE

[4] Actually, the bias expressions derived by Newey and Smith (2000) are not strictly applicable here because in (3.22) m depends on the sample estimate of the mean of the observations (3.20). However, this should not affect significantly the behaviour of the estimators for θ_0, as additional experiments, not reported here, assuming a zero mean in (3.21), confirmed.

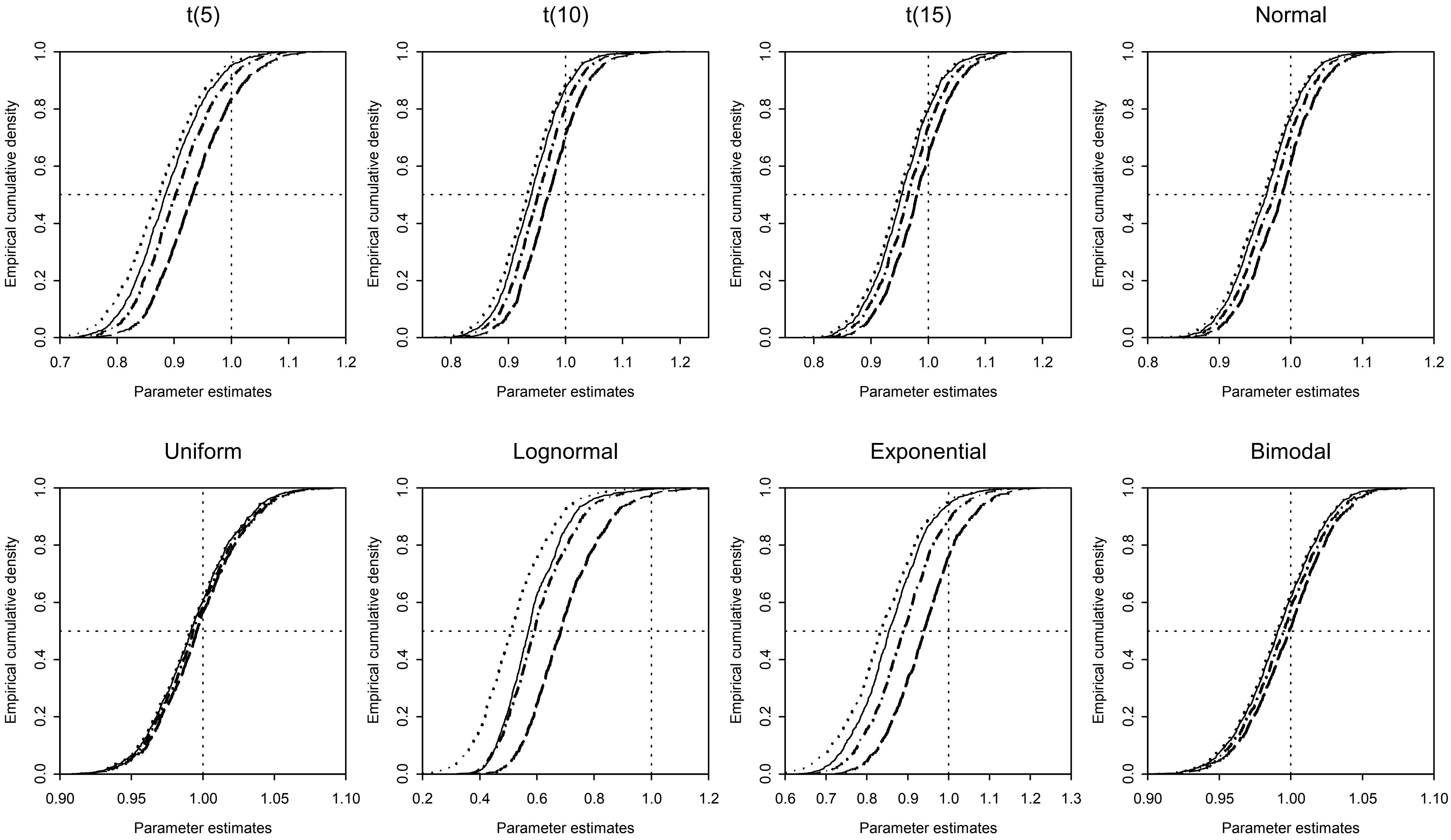
and from 1% (uniform) to 16% (exponential) for the RMSE. Again, using Newey and Smith (2000) results, we can explain theoretically why this happens. According to expressions (2.45), (2.46) and (2.73), and as $G = -\iota$ in this example, the asymptotic bias for two-step GMM, CU-GMM, EL and EI estimators are given by $b_{2s} = b_{cu} = \frac{1}{n} HE\left(g_i g_i' P g_i\right)$, $b_{ei} = 0.5 b_{2s}$ and $b_{el} = 0$. Clearly, this is the main reason for the superior performance of the EL method in this Monte Carlo experiment and for the less significant improvements resulting from application of the EI method.

The conclusions just drawn in the previous paragraphs are clearly confirmed by Figures 3.2 and 3.3 which show, respectively, the sampling cumulative and probability density functions for all estimators for the $n = 100$ case. As can be seen from both figures, whichever distribution for the data is considered, the performance of the estimation methods are ranked the same: the best is the EL method (dashed line), followed by the EI technique (dot-dashed line), the two-step GMM (solid line) and, finally, the CU-GMM (dotted line). Only in terms of dispersion is the behaviour of all methods very similar, with the exception of the lognormal case, where the distribution of the two-step GMM estimator is slightly more concentrated.

We also analyzed the ability of the alternative bootstrap techniques discussed in section 3.3 to improve the finite sample properties of two-step GMM estimators. The FSGEL, GEL, RGEL and PHGEL bootstrap methods were implemented using the EL criterion function. Although the observations are independent across both individuals and time periods, we adopted the resampling scheme usual in the panel data context, i.e. we sampled with replacement from the set of n individuals. The results reported in Table 3.2 were computed using 100 bootstrap samples in each replication.

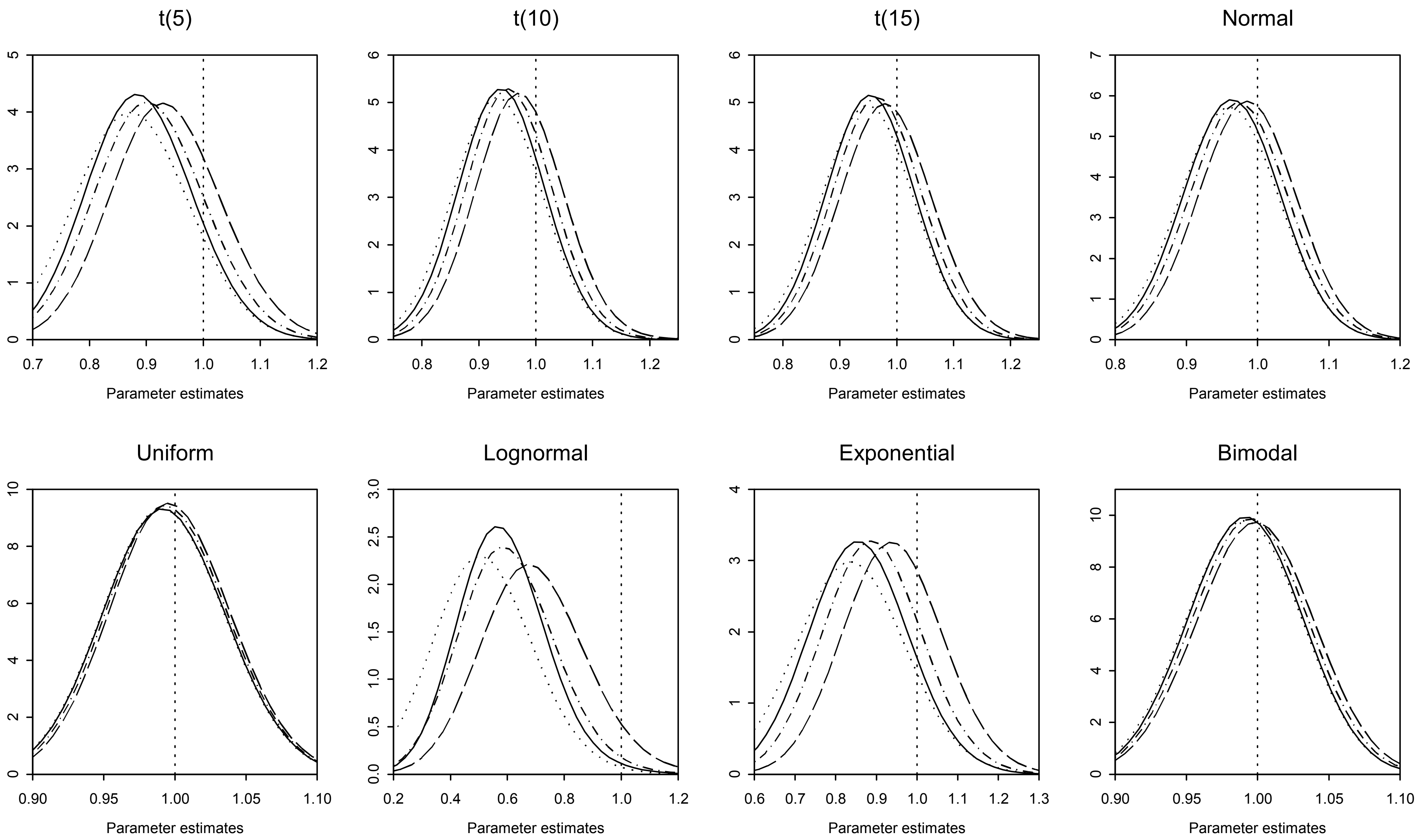
As we can see, in all cases the utilization of any one of the bootstrap methods allows the bias of the GMM estimator to be substantially reduced, although at the expense of an increment in its dispersion. However, the behaviour of these methods is not at all uniform. Analyzing firstly the three methods previously suggested by

Figure 3.2: Sampling cumulative density functions for GMM, CU-GMM, EI, and EL estimators of a variance parameter using ten sample variances from one distribution (n=100; 1000 replications)



Notes: Two-step GMM (solid line), CU-GMM (dotted line), EI (dot-dashed line), EL (dashed line).

Figure 3.3: Sampling probability density functions for GMM, CU-GMM, EI, and EL estimators of a variance parameter using ten sample variances from one distribution (n=100; 1000 replications)



Notes: Two-step GMM (solid line), CU-GMM (dotted line), EI (dot-dashed line), EL (dashed line).

Table 3.2: Monte Carlo results for bootstrap GMM estimators of a variance parameter using ten sample variances from one distribution (1000 replications)

Estimator	n=100				n=500			
	Mean bias	SE	RMSE	MAE	Mean bias	SE	RMSE	MAE
				t(5)				
NP-BOOT-GMM	-.073	.076	.105	.084	-.020	.038	.042	.029
RNP-BOOT-GMM	-.050	.077	.091	.068	-.014	.039	.041	.028
FSEL-BOOT-GMM	-.044	.075	.086	.065	-.014	.038	.040	.028
EL-BOOT-GMM	-.088	.076	.116	.094	-.038	.040	.056	.041
REL-BOOT-GMM	-.041	.075	.086	.065	-.013	.038	.040	.027
PHEL-BOOT-GMM	-.042	.075	.086	.064	-.014	.038	.040	.028
				t(10)				
NP-BOOT-GMM	-.026	.060	.065	.046	-.003	.026	.026	.018
RNP-BOOT-GMM	-.017	.059	.061	.044	-.002	.026	.026	.018
FSEL-BOOT-GMM	-.011	.058	.059	.040	-.002	.026	.026	.018
EL-BOOT-GMM	-.042	.059	.073	.052	-.013	.027	.030	.021
REL-BOOT-GMM	-.011	.058	.059	.041	-.001	.026	.026	.018
PHEL-BOOT-GMM	-.011	.058	.059	.041	-.002	.026	.026	.018
				t(15)				
NP-BOOT-GMM	-.014	.061	.062	.042	-.001	.024	.024	.016
RNP-BOOT-GMM	-.008	.060	.060	.041	.000	.024	.024	.016
FSEL-BOOT-GMM	-.002	.059	.059	.039	.000	.024	.024	.016
EL-BOOT-GMM	-.030	.061	.067	.046	-.010	.025	.026	.018
REL-BOOT-GMM	-.002	.059	.059	.040	.000	.024	.024	.016
PHEL-BOOT-GMM	-.003	.059	.059	.039	.000	.023	.023	.016
				Normal				
NP-BOOT-GMM	-.008	.050	.051	.036	.000	.021	.021	.014
RNP-BOOT-GMM	-.005	.050	.050	.035	.001	.021	.021	.014
FSEL-BOOT-GMM	-.001	.049	.049	.033	.001	.021	.021	.014
EL-BOOT-GMM	-.022	.050	.055	.038	-.006	.021	.022	.015
REL-BOOT-GMM	-.001	.049	.049	.034	.001	.021	.021	.013
PHEL-BOOT-GMM	-.001	.049	.049	.034	.001	.021	.021	.013
				Uniform				
NP-BOOT-GMM	.006	.030	.030	.020	.001	.013	.013	.009
RNP-BOOT-GMM	.005	.030	.030	.020	.001	.013	.013	.009
FSEL-BOOT-GMM	.007	.030	.030	.020	.001	.013	.013	.009
EL-BOOT-GMM	.003	.030	.030	.020	.000	.013	.013	.009
REL-BOOT-GMM	.007	.030	.030	.020	.001	.013	.013	.009
PHEL-BOOT-GMM	.007	.030	.030	.020	.001	.013	.013	.008
				Lognormal				
NP-BOOT-GMM	-.380	.145	.407	.403	-.161	.108	.194	.168
RNP-BOOT-GMM	-.230	.453	.508	.289	-.107	.123	.163	.129
FSEL-BOOT-GMM	-.264	.158	.308	.292	-.121	.106	.161	.131
EL-BOOT-GMM	-.353	.157	.387	.378	-.209	.123	.243	.216
REL-BOOT-GMM	-.242	.165	.293	.271	-.103	.109	.150	.121
PHEL-BOOT-GMM	-.244	.165	.294	.273	-.104	.109	.150	.121
				Exponential				
NP-BOOT-GMM	-.089	.108	.140	.107	-.012	.044	.046	.032
RNP-BOOT-GMM	-.060	.105	.122	.085	-.009	.044	.045	.031
FSEL-BOOT-GMM	-.042	.102	.110	.077	-.007	.043	.044	.030
EL-BOOT-GMM	-.122	.110	.164	.133	-.040	.049	.063	.046
REL-BOOT-GMM	-.039	.103	.110	.077	-.006	.043	.044	.031
PHEL-BOOT-GMM	-.040	.103	.110	.076	-.006	.043	.044	.030
				Bimodal				
NP-BOOT-GMM	.006	.029	.029	.021	.002	.013	.013	.008
RNP-BOOT-GMM	.006	.028	.029	.020	.002	.013	.013	.008
FSEL-BOOT-GMM	.008	.028	.029	.020	.002	.012	.013	.008
EL-BOOT-GMM	.001	.029	.029	.020	.000	.013	.013	.009
REL-BOOT-GMM	.008	.028	.029	.020	.002	.012	.013	.008
PHEL-BOOT-GMM	.008	.028	.029	.021	.002	.012	.013	.008

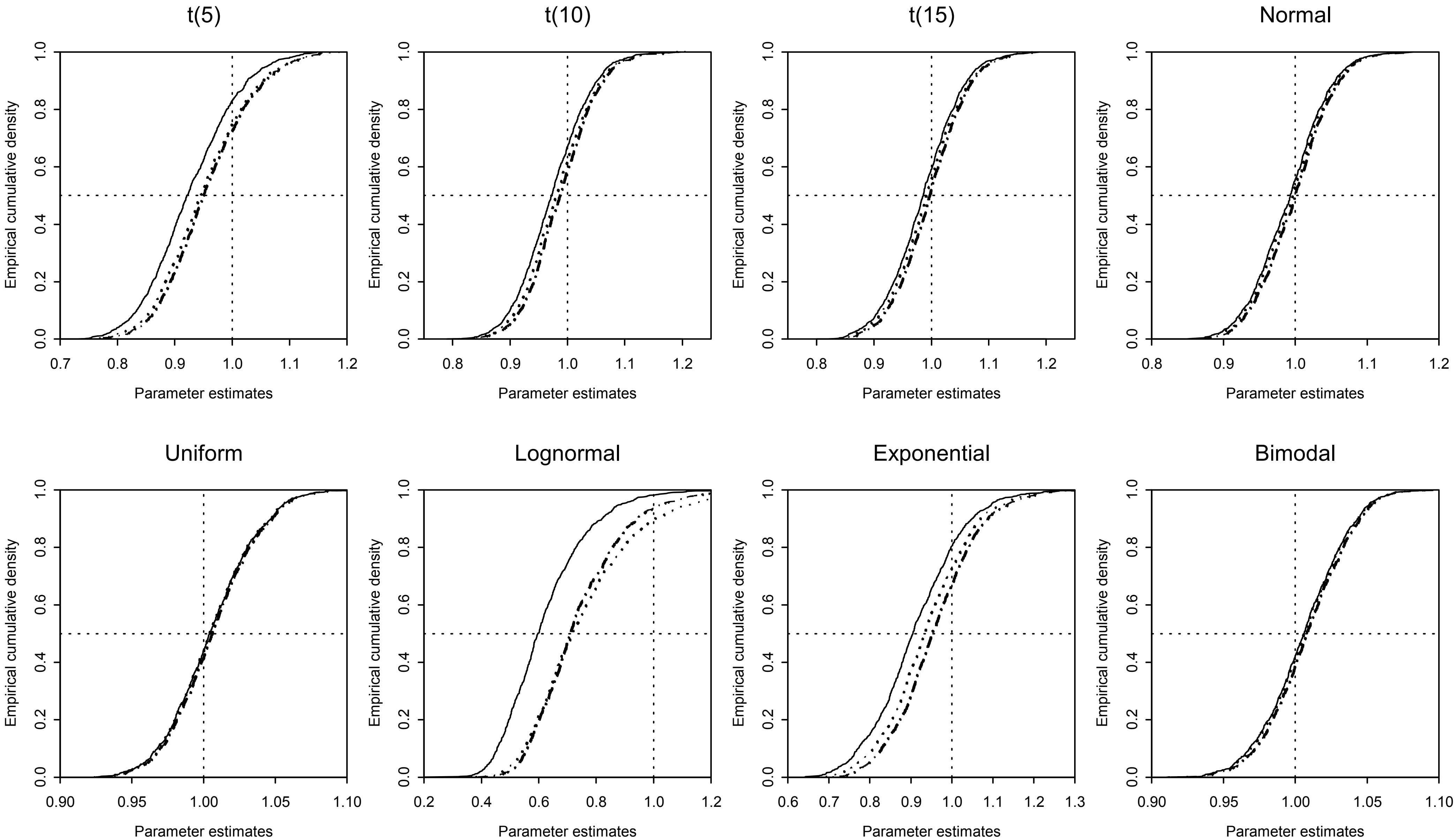
other authors, we see that they produce estimators with less bias, RMSE and MAE than the GMM estimator, but the improvements are much less significant for the NP bootstrap, as expected. The RNP and FSEL methods yielded very similar results for $n = 500$ but, for the smaller sample size considered (see also Figures 3.4 and 3.5), the FSEL bootstrap in general performed better, which is not surprising, since, as already referred to, Brown, Newey and May (1997) demonstrated that this method is efficient relative to any bootstrap method based on the empirical distribution.

With regard to the methods proposed in this chapter, the EL bootstrap, for the reasons argued in section 3.3.4, systematically under-estimated the bias of the GMM estimator, so no significant improvements were achieved. Thus, this is the only bootstrap estimator which is sometimes characterized by a larger RMSE than that of the two-step GMM estimator. In contradistinction, both the REL and the PHEL bootstrap methods produced very promising (and almost identical) results (see also Figures 3.6 and 3.7, where the lines for these methods are indistinguishable). Apart from the over-correction produced for uniform and bimodal distributions at $n = 100$ (a problem shared by all bootstrap methods), in the remaining cases for this sample size the improvement in terms of bias over the EL bootstrap GMM estimator ranges from 31% (lognormal) to 94% (normal) and, in relation to the two-step GMM estimator, from 41% (lognormal) to 96% (normal), which is quite impressive. Furthermore, certainly due to the employment of a more efficient estimator of the distribution of the data, the performance of these two bootstrap GMM estimators was clearly superior to that of RNP bootstrap estimators for all criteria in almost all cases[5] and slightly better than that of FSEL bootstrap estimators for the "badly-behaved" distributions.

With the exception of the NP and EL bootstrap methods, all others behave better in all experiments in terms of bias than EL, the method that produced the best

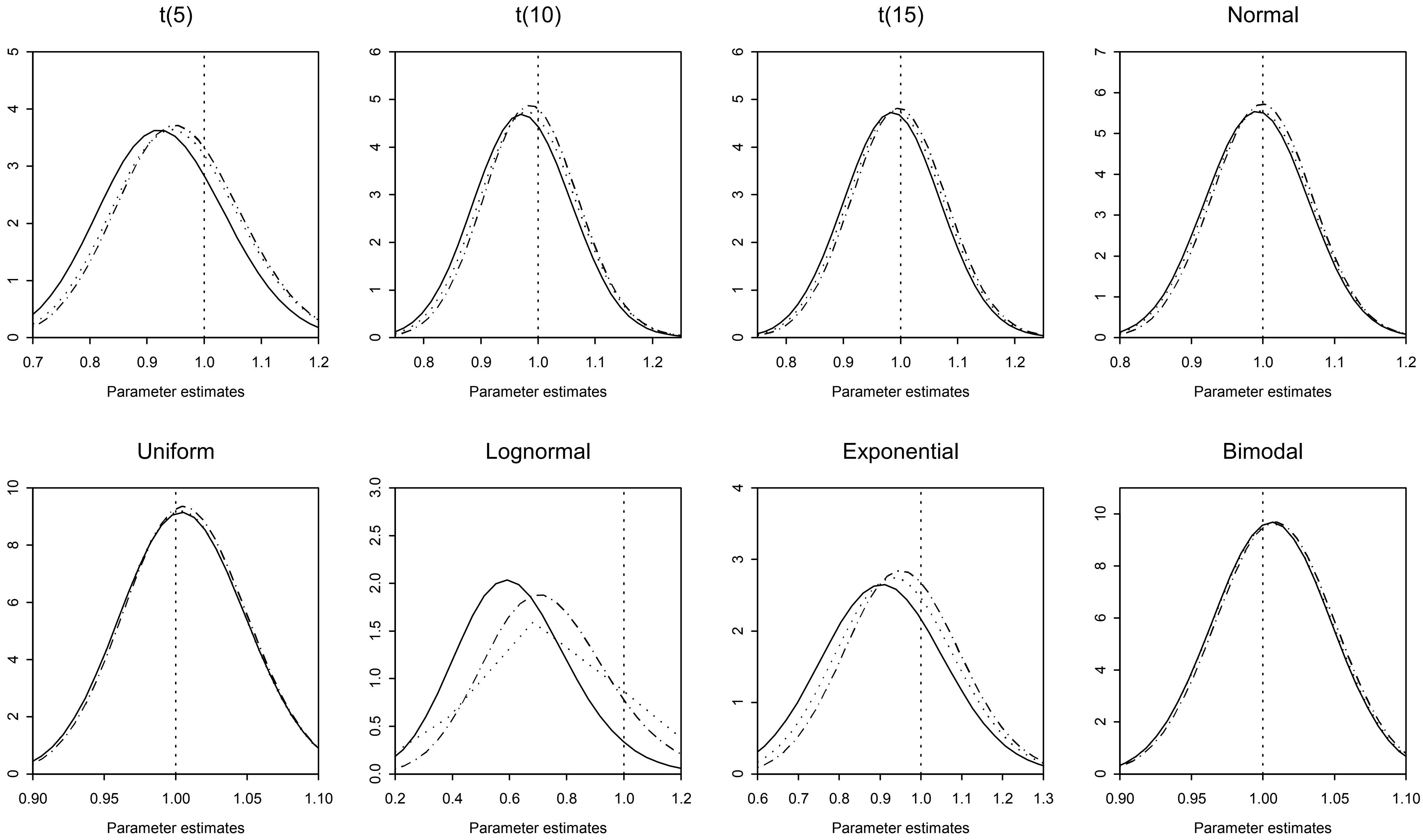
[5]Note that the exception found for the mean bias for the lognormal and $n = 100$ case was due to an outlier, as the enormous standard error of the RNP bootstrap GMM estimator indicates. In terms of median bias, the values are 0.282 (RNP), 0.290 (FSEL), 0.267 (REL) and 0.270 (PHEL).

Figure 3.4: Sampling cumulative density functions for NP, RNP and FSEL bootstrap GMM estimators of a variance parameter using ten sample variances from one distribution (n=100; 1000 replications)



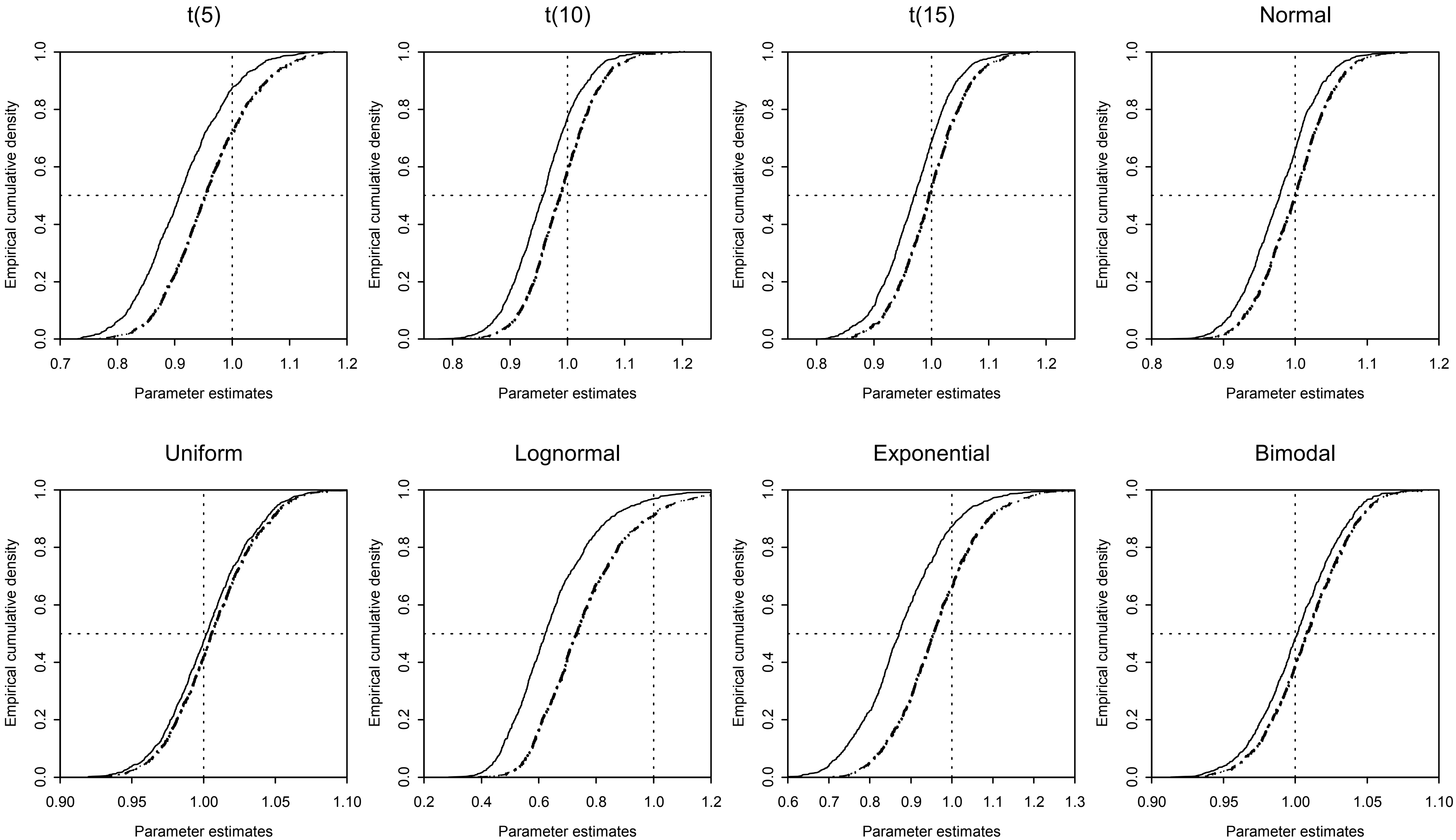
Notes: NP bootstrap GMM (solid line), RNP bootstrap GMM (dotted line), FSEL bootstrap GMM (dot-dashed line).

Figure 3.5: Sampling probability density functions for NP, RNP and FSEL bootstrap GMM estimators of a variance parameter using ten sample variances from one distribution (n=100; 1000 replications)



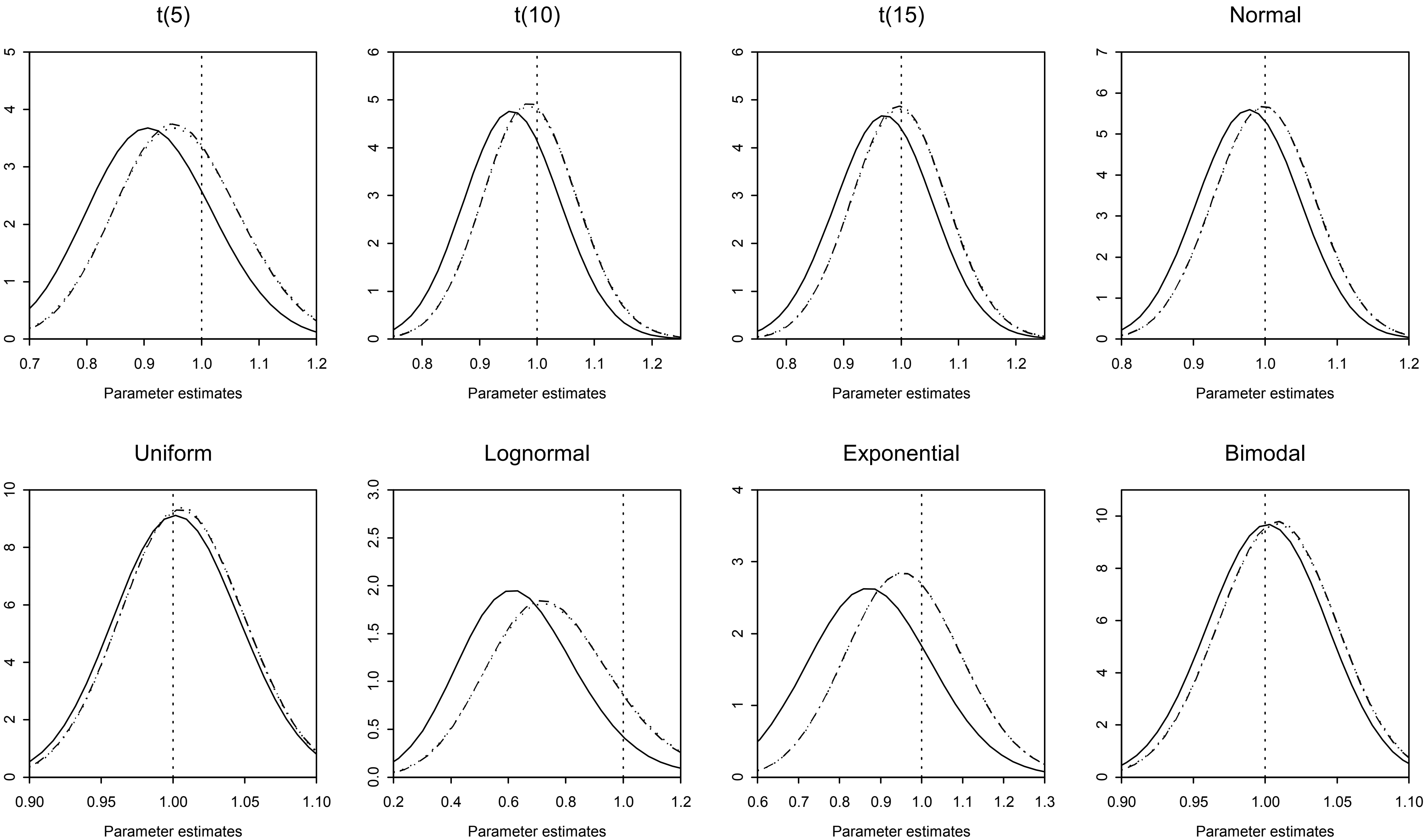
Notes: NP bootstrap GMM (solid line), RNP bootstrap GMM (dotted line), FSEL bootstrap GMM (dot-dashed line).

Figure 3.6: Sampling cumulative density functions for EL, REL and PHEL bootstrap GMM estimators of a variance parameter using ten sample variances from one distribution (n=100; 1000 replications)



Notes: EL bootstrap GMM (solid line), REL bootstrap GMM (dotted line), PHEL bootstrap GMM (dot-dashed line).

Figure 3.7: Sampling probability density functions for EL, REL and PHEL bootstrap GMM estimators of a variance parameter using ten sample variances from one distribution (n=100; 1000 replications)



Notes: EL bootstrap GMM (solid line), REL bootstrap GMM (dotted line), PHEL bootstrap GMM (dot-dashed line).

results in Table 3.1. However, they sometimes have a larger RMSE due to the greater dispersion usually exhibited by bootstrap estimators.

3.5 Monte Carlo simulation study II: instrumental variable models

In this second Monte Carlo investigation we consider instrumental variable models, one of the most wide spread applications of GMM. There are numerous studies showing that, in small samples, GMM estimators are not unbiased, especially when the number of instruments is large [e.g. Tauchen (1986b), Kocherlakota (1990) and Andersen and Sorensen (1996)] or the correlation between regressors and instruments is weak [e.g. Nelson and Startz (1990) and Bound, Jaeger and Baker (1995)]. In this section we present additional evidence confirming those results and examine how the alternative estimation methods under analysis perform in this framework.

3.5.1 Data generating process

Consider the linear model described by the equation

$$y = X\theta_0 + u, \tag{3.28}$$

where y and X are n-vectors of observations on a dependent variable and a regressor variable, respectively, and u is a n-vector of normal errors with mean zero and variance one. Analogously to Nelson and Startz (1990), we generate the regressor X and the s instruments Z_j, $j = 1, ..., s$, that constitute the matrix of instruments Z from

$$X = \lambda u + \epsilon \tag{3.29}$$

and

$$Z_j = \gamma_j \epsilon + v_j, \tag{3.30}$$

$j = 1, ..., s$, where ϵ and v_j are random disturbances independently generated from a $N(0, I)$ distribution and λ and γ_j are fixed parameters that allow the correlations ρ_{xu} between X and u and ρ_{xz_j} between X and the instrument Z_j to be controlled according to the equations

$$\lambda = \frac{\rho_{xu}}{\sqrt{1-\rho_{xu}^2}} \tag{3.31}$$

and

$$\gamma_j = \rho_{xz_j}\sqrt{\frac{1+\lambda^2}{1-\left(1+\lambda^2\right)\rho_{xz_j}^2}}. \tag{3.32}$$

As we are assuming homoskedasticity, the two-step GMM estimator is given by

$$\hat{\theta} = \left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1}X'Z\left(Z'Z\right)^{-1}Z'y, \tag{3.33}$$

while GEL estimators can be expressed as [see Smith (1997), p. 517]

$$\hat{\theta} = \left[X'\hat{P}Z\left(Z'Z\right)^{-1}Z'\hat{P}X\right]^{-1}X'\hat{P}Z\left(Z'Z\right)^{-1}Z'\hat{P}y, \tag{3.34}$$

where $\hat{P}$ is a $(n \times n)$ diagonal matrix with typical element $\hat{p}_i^{gel}$, $i = 1, ..., n$. Comparing expressions (3.33) and (3.34), we see that, again, the difference between these estimators results from the weights applied to the matrices $Z'X$ and $Z'y$: the two-step GMM estimator applies unit weight whereas the GEL estimators weight each component of those matrices using the GEL implied probabilities.

Five different experiments were performed, as described in Table 3.3. In the first case, we have just a single overidentifying moment condition, where one of the instruments utilized in estimation is worthless. The second experiment is similar, with the modification that there is a large number of instruments relative to the number of regressors. The nine instruments added are also useless. This characteristic was kept in experiments 3 and 4, which are simple extensions of experiment 2. In the first case we investigate the effects of increasing the correlation between the explanatory variable and the instrument Z_1. In the other case we examine the consequences of lower feedbacks from u to X in equation (3.28) over the parameter estimates.

Table 3.3: Monte Carlo experiments for instrumental variable models

Experiment	s	ρ_{xu}	ρ_{xz_1}	ρ_{xz_2}	ρ_{xz_j} $(j = 3, ..., 11)$
1	2	0.7	0.3	0	-
2	11	0.7	0.3	0	0
3	11	0.7	0.7	0	0
4	11	0.3	0.7	0	0
5	11	0.7	0.3	0	0.3

The latter effect is not usually analyzed [the only exception seems to be Blomquist and Dahlberg (1999)] but, as Nelson and Startz (1990) implicitly acknowledge, the correlation between the error term u and the regressor X is one of the most important determinants of the accuracy with which an IV model may be estimated, because high feedbacks from u to X make the model poorly identified even when the correlation between regressors and IVs is relatively important. Finally, in experiment 5, we repeat experiment 2 but now the additional nine instruments utilized convey information about X.

For each experiment, 1000 replications of samples of both 100 and 500 observations were generated. The parameter θ_0 was fixed at 1. Once again, all bootstrap methods using a GEL (or FSGEL) distribution to resample the data were based on the EL implied probabilities and on 100 bootstrap samples for each replication. We resampled with replacement from the original (y, X, Z) sample.

3.5.2 Results

The results obtained for $n = 100$ are presented in Table 3.4. In addition to the statistics reported in the previous section we also report the bias in terms of median and the values of the 0.05 and 0.95 quantiles of the distribution of the various estimators of the parameter θ_0. Indeed, unlike before, the median is substantially different from the mean for some of the estimation methods considered. Moreover, the tails of some of the distributions of the estimators are now much heavier for some methods, while in the previous study all of them were characterized by very similar standard errors, apart from the expected larger dispersion for GMM bootstrap estimators. In Figures

3.8 and 3.9 we show also the sampling cumulative distribution functions for some estimation methods.

Similarly to the results widely reported by other simulation studies, the two-step GMM estimator [which, in this context, is numerically equal to Hansen, Heaton and Yaron's (1996) repeatedly-iterated GMM estimator] is significantly biased in all experiments. Its best (least bad) performance in terms of bias occurs when only two instruments are used (experiment 1), precisely the case where it exhibits more dispersion, which reflects the traditional trade-off between bias and efficiency that usually happens when the number of moment conditions is increased and the two-step GMM estimator is employed. Note that this effect occurs not only when the nine instruments added are useless (experiments 2-4) but also in experiment 5, where each one of the new instruments has the same correlation with X as the instrument Z_1 in experiment 1. Notice also that the decrease in the dispersion of the two-step GMM estimator when new instruments are added is such that its RMSE is substantially lower in experiments 2-5. In all cases, this estimator has the smallest standard error of all estimation methods considered.

The bias of the two-step GMM is particularly significant in experiment 2, where this method clearly overestimates the parameter θ_0, producing estimates greater than 1, the true value of θ_0, in 96.8% of the replications realized. In experiment 3 the two-step GMM estimator is still very biased but there is an important improvement in its small sample properties, which shows clearly the beneficial effects of high correlations between instruments and regressors on the performance of this estimator. In fact, although 10 instruments are still worthless, the mean bias of the two-step GMM estimator is reduced by 68.6% and its standard error by 39.2% by merely increasing the correlation between the regressor and the remaining instrument from 0.3 to 0.7. With regard to the feedback from u to X in equation (3.28), its decrease seems to have two distinct consequences for the GMM estimator, as shown by the results obtained with experiment 4. On the one hand, its bias diminishes considerably, which was expected

Table 3.4: Performance of alternative estimators for instrumental variable models (1000 Monte Carlo replications; n = 100)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
Model 1							
2S-GMM	.023	.066	0.604	1.347	.315	.316	.159
CU-GMM	.254	.008	0.264	1.317	12.490	12.492	.169
EI	-.111	.004	0.206	1.320	.698	.707	.173
EL	-.124	.007	0.178	1.328	.866	.874	.172
NP-BOOT-GMM	-.021	.060	0.397	1.358	.504	.504	.164
RNP-BOOT-GMM	-.022	.061	0.395	1.358	.503	.504	.165
FSEL-BOOT-GMM	-.015	.060	0.465	1.351	.504	.504	.164
EL-BOOT-GMM	.036	.102	0.444	1.424	.513	.514	.187
REL-BOOT-GMM	-.032	.053	0.376	1.349	.511	.512	.167
PHEL-BOOT-GMM	-.097	.044	0.161	1.340	.764	.770	.171
Model 2							
2S-GMM	.280	.278	1.042	1.497	.143	.314	.278
CU-GMM	.091	.009	-0.033	1.405	3.976	3.977	.192
EI	-.233	.005	-0.269	1.407	1.610	1.627	.205
EL	-.201	.004	-0.228	1.406	1.313	1.329	.202
NP-BOOT-GMM	.192	.200	0.829	1.477	.204	.280	.218
RNP-BOOT-GMM	.193	.202	0.829	1.478	.203	.280	.218
FSEL-BOOT-GMM	.201	.206	0.865	1.480	.194	.279	.217
EL-BOOT-GMM	.315	.320	0.948	1.666	.222	.385	.325
REL-BOOT-GMM	.163	.175	0.776	1.465	.221	.274	.204
PHEL-BOOT-GMM	-.087	.047	-0.105	1.416	.744	.749	.210
Model 3							
2S-GMM	.088	.098	0.928	1.217	.087	.124	.103
CU-GMM	-.010	.007	0.798	1.136	.107	.108	.072
EI	-.016	.003	0.751	1.157	.123	.124	.077
EL	-.015	.002	0.759	1.153	.121	.122	.074
NP-BOOT-GMM	.020	.035	0.826	1.175	.106	.108	.079
RNP-BOOT-GMM	.020	.035	0.824	1.173	.106	.108	.079
FSEL-BOOT-GMM	.023	.037	0.827	1.174	.105	.108	.080
EL-BOOT-GMM	.102	.108	0.892	1.288	.118	.156	.120
REL-BOOT-GMM	.008	.024	0.801	1.164	.110	.110	.077
PHEL-BOOT-GMM	-.000	.017	0.789	1.158	.114	.114	.076
Model 4							
2S-GMM	.049	.060	0.825	1.255	.129	.138	.099
CU-GMM	-.010	.002	0.728	1.222	.150	.151	.092
EI	-.016	-.005	0.695	1.236	.169	.170	.109
EL	-.016	-.003	0.698	1.237	.169	.170	.107
NP-BOOT-GMM	.010	.020	0.758	1.231	.142	.143	.091
RNP-BOOT-GMM	.010	.019	0.758	1.233	.142	.142	.091
FSEL-BOOT-GMM	.008	.020	0.758	1.231	.142	.142	.093
EL-BOOT-GMM	.065	.068	0.825	1.307	.147	.161	.114
REL-BOOT-GMM	.005	.017	0.755	1.232	.144	.144	.094
PHEL-BOOT-GMM	.000	.012	0.744	1.231	.147	.147	.093
Model 5							
2S-GMM	.117	.129	0.938	1.262	.099	.153	.132
CU-GMM	-.018	-.000	0.733	1.168	.136	.137	.083
EI	-.029	-.003	0.669	1.181	.160	.163	.096
EL	-.028	.000	0.676	1.185	.158	.161	.093
NP-BOOT-GMM	.035	.049	0.808	1.211	.128	.133	.097
RNP-BOOT-GMM	.036	.051	0.804	1.209	.128	.133	.095
FSEL-BOOT-GMM	.041	.056	0.818	1.217	.124	.131	.094
EL-BOOT-GMM	.138	.146	0.906	1.360	.145	.200	.155
REL-BOOT-GMM	.018	.037	0.782	1.199	.134	.135	.094
PHEL-BOOT-GMM	-.005	.021	0.718	1.188	.148	.148	.092

Figure 3.8: Sampling cumulative density functions for GMM, CU-GMM, EI, and EL estimators of instrumental variable models (n=100; 1000 replications)

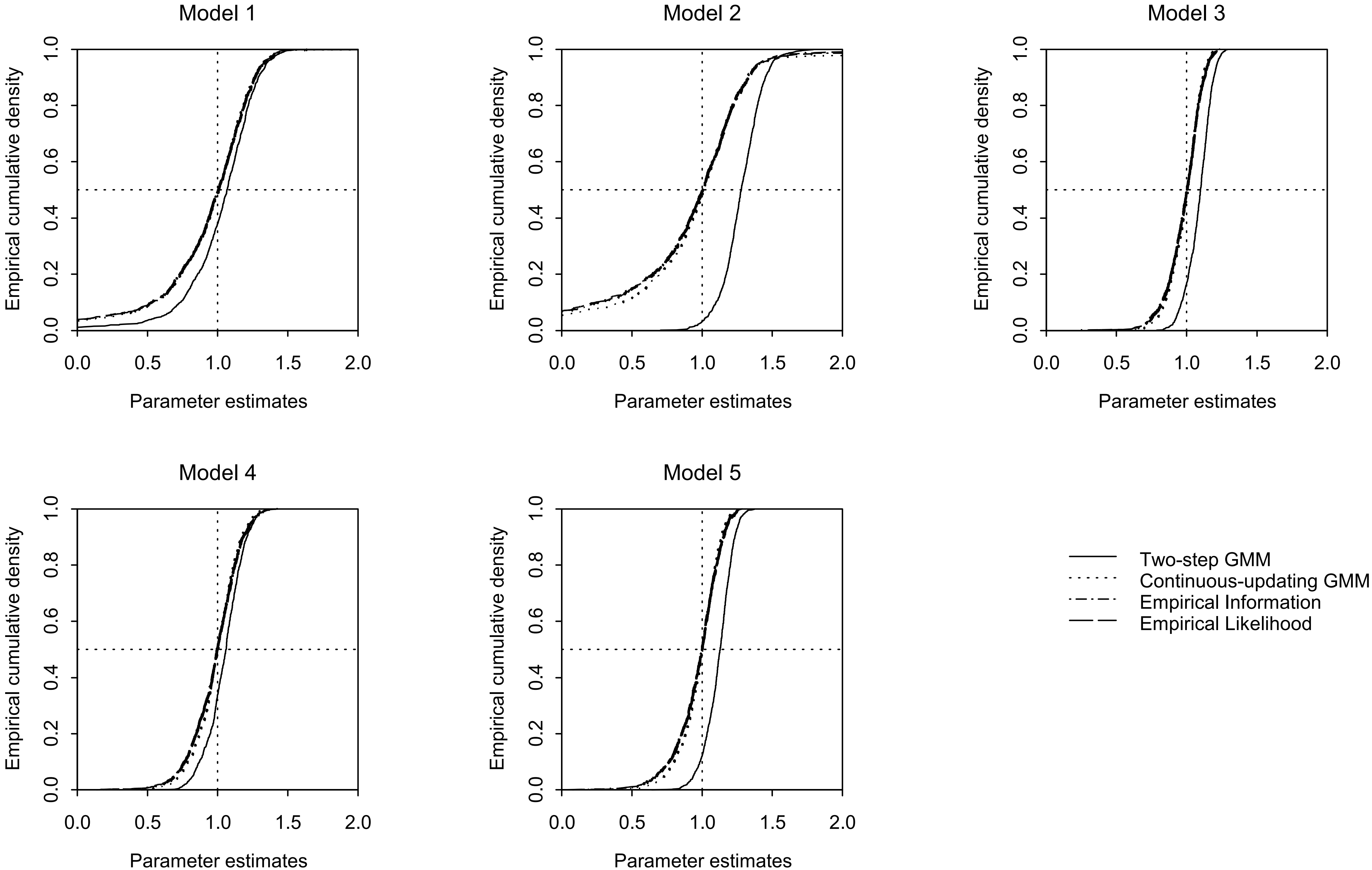
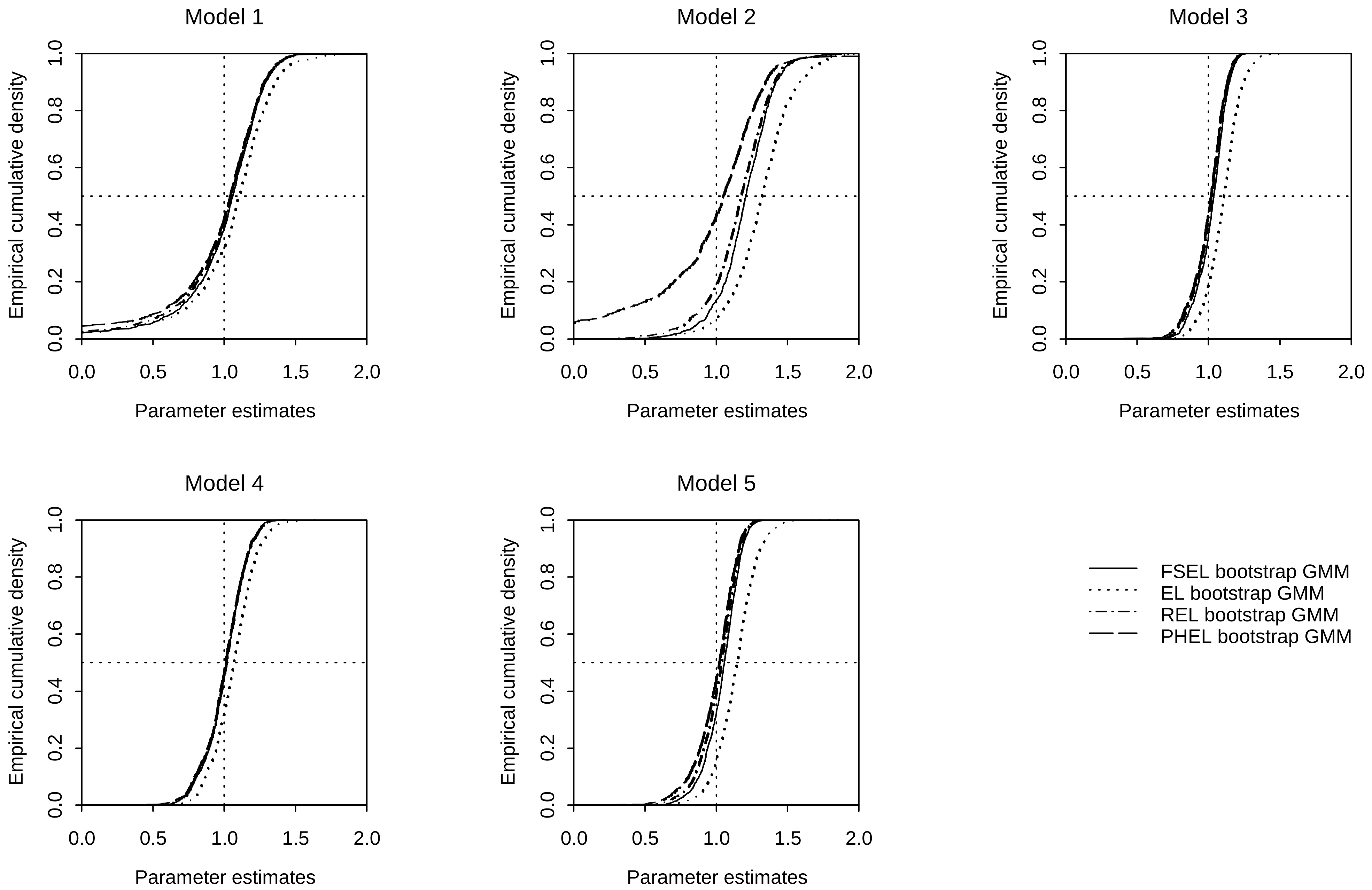


Figure 3.9: Sampling cumulative density functions for FSEL, EL, REL and PHEL bootstrap GMM estimators of instrumental variable models (n=100; 1000 replications)



because, although the correlation between Z_1 and X is still 0.7, the component of the regressor not correlated with the error term now has a greater influence over the behaviour of the dependent variable. On the other hand, there was an increase in its dispersion, probably due to the higher variability of y, which in turn results directly from X and u being less dependent. Finally, the results obtained in experiment 5, although better than those achieved for experiment 2 as expected, are worse than those of experiment 3, which emphasizes the importance of high correlations between instruments and regressors in this framework. Indeed, despite the existence of 10 useless instruments in experiment 3 and only 1 in experiment 5, the presence of a single good instrument in the former model is sufficient for better results than those obtained when 10 reasonable instruments are used in the latter.

Unlike the previous section, the CU-GMM, EI and EL estimators now exhibit a very similar behaviour in all experiments, as can be immediately seen from Figure 3.8, where their sampling cumulative density functions are almost indistinguishable. This happens because, in this case of moments consisting of products of instruments with a Gaussian residual, the third moments of g_i are zero, so the last term of (2.45) and (2.73) disappears and, hence, the asymptotic biases of these three estimators become equal. It can also be seen that, while the two-step GMM estimator is severely biased in all cases, the other three are always nearly median unbiased, a property which is independent of the quality and the number of instruments used in estimation.[6] However, for the poorest identified models (experiments 1 and 2), the Monte Carlo distributions of their estimators are quite disperse, having very heavy left tails. The tremendous standard errors in these two cases, especially of the CU-GMM estimator, are due to the occurrence of extreme values in some replications. These results conform with those obtained by Hansen, Heaton and Yaron (1996), which showed that the criterion function for the CU-GMM estimator can sometimes lead to extreme out-

[6]This confirms empirically the theoretical results of Newey and Smith (2000), which derive bias expressions for the two-step and CU-GMM estimators for a instrumental variable model of the kind considered here. They show that while the bias of the former estimator increases linearly with the number of moment conditions, the bias of the latter does not depend on it.

liers for the minimizing value of θ but that, in general, this estimator will be median unbiased [see also the results reported by Stock and Wright (1996)].[7] By increasing the correlation between instruments and regressors, much more concentrated sampling distributions for these three estimators are obtained, without extreme values. For this reason, only small mean biases are present in experiments 3-5, substantially less than that of the two-step GMM estimator.

With regard to the bootstrap methods, the first aspect to note is that those whose resampling is based on the empirical distribution $F_n(y)$, the NP and RNP bootstrap methods, produced almost identical results in all experiments. The reason for this behaviour in this setting seems to be the following. As can be inferred from section 3.3.2, the more distant from zero are the sample moment conditions evaluated at the two-step GMM estimator, the more significant are the differences between these two bootstrap techniques. With covariance structure models, for the reasons stated earlier, the estimated value of the moments was significantly different from zero, so the improvement produced by the RNP bootstrap was substantial. Here, the sample moment conditions evaluated at the two-step GMM estimator are nearly zero in all cases. Indeed, although most are not good instruments, once they convey little or no information at all about the explanatory variable (and this is the main reason why the estimators are strongly biased in some cases), they are not correlated with the error term, so the sample moment conditions are very close to zero, attenuating the differences between these bootstrap methods. For the same reason, the FSEL implied probabilities are approximately equal to $\frac{1}{n}$ for all observations, in all cases. Thus, also the FSEL bootstrap produced very similar results to the NP and RNP bootstrap methods.

In the first two experiments, which concern the most poorly identified models, the performance of these three bootstrap methods is not particularly promising. In the first case, they are only very slightly less biased than the two-step GMM estimator

[7] Note that median bias is more appropriate than mean bias to assess the performance of the CU-GMM estimator because, in this example, it coincides with the limited information maximum likelihood estimator, which is known to have no finite moments [see *inter alia* Mariano (1982)].

itself and their sampling distributions are much more variable. In the second case, although they cut the bias of the two-step GMM estimator by about 30%, their bias is still very high (around 20%). However, their behaviour improves substantially in the remaining experiments. In these models, using the FSEL bootstrap method or bootstrapping the two-step GMM estimator utilizing the empirical distribution $F_n(y)$ is an effective way of largely, although still not entirely, correcting its bias.

While the standard form of the EL bootstrap method is not able to improve at all the properties of two-step GMM estimators in any case (the results are worse according to all criteria), the two proposed adjustments work very well, as shown clearly by Figure 3.9. Indeed, the REL and PHEL are the two best bootstrap methods in terms of median bias in all cases, with the latter being always superior. Furthermore, while in the two first experiments the performance of the PHEL bootstrap was affected negatively by the great variability exhibited by EL estimators in those cases, as soon as this problem disappears (experiments 3-5) the PHEL bootstrap becomes the only estimation method which appears mean unbiased. However, the best in terms of median bias are still the CU-GMM, EL and EI estimators.

Table 3.5 presents the results for $n = 500$. There is a significant improvement in the properties of all estimation methods but various points should be noted. First, even for this sample size, the two-step GMM estimator exhibits important biases, particularly in experiment 2. Thus, it seems that it would be necessary to dramatically increase the number of observations to avoid this. Second, the CU-GMM, EI and EL estimators appear even more similar. Their variability is much less for this sample size, so they are now also approximately mean unbiased. Comparing the results obtained for experiments 1 and 2, we can confirm that these methods are relatively indifferent to the addition of worthless instruments, unlike the two-step GMM estimator that continues to present the habitual trade-off between bias and efficiency. With respect to bootstrap techniques, the EL bootstrap method continues to provide no improvement over the two-step GMM estimator, while the estimators based on its post-hoc adjusted version are clearly unbiased in all cases. The NP, RNP, FSEL and REL bootstrap

GMM estimators also perform well, but in experiment 2 still exhibit some bias.

3.6 Conclusion

In this chapter we investigated through some Monte Carlo experiments the finite sample properties of various methods which are theoretically appropriate for the estimation of moment condition models. Two different settings, where two-step GMM is known to produce biased estimators, were considered. Clearly, our results showed that there are better alternatives to estimate both covariance structure and instrumental variable models. Indeed, very promising results were obtained by particularly the PHEL bootstrap and EL estimation methods.

In covariance structure models, apart from the CU-GMM estimator, whose behaviour appeared even worse, all the other seven alternative methods clearly performed better than the two-step GMM estimator in all circumstances. The REL and the PHEL bootstrap methods, suggested in sections 3.3.5 and 3.3.6, produced the best results, leading to the least biased estimators in almost all cases simulated. Moreover, in spite of the usual greater dispersion exhibited by bootstrap estimators, they also behaved very well in terms of RMSE, sharing with the EL method the best performance according to this criterion.

For instrumental variable models, two-step GMM proved again to be completely inadequate, producing very biased estimators for models using large number of instruments, even for moderate sample sizes (500 observations). Also the finite sample properties of the EL bootstrap GMM estimator were not satisfactory, which emphasizes the necessity for a correction of the kind proposed in this chapter. All other estimation methods possess better finite sample properties. The CU-GMM, EI and EL methods behaved in a very similar way, being always nearly unbiased in terms of median bias and also mean unbiased for larger sample sizes. However, in poorly identified models, they exhibited great variability which suggests that some care must be taken in their application in small samples and when there are doubts about the

Table 3.5: Performance of alternative estimators for instrumental variable models (1000 Monte Carlo replications; n = 500)

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
Model 1							
2S-GMM	.001	.015	0.793	1.159	.113	.113	.072
CU-GMM	-.011	.004	0.780	1.153	.118	.118	.073
EI	-.011	.004	0.776	1.155	.118	.118	.073
EL	-.011	.004	0.776	1.155	.118	.118	.073
NP-BOOT-GMM	.001	.015	0.794	1.159	.114	.114	.073
RNP-BOOT-GMM	.001	.015	0.794	1.158	.114	.114	.073
FSEL-BOOT-GMM	.001	.012	0.793	1.159	.114	.114	.072
EL-BOOT-GMM	.013	.027	0.804	1.177	.114	.115	.074
REL-BOOT-GMM	.001	.013	0.790	1.157	.114	.114	.071
PHEL-BOOT-GMM	.001	.015	0.798	1.158	.113	.113	.073
Model 2							
2S-GMM	.088	.093	0.947	1.216	.086	.123	.097
CU-GMM	-.008	.007	0.780	1.157	.118	.118	.075
EI	-.008	.005	0.787	1.160	.122	.123	.078
EL	-.008	.006	0.786	1.160	.122	.123	.078
NP-BOOT-GMM	.024	.034	0.841	1.179	.109	.111	.076
RNP-BOOT-GMM	.024	.035	0.837	1.180	.109	.111	.077
FSEL-BOOT-GMM	.028	.037	0.850	1.180	.106	.110	.077
EL-BOOT-GMM	.096	.102	0.917	1.260	.108	.145	.108
REL-BOOT-GMM	.016	.027	0.824	1.178	.113	.114	.077
PHEL-BOOT-GMM	.001	.013	0.798	1.171	.123	.123	.077
Model 3							
2S-GMM	.018	.019	0.946	1.081	.043	.047	.032
CU-GMM	-.002	-.001	0.921	1.066	.045	.045	.030
EI	-.001	-.001	0.921	1.068	.046	.046	.029
EL	-.001	-.001	0.921	1.068	.046	.046	.029
NP-BOOT-GMM	.001	.003	0.926	1.069	.045	.045	.030
RNP-BOOT-GMM	.001	.003	0.925	1.069	.045	.045	.031
FSEL-BOOT-GMM	.001	.004	0.924	1.070	.046	.046	.030
EL-BOOT-GMM	.019	.022	0.942	1.092	.046	.050	.034
REL-BOOT-GMM	.000	.003	0.922	1.069	.046	.046	.030
PHEL-BOOT-GMM	.000	.002	0.923	1.066	.045	.045	.030
Model 4							
2S-GMM	.012	.013	0.917	1.103	.058	.059	.040
CU-GMM	.001	.003	0.902	1.095	.060	.060	.040
EI	.001	.002	0.898	1.096	.061	.061	.039
EL	.001	.002	0.900	1.098	.061	.061	.039
NP-BOOT-GMM	.002	.004	0.905	1.096	.060	.060	.041
RNP-BOOT-GMM	.002	.003	0.905	1.096	.060	.060	.041
FSEL-BOOT-GMM	.002	.004	0.901	1.098	.060	.060	.041
EL-BOOT-GMM	.012	.012	0.914	1.110	.060	.062	.042
REL-BOOT-GMM	.002	.004	0.904	1.097	.060	.060	.040
PHEL-BOOT-GMM	.002	.004	0.904	1.097	.060	.060	.040
Model 5							
2S-GMM	.026	.028	0.942	1.102	.049	.056	.040
CU-GMM	-.002	-.001	0.907	1.081	.054	.054	.034
EI	-.002	.001	0.908	1.083	.055	.055	.035
EL	-.002	.002	0.909	1.084	.055	.055	.035
NP-BOOT-GMM	.002	.005	0.912	1.086	.053	.054	.036
RNP-BOOT-GMM	.002	.005	0.912	1.086	.053	.054	.035
FSEL-BOOT-GMM	.002	.005	0.913	1.086	.053	.053	.036
EL-BOOT-GMM	.028	.031	0.937	1.112	.054	.061	.044
REL-BOOT-GMM	.001	.004	0.910	1.085	.054	.054	.035
PHEL-BOOT-GMM	.000	.003	0.910	1.083	.054	.054	.034

quality of the instruments. Apart from the experiments where it was affected by the large dispersion of the EL estimator, the PHEL bootstrap worked very well, being the only mean unbiased estimation method in the smallest sample size considered.

We also found that Newey and Smith's (2000) results seem to be a good guide for the small sample behaviour of the non-bootstrap estimators. Thus, a natural and interesting extension of the investigation undertaken in this chapter would be the study of the finite sample properties of bias-corrected GMM estimators based on the expression derived by those authors for the bias of the GMM estimator. Another potential avenue for future research is the analysis of the ability of the bootstrap and Newey and Smith's (2000) theoretical results to correct the bias of CU-GMM and GEL estimators.[8]

3.7 Appendix

The first-order conditions (2.64) defining GEL estimators can also be written as:

$$\sum_{i=1}^{n} \hat{p}_i^{gel} \begin{bmatrix} g_i\left(\hat{\theta}\right) \\ G_i\left(\hat{\theta}\right)' \hat{\phi} \end{bmatrix} = 0.$$

For covariance structure models, $G_i\left(\hat{\theta}\right) = -\iota$. Therefore,

$$\begin{aligned} \sum_{i=1}^{n} \hat{p}_i^{gel} G_i\left(\hat{\theta}\right)' \hat{\phi} &= -\iota' \hat{\phi} \sum_{i=1}^{n} \hat{p}_i^{gel} \\ &= -\iota' \hat{\phi} \\ &= -\sum_{t=1}^{s} \hat{\phi}_t, \end{aligned}$$

and the second first-order condition can be written as $\sum_{t=1}^{s} \hat{\phi}_t = 0$.

[8] A joint paper with R. J. Smith and A. D. Chesher examining some of these issues is currently in preparation.

Furthermore, as

$$\sum_{i=1}^{n} \hat{p}_i^{gel} g\left(\hat{\theta}\right) = \sum_{i=1}^{n} \hat{p}_i^{gel} \left(m_i - \iota\hat{\theta}\right) = 0,$$

then

$$\begin{aligned}
\sum_{i=1}^{n} \hat{p}_i^{gel} m_i &= \iota\hat{\theta} \sum_{i=1}^{n} \hat{p}_i^{gel} \\
\iota\hat{\theta} &= \sum_{i=1}^{n} \hat{p}_i^{gel} m_i \\
\hat{\theta} &= \frac{1}{10} \sum_{t=1}^{10} \sum_{i=1}^{n} \hat{p}_i^{gel} m_{ti}.
\end{aligned}$$

Chapter 4

GEL Pearson-type specification tests

4.1 Introduction

In this chapter we propose Pearson-type statistics suitable to test overidentifying moment conditions and parametric restrictions in models estimated by GEL methods. These new statistics are based on the comparison of two consistent estimators, under the corresponding null hypothesis, of the unknown distribution of the data. For the former class of tests those estimators are the empirical and the GEL distribution functions, while in the latter case two GEL distributions estimated under different assumptions are contrasted. We derive two types of Pearson-type tests. First, we show that the classical Pearson χ^2 statistic is directly applicable in the GEL framework, after minor adaptations. The other approach involves the partition of the sample space into several sets and the contrast between the empirical and the GEL implied probabilities (or two GEL implied probabilities) estimated for each set, which forms the basis for the second Pearson-type statistic we develop.

In the second part of this chapter we investigate, through a Monte Carlo simulation analysis based on two of the settings considered by Imbens, Spady and Johnson (1998), how Pearson-type statistics for overidentifying moment conditions perform in finite samples. We examine their size behaviour and compare it with some of the existing alternatives: Hansen's (1982) J test, cf. section 2.3.4, and the distance metric and Wald statistics discussed in section 2.5.6. In the case of the J test evaluated at the two-step GMM estimator, we consider also bootstrap approximations to its small sample distribution. In particular, five bootstrap techniques already studied in chapter 3 are utilized: the nonparametric (NP), recentered nonparametric (RNP), first-stage GEL (FSGEL), GEL and recentered GEL (RGEL) bootstrap methods. Note that the post-hoc GEL bootstrap, also analyzed in that chapter, is not applicable in this context.

This chapter is organized as follows. Section 4.2 briefly reviews the concept of GEL implied probabilities and formalizes its asymptotic relation to the empirical distribution function. The Pearson-type tests for overidentifying moment conditions are derived in section 4.3 while the case of parametric restrictions is considered in section 4.4. The Monte Carlo simulation studies are discussed in section 4.5. Section 4.6 concludes.

4.2 GEL implied probabilities

Consider the moment conditions

$$E_F\left[g\left(y, \theta_0\right)\right] = 0, \tag{4.1}$$

where the distribution $F \equiv F(y)$ with respect to which the expected value is taken is assumed unknown. As discussed previously, in the GEL context there exists two different ways of consistently estimating F. One of those estimators is the empirical cumulative distribution function,

$$F_n\left(y\right) = \frac{1}{n} \sum_{i=1}^{n} 1\left(y_i \leq y\right), \tag{4.2}$$

which gives constant weights $dF_n(y) = \frac{1}{n}$ to each observation. A more efficient estimator is the GEL distribution,

$$\hat{F}_{gel}(y) = \sum_{i=1}^{n} 1(y_i \leq y)\, p_i\left(\hat{\theta}, \hat{\phi}\right), \tag{4.3}$$

where $p_i(\cdot) \equiv p_i^{gel}(\cdot)$, since it exploits the information contained in (4.1) by reweighting each observation in such a way that the moment conditions are numerically imposed in the sample; cf. section 2.5.4. The weights assigned to each observation, the GEL implied probabilities $d\hat{F}_{gel}(y) = p_i\left(\hat{\theta}, \hat{\phi}\right)$, $i = 1, ..., n$, are estimated using the formula

$$\hat{p}_i \equiv p_i\left(\hat{\theta}, \hat{\phi}\right) = \frac{\pi_i\left(\hat{\theta}, \hat{\phi}\right)}{\sum_{i=1}^{n} \pi_i\left(\hat{\theta}, \hat{\phi}\right)}, \tag{4.4}$$

where $\pi_i\left(\hat{\theta}, \hat{\phi}\right) \equiv \nabla h\left[\hat{\phi}' g\left(y_i, \hat{\theta}\right)\right]$; cf. (2.69) and (2.65). Under the null hypothesis that the moment conditions (4.1) hold in the population of interest, the probability limit of $\hat{\phi}$ is 0, so $\pi_i\left(\hat{\theta}, 0\right) = \nabla h(0)$, and, hence, the GEL probabilities $p_i\left(\hat{\theta}, 0\right)$, $i = 1, ..., n$, are equal to the empirical measures $dF_n(y) = \frac{1}{n}$.

More rigorously, let $\hat{g}_i \equiv g\left(y_i, \hat{\theta}\right)$ and $b_i \equiv \sup_{\theta \in \Theta} \|g(y_i, \theta)\|$. From Newey and Smith (2001, Proof of Lemma A1), $\max_{1 \leq i \leq n} b_i \equiv O_p\left(n^{\frac{1}{\alpha}}\right)$, where $\alpha > 2$ is such that $E\left[\sup_{\theta \in \Theta} \|g(y_i, \theta)\|^{\alpha}\right] < \infty$. A second-order expansion Taylor series expansion for $\pi_i\left(\hat{\theta}, \hat{\phi}\right)$ yields

$$\nabla h\left(\hat{\phi}' \hat{g}_i\right) = \nabla h(0) + \nabla^2 h(0)\, \hat{\phi}' \hat{g}_i + \frac{1}{2} \nabla^3 h\left(\dot{\phi}' \hat{g}_i\right)\left(\hat{\phi}' \hat{g}_i\right)^2, \tag{4.5}$$

where $0 < \dot{\phi} < \hat{\phi}$. Now, $\max_{1 \leq i \leq n} \left|\nabla^3 h\left(\dot{\phi}' \hat{g}_i\right) - \nabla^3 h(0)\right| \xrightarrow{p} 0$ as $\sup_{\theta \in \Theta, 1 \leq i \leq n} \left|\dot{\phi}' g(y_i, \theta)\right| \xrightarrow{p} 0$, see Newey and Smith (2001, Proof of Lemma A1). Therefore,

$$\nabla^3 h\left(\dot{\phi}' \hat{g}_i\right)\left(\hat{\phi}' \hat{g}_i\right)^2 = O_p\left(n^{-\left(1 - \frac{2}{\alpha}\right)}\right), \tag{4.6}$$

as $\hat{\phi} = O_p\left(n^{-\frac{1}{2}}\right)$ and $\|\hat{g}_i\| = O_p\left(n^{\frac{1}{\alpha}}\right)$. On the other hand, a first-order Taylor expansion gives

$$\frac{1}{\sum_{i=1}^{n}\nabla h\left(\dot{\phi}'\hat{g}_i\right)} = \frac{1}{n\nabla h(0)} - \frac{1}{n\left[\nabla h(0)\right]^2}\frac{1}{n}\sum_{i=1}^{n}\nabla^2 h\left(\dot{\phi}'\hat{g}_i\right)\hat{g}_i'\hat{\phi}$$

$$= \frac{1}{n\nabla h(0)}\left[1+O_p\left(n^{-1}\right)\right], \quad (4.7)$$

as $\max_{1\le i\le n}\left|\nabla^2 h\left(\dot{\phi}'\hat{g}_i\right)-\nabla^2 h(0)\right| \xrightarrow{p} 0$ and $\frac{1}{n}\sum_{i=1}^{n} g\left(y_i,\hat{\theta}\right) = O_p\left(n^{-\frac{1}{2}}\right)$.

Hence,

$$\hat{p}_i = \left[\nabla h(0) + \nabla^2 h(0)\,\hat{\phi}'\hat{g}_i + \frac{1}{2}\nabla^3 h\left(\dot{\phi}'\hat{g}_i\right)\left(\hat{\phi}'\hat{g}_i\right)^2\right]\frac{1}{n\nabla h(0)}\left[1+O_p\left(n^{-1}\right)\right]$$

$$= \frac{1}{n} + \frac{1}{n}\frac{\nabla^2 h(0)}{\nabla h(0)}\hat{\phi}'\hat{g}_i + \frac{1}{2}\frac{\nabla^3 h\left(\dot{\phi}'\hat{g}_i\right)}{n\nabla h(0)}\left(\hat{\phi}'\hat{g}_i\right)^2 + O_p\left(n^{-2}\right) \quad (4.8)$$

and, using (4.6),

$$\sqrt{n}\left(\hat{p}_i - \frac{1}{n}\right) = \frac{\nabla^2 h(0)}{\nabla h(0)}\frac{1}{n}\hat{g}_i'\sqrt{n}\hat{\phi} + O_p\left(n^{-\left(\frac{3}{2}-\frac{2}{\alpha}\right)}\right). \quad (4.9)$$

Equations (4.8) and (4.9), by expressing the asymptotic relationship between the empirical and GEL probability density functions, form the basis for the construction of the Pearson-type test statistics derived in the next sections.

4.3 Tests of overidentifying moment conditions

In this section we develop two classes of Pearson-type test statistics appropriate for testing the moment conditions (4.1). First, we show that a very simple adaptation of the standard Pearson statistic utilized in the parametric context allows its employment in the GEL framework as a test of overidentifying moment conditions. Then, we derive an alternative Pearson-type statistic which is based on the contrast between empirical and GEL probabilities estimated for each set into which the sample space is divided.

4.3.1 Classical Pearson statistics

Suppose that we have a dataset containing some ties, where the distinct value y_i arises $n_i \geq 1$ times. Let u be the number of ties. In a parametric context, we may wish to test whether a given distribution function $\bar{F}(y)$ correctly describes the data. To this end, there are two versions of the Pearson statistic that are usually applied:

$$P_1^* = \sum_{i=1}^{u} \frac{(e_i - n_i)^2}{n_i} \tag{4.10}$$

and

$$P_2^* = \sum_{i=1}^{u} \frac{(e_i - n_i)^2}{e_i}, \tag{4.11}$$

where n_i and $e_i \equiv n \cdot d\bar{F}(y_i)$ denote, respectively, the actual and the expected number of observations of the distinct value y_i, $i = 1, ..., u$, under $\bar{F}(y)$. In (4.11) it is assumed that $e_i > 0$ for all $i = 1, ..., u$. Both statistics have a limiting chi-square distribution when $\bar{F}(y)$ is indeed the true distribution of the data.

In the GEL framework, we can ignore the ties in the data and deal with the probability associated with an observation, not a value; see *inter alia* Owen (2001). In other words, we can act as if a single data point was observed in each cell of a n-cell contingency table, that is, a GEL version of the above statistics may be directly obtained by setting $n_i = 1$, $u = n$ and $e_i = n\hat{p}_i$, $i = 1, ..., n$. In fact, as we show next, the corresponding versions of (4.10) and (4.11) that allow the hypothesis (4.1) to be tested in models estimated by GEL methods are given by

$$P_1 = \sum_{i=1}^{n} (n\hat{p}_i - 1)^2 \tag{4.12}$$

and

$$P_2 = \sum_{i=1}^{n} \frac{(n\hat{p}_i - 1)^2}{n\hat{p}_i}, \tag{4.13}$$

which have a limiting chi-square distribution with $s - k$ degrees of freedom.[1] Note

[1] These Pearson statistics could also be used as distance metric in (2.48) and, therefore, be

that, from (4.8), it follows that $n\hat{p}_i = 1 + O_p\left(n^{-\left(\frac{1}{2}-\frac{1}{\alpha}\right)}\right)$, so (4.12) and (4.13) are asymptotically equivalent.

To demonstrate that these statistics are appropriated for testing the moment conditions (4.1), we show the asymptotic equivalence of P_1 to a Wald test of overidentifying moment conditions. The proof is very simple. In fact, from (4.8), and using (4.6), it follows that

$$\begin{aligned}
(n\hat{p}_i - 1)^2 &= \left[\frac{\nabla^2 h(0)}{\nabla h(0)}\hat{\phi}'\hat{g}_i + \frac{1}{2}\frac{\nabla^3 h\left(\dot{\phi}'\hat{g}_i\right)}{\nabla h(0)}\left(\hat{\phi}'\hat{g}_i\right)^2 + O_p\left(n^{-1}\right)\right]^2 \\
&= \left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2\left(\hat{\phi}'\hat{g}_i\right)^2 + \frac{\nabla^2 h(0)}{\nabla h(0)}\hat{\phi}'\hat{g}_i O_p\left(n^{-\left(1-\frac{2}{\alpha}\right)}\right) + \\
&\quad \frac{1}{4}\frac{\nabla^3 h\left(\dot{\phi}'\hat{g}_i\right)}{\nabla h(0)}\left(\hat{\phi}'\hat{g}_i\right)^2 O_p\left(n^{-\left(1-\frac{2}{\alpha}\right)}\right).
\end{aligned}$$

Summing over $i = 1, ..., n$,

$$\begin{aligned}
\sum_{i=1}^{n}(n\hat{p}_i - 1)^2 &= n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2\hat{\phi}'\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'\hat{\phi} + \frac{\nabla^2 h(0)}{\nabla h(0)}\sqrt{n}\hat{\phi}'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{g}_i O_p\left(n^{-\left(1-\frac{2}{\alpha}\right)}\right) + \\
&\quad \frac{1}{4}n\hat{\phi}'\frac{1}{n}\left[\sum_{i=1}^{n}\frac{\nabla^3 h\left(\dot{\phi}'\hat{g}_i\right)}{\nabla h(0)}\hat{g}_i\hat{g}_i'\right]\hat{\phi}O_p\left(n^{-\left(1-\frac{2}{\alpha}\right)}\right) \\
&= n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2\hat{\phi}'\hat{V}_n\hat{\phi} + O_p\left(n^{-\left(1-\frac{2}{\alpha}\right)}\right),
\end{aligned}$$

as $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{g}_i = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n}\left[\nabla^3 h\left(\dot{\phi}'\hat{g}_i\right)\hat{g}_i\hat{g}_i' - \nabla^3 h(0)\right] \xrightarrow{p} 0$. Hence, $P_1 = W_n + O_p\left(n^{-\left(1-\frac{2}{\alpha}\right)}\right)$,where W_n denotes the Wald test statistic of overidentifying moment conditions given in (2.79).

applied to produce estimators for the parameters of interest in moment condition models. Actually, the optimization of the program (2.48) based on P_1 and P_2 would yield the same estimators as those produced by the choices $\mathcal{M}_1(F_{md}, F_n)$ and $\mathcal{M}_1(F_n, F_{md})$ in the Cressie-Read statistic (2.49), respectively. Note that (4.13) may be written as $P_2 = \sum_{i=1}^{n}\left(\frac{1}{n\hat{p}_i} - 1\right)$.

4.3.2 Alternative Pearson-type tests

In this sub-section we develop an alternative Pearson-type test of overidentifying moment conditions. As discussed in section 4.2, the distribution F in (4.1) can be consistently estimated, under the hypothesis that those moment conditions hold in the population of interest, by either $F_n(y)$ of (4.2) or $\hat{F}_{gel}(y)$ of (4.3). Therefore, we can think of testing the validity of the overidentifying moment conditions (4.1) by testing for $H_0 : \hat{F}_{gel}(y) - F_n(y) = 0$. Indeed, if the null model is correctly specified, the limiting distribution of a test statistic based on the contrast $\hat{F}_{gel}(y) - F_n(y)$ should be centred at zero.

Derivation

Consider a first-order Taylor series expansion of $\sqrt{n}\hat{F}_{gel}(y)$ around $\phi = 0$:

$$\sqrt{n}\hat{F}_{gel}(y) = \sqrt{n}F_n(y) + \sum_{i=1}^{n} 1(y_i \leq y) \frac{\partial p_i\left(\hat{\theta}, 0\right)'}{\partial \phi'} \sqrt{n}\hat{\phi} + O_p\left(n^{-\frac{1}{2}}\right) \qquad (4.14)$$

As

$$\frac{\partial p_i(\theta, \phi)}{\partial \phi'} = \frac{\nabla \pi_i(\theta, \phi) g(y_i, \theta) \sum_{i=1}^{n} \pi_i(\theta, \phi) - \pi_i(\theta, \phi) \sum_{i=1}^{n} \nabla \pi_i(\theta, \phi) g(y_i, \theta)}{\left[\sum_{i=1}^{n} \pi_i(\theta, \phi)\right]^2} \qquad (4.15)$$

and, thus,

$$\begin{aligned} \frac{\partial p_i\left(\hat{\theta}, 0\right)}{\partial \phi'} &= \frac{\nabla^2 h(0)}{\nabla h(0)} \frac{1}{n} \left[g\left(y_i, \hat{\theta}\right)' - \frac{1}{n} \sum_{i=1}^{n} g\left(y_i, \hat{\theta}\right)' \right] \\ &= \frac{\nabla^2 h(0)}{\nabla h(0)} \frac{1}{n} g\left(y_i, \hat{\theta}\right)', \end{aligned} \qquad (4.16)$$

since $\frac{1}{n}\sum_{i=1}^{n} g\left(y_i, \hat{\theta}\right) = O_p\left(n^{-\frac{1}{2}}\right)$, it follows that

$$\sqrt{n}\hat{F}_{gel}(y) = \sqrt{n}F_n(y) + \frac{\nabla^2 h(0)}{\nabla h(0)} \frac{1}{n} \sum_{i=1}^{n} 1(y_i \leq y)\, g\left(y_i, \hat{\theta}\right)' \sqrt{n}\hat{\phi}$$

$$
\begin{aligned}
& +O_p\left(n^{-\frac{1}{2}}\right) \\
= \quad & \sqrt{n}F_n(y) + \frac{\nabla^2 h(0)}{\nabla h(0)} b' \sqrt{n}\hat{\phi} + O_p\left(n^{-\frac{1}{2}}\right), \qquad (4.17)
\end{aligned}
$$

where the s-vector $b \equiv E_F\left[1\left(y_i \leq y\right) g\left(y_i, \theta_0\right)\right]$ is assumed to be nonzero. Moreover, as $\sqrt{n}\hat{\phi} = -\frac{\nabla h(0)}{\nabla^2 h(0)} V^{-1} M \sqrt{n} g_n\left(\theta_0\right) + O_p\left(n^{-\frac{1}{2}}\right)$, see expression (2.66), equation (4.17) can be written as:

$$
\sqrt{n}\left[\hat{F}_{gel}(y) - F_n(y)\right] = -b' V^{-1} M \sqrt{n} g_n\left(\theta_0\right) + O_p\left(n^{-\frac{1}{2}}\right). \qquad (4.18)
$$

Now, consider a partition of the sample space of y into the sets C_j, $j = 1, ..., L$, where L is finite. Define

$$
F_n\left(C_j\right) = \frac{1}{n}\sum_{i=1}^{n} 1\left(y_i \in C_j\right) \qquad (4.19)
$$

and

$$
\hat{F}_{gel}\left(C_j\right) = \sum_{i=1}^{n} 1\left(y_i \in C_j\right) p_i\left(\hat{\theta}, \hat{\phi}\right). \qquad (4.20)
$$

Using a similar argument to that above, we have, corresponding to (4.18),

$$
\sqrt{n}\left[\hat{F}_{gel}\left(C_j\right) - F_n\left(C_j\right)\right] = -b_j' V^{-1} M \sqrt{n} g_n\left(\theta_0\right) + O_p\left(n^{-\frac{1}{2}}\right), \qquad (4.21)
$$

$j = 1, ..., L$. Stacking $B \equiv \left(b_1, ..., b_L\right)$, an $(s \times L)$ matrix, and

$$
\hat{F}_{gel} - F_n \equiv \begin{bmatrix} \hat{F}_{gel}\left(C_1\right) - F_n\left(C_1\right) \\ \dots \\ \hat{F}_{gel}\left(C_L\right) - F_n\left(C_L\right) \end{bmatrix}, \qquad (4.22)
$$

an L-vector, it follows that

$$
\sqrt{n}\left(\hat{F}_{gel} - F_n\right) = -B' V^{-1} M \sqrt{n} g_n\left(\theta_0\right) + O_p\left(n^{-\frac{1}{2}}\right). \qquad (4.23)
$$

Hence, under the null hypothesis that the moment conditions (4.1) hold in the population,

$$\sqrt{n}\left(\hat{F}_{gel} - F_n\right) \xrightarrow{d} N(0, \Psi), \tag{4.24}$$

where $\Psi \equiv B'M'V^{-1}MB$. Thus,

$$n\left(\hat{F}_{gel} - F_n\right)' \Psi_n^- \left(\hat{F}_{gel} - F_n\right) \xrightarrow{d} \chi^2_v, \tag{4.25}$$

where $\Psi_n^- \equiv \left(\hat{B}_n'\hat{M}_n'\hat{V}_n^{-1}\hat{M}_n\hat{B}_n\right)^-$ denotes a consistent estimator for a *g-inverse* of Ψ, $\hat{B}_n$, $\hat{V}_n$ and $\hat{M}_n$ are consistent estimators for B, V and M, respectively, and $v = rk\left(B'M'V^{-1}MB\right)$.

Let $L \geq s$ and assume that B is full row rank s. Then, $v = s - k$ and a generalized inverse for Ψ is $B'\left(BB'\right)^{-1} V \left(BB'\right)^{-1} B$. Therefore, the Pearson-type test statistic proposed in this section is given by

$$P_3 = n\left(\hat{F}_{gel} - F_n\right)' \hat{B}_n' \left(\hat{B}_n\hat{B}_n'\right)^{-1} \hat{V}_n \left(\hat{B}_n\hat{B}_n'\right)^{-1} \hat{B}_n \left(\hat{F}_{gel} - F_n\right) \xrightarrow{d} \chi^2_{s-k}. \tag{4.26}$$

If $L = s$, the matrix B will be invertible and this test statistic can be simplified to

$$P_3 = n\left(\hat{F}_{gel} - F_n\right)' \hat{B}_n^{-1}\hat{V}_n\hat{B}_n^{-1} \left(\hat{F}_{gel} - F_n\right) \xrightarrow{d} \chi^2_{s-k}. \tag{4.27}$$

Asymptotic equivalence to alternative tests

In this sub-section we show that the Pearson-type test statistic P_3 above developed is asymptotically equivalent to all the other GEL tests of overidentifying moment conditions discussed before. First, note that (4.23) can be rewritten both as

$$\sqrt{n}\left(\hat{F}_{gel} - F_n\right) = -B'V^{-1}\sqrt{n}g_n\left(\hat{\theta}\right) + O_p\left(n^{-\frac{1}{2}}\right) \tag{4.28}$$

and

$$\sqrt{n}\left(\hat{F}_{gel} - F_n\right) = -\frac{\nabla^2 h(0)}{\nabla h(0)} B'\sqrt{n}\hat{\phi} + O_p\left(n^{-\frac{1}{2}}\right). \tag{4.29}$$

Expression (4.28) follows from a Taylor series expansion of $\sqrt{n}g_n\left(\hat{\theta}\right)$ around $\sqrt{n}g_n\left(\theta_0\right)$, $\sqrt{n}g_n\left(\hat{\theta}\right) = \sqrt{n}g_n\left(\theta_0\right) + G\sqrt{n}\left(\hat{\theta}-\theta_0\right) + O_p\left(n^{-\frac{1}{2}}\right)$, where $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ is replaced by $-\Sigma G'V^{-1}\sqrt{n}g_n\left(\theta_0\right)$, see (2.66), yielding

$$\sqrt{n}g_n\left(\hat{\theta}\right) = M\sqrt{n}g_n\left(\theta_0\right) + O_p\left(n^{-\frac{1}{2}}\right). \tag{4.30}$$

To obtain (4.29), note that $\sqrt{n}\frac{\nabla^2 h(0)}{\nabla h(0)}\hat{\phi} = -V^{-1}M\sqrt{n}g_n\left(\theta_0\right) + O_p\left(n^{-\frac{1}{2}}\right)$, see also (2.66).

Using (4.25) and (4.28), we can demonstrate the asymptotic equivalence of the P_3 and J tests. Indeed, substituting the latter expression into the former, we obtain:

$$P_3 = n\hat{g}_n'\hat{V}_n^{-1}\hat{B}_n\left(\hat{B}_n'\hat{M}_n'\hat{V}_n^{-1}\hat{M}_n\hat{B}_n\right)^{-}\hat{B}_n'\hat{V}_n^{-1}\hat{g}_n + O_p\left(n^{-\frac{1}{2}}\right). \tag{4.31}$$

Following Lemma 2.2.5d) of Rao and Mitra (1971), $\hat{B}_n\left(\hat{B}_n'\hat{M}_n'\hat{V}_n^{-1}\hat{M}_n\hat{B}_n\right)^{-}\hat{B}_n'$ is a generalized inverse for $\hat{M}_n'\hat{V}_n^{-1}\hat{M}_n$, since $rk\left(\hat{B}_n'\hat{M}_n'\hat{V}_n^{-1}\hat{M}_n\hat{B}_n\right) = rk\left(\hat{M}_n'\hat{V}_n^{-1}\hat{M}_n\right)$. Thus, as $\hat{V}_n$ is just another generalized inverse for $\hat{M}_n'\hat{V}_n^{-1}\hat{M}_n$, it follows that

$$\begin{aligned} P_3 &= n\hat{g}_n'\hat{V}_n^{-1}\hat{g}_n + O_p\left(n^{-\frac{1}{2}}\right) \\ &= J_n + O_p\left(n^{-\frac{1}{2}}\right). \end{aligned} \tag{4.32}$$

Similarly, substituting (4.29) into (4.25) and applying the same Lemma of Rao and Mitra (1971), the asymptotic equivalence of the Pearson-type test to the Wald test presented in (2.79) (and, hence, to the Pearson statistics P_1 and P_2) is proven:

$$\begin{aligned} P_3 &= n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2 \hat{\phi}'\hat{V}_n\hat{\phi} + O_p\left(n^{-\frac{1}{2}}\right) \\ &= W_n + O_p\left(n^{-\frac{1}{2}}\right). \end{aligned} \tag{4.33}$$

The asymptotic equivalence of the Pearson-type test to the distance metric test of (2.78) can be shown by demonstrating the equivalence of the latter to the Wald statistic; see Smith (1997, pp. 510-511) for a proof.

4.4 Tests of parametric restrictions

The same principles used to construct Pearson-type tests of overidentifying moment conditions can be applied in other contexts. In this section we show how to develop GEL Pearson-type statistics appropriate for testing parametric restrictions.

4.4.1 Constrained GEL estimation

Consider the null hypothesis

$$H_0 : r(\theta_0) = 0, \tag{4.34}$$

where $r(\cdot)$ is a known continuously differentiable q-vector of parametric restrictions, where $q < k$. The $(q \times k)$ derivative matrix $R(\theta) \equiv \nabla_\theta r(\theta)$ is assumed full row rank q. Let $\left(\hat{\theta}, \hat{\phi}\right)$ be the unconstrained estimators resulting from the optimization of the GEL criterion $Q_{gel}(\theta, \phi) = h[\phi' g(y_i, \theta)]$ and $\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right)$ the estimators of the constrained model incorporating H_0, which are obtained by optimizing the modified GEL function $Q^*_{gel}(\theta, \phi, \psi) = h[\phi' g(y_i, \theta) + \psi' r(\theta)]$; see Smith (1997) and section 2.5.6. Define

$$\tilde{p}^*_i \equiv p^*_i\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right) = \frac{\pi_i\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right)}{\sum_{i=1}^n \pi_i\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right)}, \tag{4.35}$$

where $\pi_i\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right) \equiv \nabla h\left[\tilde{\phi}' g\left(y_i, \tilde{\theta}\right) + \tilde{\psi}' r\left(\tilde{\theta}\right)\right]$, $i = 1, ..., n$.

In this setting, assuming that the moment conditions (4.1) hold in the population, the empirical distribution function $F_n(y)$ and the unconstrained GEL distribution $\hat{F}_{gel}(y)$ are still consistent estimators of the distribution F in (4.1), whether or not H_0 (4.34) holds. However, under H_0 (4.34), a more efficient estimator is given by

$$\tilde{F}^*_{gel}(y) = \sum_{i=1}^n 1(y_i \leq y)\, p^*_i\left(\tilde{\theta}, \tilde{\phi}, \tilde{\psi}\right). \tag{4.36}$$

The statistics suggested below for testing the parametric restrictions (4.34) are based on the contrasts $\hat{F}_{gel}(y) - \tilde{F}^*_{gel}(y)$ or $\hat{p}_i - \tilde{p}^*_i$, $i = 1, ..., n$. Before presenting them, we derive in the remaining of this sub-section the asymptotic relationship that occurs between the GEL implied probabilities $\hat{p}_i$ and $\tilde{p}_i^*$, $i = 1, ..., n$.

Consider a Taylor expansion of $\sqrt{n}\tilde{p}_i^*$ about $\left(\tilde{\theta}, 0, 0\right)$:

$$\sqrt{n}\tilde{p}_i^* = \sqrt{n}p_i^*\left(\tilde{\theta}, 0, 0\right) + \frac{\partial p_i^*\left(\tilde{\theta}, 0, 0\right)'}{\partial \phi'}\sqrt{n}\tilde{\phi} + \frac{\partial p_i^*\left(\tilde{\theta}, 0, 0\right)'}{\partial \psi'}\sqrt{n}\tilde{\psi} + O_p\left(n^{-\frac{3}{2}}\right). \quad (4.37)$$

As

$$\frac{\partial p_i^*\left(\theta, \phi, \psi\right)}{\partial \phi'} = \frac{\nabla \pi_i^*\left(\theta, \phi, \psi\right) g\left(y_i, \theta\right) \sum_{i=1}^n \pi_i^*\left(\theta, \phi, \psi\right) - \pi_i^*\left(\theta, \phi, \psi\right) \sum_{i=1}^n \nabla \pi_i^*\left(\theta, \phi, \psi\right) g\left(y_i, \theta\right)}{\left[\sum_{i=1}^n \pi_i^*\left(\theta, \phi, \psi\right)\right]^2}$$

and

$$\frac{\partial p_i^*\left(\theta, \phi, \psi\right)}{\partial \psi'} = \frac{\nabla \pi_i^*\left(\theta, \phi, \psi\right) \sum_{i=1}^n \pi_i^*\left(\theta, \phi, \psi\right) - \pi_i^*\left(\theta, \phi, \psi\right) \sum_{i=1}^n \nabla \pi_i^*\left(\theta, \phi, \psi\right)}{\left[\sum_{i=1}^n \pi_i^*\left(\theta, \phi, \psi\right)\right]^2} r\left(\theta\right),$$

it follows that, since $\pi_i^*\left(\theta, \phi, 0\right) = \pi_i\left(\theta, \phi\right)$,

$$\begin{aligned} \frac{\partial p_i^*\left(\tilde{\theta}, 0, 0\right)}{\partial \phi'} &= \frac{\nabla \pi_i\left(\tilde{\theta}, 0\right) g\left(y_i, \tilde{\theta}\right) \sum_{i=1}^n \pi_i\left(\tilde{\theta}, 0\right) - \pi_i\left(\tilde{\theta}, 0\right) \sum_{i=1}^n \nabla \pi_i\left(\tilde{\theta}, 0\right) g\left(y_i, \tilde{\theta}\right)}{\left[\sum_{i=1}^n \pi_i\left(\tilde{\theta}, 0\right)\right]^2} \\ &= \frac{\partial p_i\left(\tilde{\theta}, 0\right)}{\partial \phi'}, \end{aligned} \quad (4.38)$$

and

$$\begin{aligned} \frac{\partial p_i^*\left(\tilde{\theta}, 0, 0\right)}{\partial \psi'} &= \frac{\nabla^2 h\left(0\right) \sum_{i=1}^n \nabla h\left(0\right) - \nabla h\left(0\right) \sum_{i=1}^n \nabla^2 h\left(0\right)}{\left[\sum_{i=1}^n \nabla h\left(0\right)\right]^2} r\left(\theta\right) \\ &= \frac{n\nabla^2 h\left(0\right) \nabla h\left(0\right) - n\nabla h\left(0\right) \nabla^2 h\left(0\right)}{\left[n\nabla h\left(0\right)\right]^2} r\left(\theta\right) \\ &= 0. \end{aligned} \quad (4.39)$$

Noting that $p_i^*\left(\tilde{\theta},0,0\right)=\frac{1}{n}$ and substituting (4.38) and (4.39) into (4.37) yields:

$$\sqrt{n}\left(\tilde{p}_i^*-\frac{1}{n}\right)=\frac{\nabla^2 h\left(0\right)}{\nabla h\left(0\right)}\frac{1}{n}g\left(y_i,\tilde{\theta}\right)'\sqrt{n}\tilde{\phi}+O_p\left(n^{-\frac{3}{2}}\right). \tag{4.40}$$

Finally, subtracting (4.40) from (4.9), with $\tilde{\theta}$ and $\hat{\theta}$ replaced by θ_0, produces:

$$\sqrt{n}\left(\hat{p}_i-\tilde{p}_i^*\right)=\frac{\nabla^2 h\left(0\right)}{\nabla h\left(0\right)}\frac{1}{n}g\left(y_i,\theta_0\right)'\sqrt{n}\left(\hat{\phi}-\tilde{\phi}\right)+O_p\left(n^{-\frac{3}{2}}\right). \tag{4.41}$$

4.4.2 Classical Pearson tests

In the present framework, classical-type Pearson statistics for testing (4.34), similar to those derived in section 4.3.1, can be constructed. Corresponding to P_1 (4.12), we propose the statistic

$$P_1^{pr}=\sum_{i=1}^{n}\left(n\hat{p}_i-n\tilde{p}_i^*\right)^2, \tag{4.42}$$

which has a limiting chi-square distribution with q degrees of freedom. Indeed, from (4.41) it follows that

$$\begin{aligned}
n\left(\hat{p}_i-\tilde{p}_i^*\right)^2 &= \left[\frac{\nabla^2 h\left(0\right)}{\nabla h\left(0\right)}\right]^2\frac{1}{n}\left(\hat{\phi}-\tilde{\phi}\right)'g\left(y_i,\theta_0\right)g\left(y_i,\theta_0\right)'\left(\hat{\phi}-\tilde{\phi}\right)+O_p\left(n^{-\frac{5}{2}}\right)\\
\sum_{i=1}^{n}n\left(\hat{p}_i-\tilde{p}_i^*\right)^2 &= \left[\frac{\nabla^2 h\left(0\right)}{\nabla h\left(0\right)}\right]^2\left(\hat{\phi}-\tilde{\phi}\right)'\frac{1}{n}\sum_{i=1}^{n}g\left(y_i,\theta_0\right)g\left(y_i,\theta_0\right)'\left(\hat{\phi}-\tilde{\phi}\right)+O_p\left(n^{-\frac{3}{2}}\right)\\
P_1^{pr} &= n\left[\frac{\nabla^2 h\left(0\right)}{\nabla h\left(0\right)}\right]^2\left(\hat{\phi}-\tilde{\phi}\right)'\hat{V}_n\left(\hat{\phi}-\tilde{\phi}\right)+O_p\left(n^{-\frac{1}{2}}\right)\\
&= MC_n+O_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where MC_n denotes the minimum chi-square statistic of parametric restrictions given in (2.97).

Corresponding to P_2, we have the following two alternatives:

$$P_{2a}^{pr}=\sum_{i=1}^{n}\frac{\left(n\hat{p}_i-n\tilde{p}_i^*\right)^2}{n\hat{p}_i} \tag{4.43}$$

and

$$P_{2b}^{pr} = \sum_{i=1}^{n} \frac{\left(n\hat{p}_i - n\tilde{p}_i^*\right)^2}{n\tilde{p}_i}. \tag{4.44}$$

Both statistics also have a limiting chi-square distribution with q degrees of freedom. The proof is similar to that presented above for P_1^{pr} since, from (4.9), $n\hat{p}_i = 1 + O_p\left(n^{-\left(\frac{3}{2}-\frac{2}{\alpha}\right)}\right)$ and, from (4.40), $n\tilde{p}_i^* = 1 + O_p\left(n^{-\frac{1}{2}}\right)$.

4.4.3 Alternative Pearson-type tests

By analogy with the overidentifying moment conditions case, a test statistic based on the normalized contrast $\sqrt{n}\left[\hat{F}_{gel}(y) - \tilde{F}_{gel}^*(y)\right]$ constitutes an alternative way of assessing the hypothesis H_0 (4.34). Expanding $\tilde{F}_{gel}^*(y)$ about $\left(\tilde{\theta}, 0, 0\right)$ yields:

$$\begin{aligned}
\sqrt{n}\tilde{F}_{gel}^*(y) &= \sqrt{n}F_n(y) + \sum_{i=1}^{n} 1(y_i \le y) \frac{\partial p_i^*\left(\tilde{\theta}, 0, 0\right)'}{\partial \phi'} \sqrt{n}\tilde{\phi} + \\
&\quad + \sum_{i=1}^{n} 1(y_i \le y) \frac{\partial p_i^*\left(\tilde{\theta}, 0, 0\right)'}{\partial \psi'} \sqrt{n}\tilde{\psi} + O_p\left(n^{-\frac{1}{2}}\right) \\
&= \sqrt{n}F_n(y) + \frac{\nabla^2 h(0)}{\nabla h(0)} \frac{1}{n} \sum_{i=1}^{n} 1(y_i \le y)\, g\left(y_i, \tilde{\theta}\right)' \sqrt{n}\tilde{\phi} + \\
&\quad + O_p\left(n^{-\frac{1}{2}}\right);
\end{aligned} \tag{4.45}$$

see (4.38) and (4.39). Hence, using the same notation as in sub-section 4.3.2, we have:

$$\sqrt{n}\left[\tilde{F}_{gel}^*(y) - F_n(y)\right] = \frac{\nabla^2 h(0)}{\nabla h(0)} b' \sqrt{n}\tilde{\phi} + O_p\left(n^{-\frac{1}{2}}\right). \tag{4.46}$$

Subtracting (4.46) from (4.17) produces:

$$\sqrt{n}\left[\hat{F}_{gel}(y) - \tilde{F}_{gel}^*(y)\right] = \frac{\nabla^2 h(0)}{\nabla h(0)} b' \sqrt{n}\left(\hat{\phi} - \tilde{\phi}\right) + O_p\left(n^{-\frac{1}{2}}\right). \tag{4.47}$$

Finally, as $\sqrt{n}\hat{\phi} = -\frac{\nabla h(0)}{\nabla^2 h(0)}\left(V^{-1} - V^{-1}G\Sigma G'V^{-1}\right)\sqrt{n}g_n(\theta_0) + O_p\left(n^{-\frac{1}{2}}\right)$, see (2.66), and $\sqrt{n}\tilde{\phi} = -\frac{\nabla h(0)}{\nabla^2 h(0)}\left(V^{-1} - V^{-1}G\Sigma PG'V^{-1}\right)\sqrt{n}g_n(\theta_0) + O_p\left(n^{-\frac{1}{2}}\right)$, see (2.88), it follows that

$$\sqrt{n}\left(\hat{\phi}-\tilde{\phi}\right)=\frac{\nabla h\left(0\right)}{\nabla^{2}h\left(0\right)}V^{-1}G\Sigma\left(I-P\right)G'V^{-1}\sqrt{n}g_{n}\left(\theta_{0}\right)+O_{p}\left(n^{-\frac{1}{2}}\right) \tag{4.48}$$

and, therefore,

$$\sqrt{n}\left[\hat{F}_{gel}\left(y\right)-\tilde{F}_{gel}^{*}\left(y\right)\right]=b'V^{-1}G\Sigma\left(I-P\right)G'V^{-1}\sqrt{n}g_{n}\left(\theta_{0}\right)+O_{p}\left(n^{-\frac{1}{2}}\right). \tag{4.49}$$

Now, consider a partition of the sample space of y into the sets C_j, $j=1,...,L$, identical to that of the previous section. Define $\tilde{F}_{gel}^{*}\left(C_j\right)=\sum_{i=1}^{n}1\left(y_i\in C_j\right)\tilde{p}_i^{*}$ and

$$\hat{F}_{gel}-\tilde{F}_{gel}^{*}\equiv\begin{bmatrix}\hat{F}_{gel}\left(C_1\right)-\tilde{F}_{gel}^{*}\left(C_1\right)\\ \cdots\\ \hat{F}_{gel}\left(C_L\right)-\tilde{F}_{gel}^{*}\left(C_L\right)\end{bmatrix}. \tag{4.50}$$

Then, noting that $\Sigma\left(I-P\right)\Sigma^{-1}G\left(I-P\right)'\Sigma=\Sigma\left(I-P\right)$, it follows from (4.49) that

$$\sqrt{n}\left(\hat{F}_{gel}-\tilde{F}_{gel}^{*}\right)\xrightarrow{d}N\left(0,\Psi\right), \tag{4.51}$$

where $\Psi\equiv B'V^{-1}G\Sigma\left(I-P\right)G'V^{-1}B$. Thus, a Pearson-type statistic for testing the parametric restrictions (4.34) is given by

$$P_{3}^{pr}=n\left(\hat{F}_{gel}-\tilde{F}_{gel}^{*}\right)'\Psi_{n}^{-}\left(\hat{F}_{gel}-\tilde{F}_{gel}^{*}\right)\xrightarrow{d}\chi_{q}^{2}, \tag{4.52}$$

where Ψ_n^{-} denotes a consistent estimator for a *g-inverse* of Ψ. Assuming that B is full row rank s, a generalized inverse for Ψ is $B'\left(BB'\right)^{-1}V\left(BB'\right)^{-1}B$. In the case that B is a square matrix ($L=s$), a generalized inverse for Ψ is simply $B^{-1}VB^{-1}$.

4.5 Finite sample properties of tests of overidentifying moment conditions: Monte Carlo investigation

In this section we investigate the finite sample properties of some of the Pearson-type tests proposed in the previous sections. In particular, we examine the size behaviour of the P_1, P_2 and P_3 test statistics of overidentifying moment conditions suggested in section 4.3 and assess how they perform comparatively to the J, Wald (W) and distance metric (DM) tests, see sections 2.3.4 and 2.5.6, and also to several bootstrap versions of the first test.

4.5.1 Experimental designs

We follow closely the simulation study realized by Imbens, Spady and Johnson (1998) to compare the finite sample properties of the aforementioned tests, using their first two experimental designs as a basis for our investigation. The first model simulated is a simplified version of an asset-pricing model, characterized by the moment indicators for unit i

$$g\left(X_i, Z_i, \theta\right) = \begin{bmatrix} \exp\left[-0.72 - \theta\left(X_i + Z_i\right) + 3Z_i\right] - 1 \\ Z_i\left\{\exp\left[-0.72 - \theta\left(X_i + Z_i\right) + 3Z_i\right] - 1\right\} \end{bmatrix}, \tag{4.53}$$

where X and Z were generated independently from a $N\left(0, 0.16\right)$ distribution and the true value of θ is 3. The second Monte Carlo experiment is based on the moment vector

$$g\left(Z_i, \theta\right) = \begin{bmatrix} Z_i - \theta \\ Z_i^2 - \theta^2 - 2\theta \end{bmatrix}, \tag{4.54}$$

where Z has a chi-square distribution with one degree of freedom and $\theta_0 = 1$. We considered samples of 100, 200, 500 and 1000 observations, each one being replicated 10000 times.

For the tests requiring evaluation at GEL estimators (W, DM, P_1, P_2 and P_3), we considered both EI and EL estimation. In both cases, consistent estimators for the matrices needed to compute the W and P_3 tests were obtained in three different ways:

- $gel(n)$: uses sample means to estimate consistently V and G, for example:

$$\hat{V}_n = \frac{1}{n}\sum_{i=1}^{n} g\left(y_i, \hat{\theta}\right) g\left(y_i, \hat{\theta}\right)'; \tag{4.55}$$

- $gel(s)$: uses the GEL implied probabilities $\hat{p}_i$, $i = 1, ..., n$, for both V and G, for example:

$$\hat{V}_n = \sum_{i=1}^{n} \hat{p}_i g\left(y_i, \hat{\theta}\right) g\left(y_i, \hat{\theta}\right)'; \tag{4.56}$$

- $gel(r)$: the matrix G is estimated as in $gel\,(s)$ and V_g is estimated robustly as:

$$\hat{V}_n = \sum_{i=1}^{n} \hat{p}_i g\left(y_i, \hat{\theta}\right) g\left(y_i, \hat{\theta}\right)' \left[n \sum_{i=1}^{n} \hat{p}_i^2 g\left(y_i, \hat{\theta}\right) g\left(y_i, \hat{\theta}\right)'\right]^{-1} \cdot$$
$$\cdot \sum_{i=1}^{n} \hat{p}_i g\left(y_i, \hat{\theta}\right) g\left(y_i, \hat{\theta}\right)'. \tag{4.57}$$

The same three procedures were followed to compute the J test but, in addition, we evaluate it also at two-step (J_{2s}), repeatedly-iterated (J_{ri}) and continuous-updating (J_{cu}) GMM estimators, in which cases we only use a consistent estimator for the matrix V based on, naturally, sample means.

In their Monte Carlo simulation study, Imbens, Spady and Johnson (1998) analyzed the finite sample behaviour of the following tests: J_{2s}, J_{ri}, J_{cu}, $J_{ei(s)}$, $W_{ei(s)}$, $W_{ei(r)}$, DM_{ei} and DM_{el}. In this section we replicate their results for the two experimental designs described above and examine whether their conclusions remain valid when other estimators are employed to evaluate the J and Wald tests. In particular, we study the effects of using EL instead of EI estimation [$J_{el(s)}$, $W_{el(s)}$ and $W_{el(r)}$ tests], confirm the conjecture that robust estimation of the matrix V does not work well in the case of the J test [$J_{ei(r)}$ and $J_{el(r)}$ tests], for reasons explained below, and investigate the consequences of using sample means to estimate that same matrix when GEL estimation is utilized [$J_{ei(n)}$, $J_{el(n)}$, $W_{ei(n)}$ and $W_{el(n)}$ tests].

The implementation of the P_3 test requires the previous partition of the sample space into L sets. In order to examine the sensitivity of this test to the number of classes into which the observations are divided, we considered two different values for L: 8 and 16. The definition of each set in each Monte Carlo sample was such that each class contains, approximately, $(100/L)\,\%$ of the observations.

4.5.2 Main results

Tables 4.1 and 4.2, for the asset-pricing model, and 4.3 and 4.4, for the chi-squared moments case, report the estimated size of each test at seven different levels of significance for the asset-pricing models. For each significance level, sample size and model considered, the actual size closest to the nominal size is underlined. For the tests analyzed there, these results conform with those presented by Imbens, Spady and Johnson (1998).[2] As can be immediately seen from tables 4.1 and 4.3, all tests are significantly oversized in almost all cases, particularly for the chi-squared moments model. Clearly, the $W_{ei(r)}$ test registered the best behaviour in most experiments, the only exceptions being the largest nominal sizes, where the J_{cu} test, in the first model, and the $W_{el(r)}$ test, in both models, achieved superior performances. However, even for $n = 1000$ the $W_{ei(r)}$ test is still slightly oversized for most significance levels. The J test evaluated at two-step GMM estimators, the most widely applied test to assess overidentifying moment condition models, has a disastrous behaviour in these experiments, in the asset-pricing model being the worst of all versions of the J test based on sample mean estimators for the matrix V [J_{2s}, J_{ri}, J_{cu}, $J_{ei(n)}$ and $J_{el(n)}$]. The DM tests also produced very modest results, with that based on the EL objective

[2]There is the following correspondence between the notation used here and that utilized by Imbens, Spady and Johnson (1998): $J_{2s} = T^{AM}_{g1}$, $J_{ri} = T^{AM}_{g2}$, $J_{cu} = T^{AM}_{g3}$, $J_{ei(s)} = T^{AM}_{et}$, $W_{ei(s)} = T^{LM}_{et(s)}$, $W_{ei(r)} = T^{LM}_{et(r)}$, $DM_{ei} = T^{CF}_{klic(et)}$ and $DM_{el} = T^{CF}_{lr(el)}$.

function performing substantially better than that using the EI criterion, particularly for the chi-squared moments model and for the smallest nominal sizes.

As noted by Imbens, Spady and Johnson (1998), robust estimation of the matrix V decisively influences the performance of the tests. However, the extraordinary benefits reported by them for the $W_{ei(r)}$ statistic do not extend to all the other tests. They do not extend even to the $W_{el(r)}$ test for the smallest nominal sizes considered. The behaviour of the J test also deteriorates considerably. Although a theoretical analysis of the effects of using the $gel\,(r)$ method is not available, it is evident why the W and the J tests are affected in opposite ways: the matrix V appears in the expression of those tests in an inverse way.

When the J test is evaluated at GEL estimators, it is relatively invariant to the use of EI or EL estimation. In both cases the robust version of the J test is clearly the worst and the actual sizes of the tests based on sample means, $J_{ei(n)}$ and $J_{el(n)}$, are very similar. The only important divergence appears when the $J_{ei(s)}$ and $J_{el(s)}$ tests are considered. In both models the utilization of the GEL implied probabilities in the estimation of the covariance matrix V produced substantially better results in the EL case. For this reason, while the $J_{el(s)}$ statistic is the best of the J tests based on EL estimators, in the EI case the best performances are shared by the $J_{ei(n)}$ (pricing-asset model) and $J_{ei(s)}$ (chi-squared model) tests.

With regard to the Wald statistics, there are significant differences between evaluation at EI or EL estimators. $W_{ei(r)}$ is undoubtedly the best performer, being surpassed by $W_{el(r)}$ only for 10% (sometimes) and 20% (always) levels of significance. For the smallest nominal sizes, the actual sizes of the $W_{el(r)}$ statistic are much higher than those of the $ei\,(r)$ version in both models. The same happens with the $el\,(n)$ version relative to the $ei\,(n)$ one. In contradistinction, EL evaluation leads to better performances when based on the $gel\,(s)$ method. Clearly, as also found for the J test, direct application of the GEL probabilities to estimate the covariance matrix V works much better in the EL case. Therefore, while robust estimation of V is always recommended when calculating Wald tests based on EI estimators, in the EL case the

Table 4.1: Monte Carlo estimated sizes for J, W and DM tests of overidentifying moment conditions: asset-pricing model (10 000 replications)

n	Size	J									W						DM	
		2s(n)	ri(n)	cu(n)	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)	ei	el
100	20.0%	26.7	26.1	24.0	25.5	29.8	29.2	25.9	28.3	28.5	29.9	25.3	26.9	28.1	28.3	24.8	27.2	27.9
	10.0%	17.5	16.7	12.8	15.9	19.6	20.1	16.4	17.6	19.3	19.8	16.1	14.8	18.7	17.6	15.3	16.9	17.0
	5.0%	12.2	11.3	7.2	10.6	13.7	14.6	11.2	11.2	14.0	13.8	11.0	8.3	13.6	11.2	10.4	11.0	11.1
	2.5%	9.5	8.5	4.5	7.8	9.9	11.5	8.3	7.4	10.8	10.3	8.3	4.6	10.4	7.4	7.6	7.7	7.3
	1.0%	6.9	5.9	2.5	5.2	6.9	8.7	5.7	4.3	8.1	7.3	5.8	2.2	7.6	4.3	5.3	5.0	4.1
	0.5%	5.7	4.4	1.6	3.7	5.4	7.2	4.2	2.9	7.0	5.6	4.4	1.3	6.1	2.9	4.0	3.6	2.8
	0.1%	3.9	2.3	0.7	1.7	3.3	4.7	2.1	1.1	4.6	3.5	2.4	0.4	4.0	1.1	2.7	1.8	1.2
200	20.0%	25.3	25.1	24.2	24.8	28.0	27.7	25.0	26.8	27.2	28.1	24.7	25.7	26.0	26.8	23.6	26.0	26.8
	10.0%	15.0	14.7	13.3	14.4	17.3	18.0	14.7	15.7	17.3	17.4	14.5	13.6	16.5	15.7	13.9	15.2	15.5
	5.0%	9.9	9.5	7.6	9.2	11.2	12.5	9.5	9.4	11.9	11.3	9.5	7.1	11.2	9.4	8.9	9.3	9.1
	2.5%	6.8	6.6	4.5	6.2	7.8	9.1	6.5	5.8	8.5	7.8	6.5	3.8	8.1	5.8	6.6	5.8	5.6
	1.0%	4.6	4.3	2.4	4.0	5.0	6.2	4.3	2.9	6.0	5.0	4.3	1.5	5.9	2.9	4.6	3.6	3.0
	0.5%	3.5	3.1	1.4	2.9	3.5	4.9	3.1	1.9	4.8	3.6	3.1	0.9	4.8	1.9	3.6	2.5	1.9
	0.1%	1.9	1.5	0.4	1.3	2.0	3.0	1.5	0.7	3.0	2.1	1.6	0.2	3.2	0.7	2.1	1.1	0.7
500	20.0%	23.1	23.0	22.7	22.8	25.6	25.4	22.9	24.7	25.1	25.7	22.9	23.7	24.2	24.7	22.1	24.0	24.4
	10.0%	13.1	13.0	12.4	12.8	15.0	15.6	13.0	13.5	15.0	15.0	13.0	12.0	14.4	13.5	12.4	13.4	13.6
	5.0%	8.0	7.9	7.3	7.6	9.3	10.1	7.8	7.9	9.7	9.4	7.8	6.3	9.3	7.9	7.9	7.8	7.7
	2.5%	5.0	4.9	4.1	4.7	6.2	6.9	4.9	4.6	6.6	6.2	4.9	3.4	6.6	4.6	5.4	4.7	4.6
	1.0%	3.0	3.0	2.3	2.9	3.7	4.4	3.0	2.4	4.2	3.7	3.0	1.3	4.5	2.4	3.7	2.5	2.2
	0.5%	2.1	2.1	1.5	1.9	2.4	3.3	2.1	1.4	3.2	2.5	2.0	0.6	3.7	1.4	2.8	1.6	1.3
	0.1%	0.8	0.9	0.4	0.7	1.1	1.7	0.8	0.5	1.8	1.2	0.8	0.1	2.4	0.5	1.8	0.5	0.4
1000	20.0%	21.8	21.8	21.6	21.7	23.5	23.5	21.9	22.6	23.2	23.5	21.8	22.0	22.3	22.6	20.8	22.5	22.7
	10.0%	11.9	11.8	11.6	11.8	13.2	13.9	11.9	12.1	13.6	13.3	11.9	11.2	12.5	12.1	11.2	12.3	12.3
	5.0%	6.7	6.8	6.5	6.7	8.0	8.5	6.8	7.1	8.3	8.1	6.8	5.9	8.0	7.1	6.6	6.9	6.8
	2.5%	4.4	4.3	4.1	4.3	4.8	5.5	4.4	3.9	5.4	5.0	4.4	3.0	5.2	3.9	4.3	4.1	3.9
	1.0%	2.4	2.4	2.3	2.4	2.5	3.5	2.4	1.8	3.4	2.6	2.4	1.2	3.4	1.8	2.5	2.2	1.9
	0.5%	1.7	1.7	1.6	1.7	1.7	2.4	1.7	1.0	2.5	1.7	1.7	0.6	2.4	1.0	1.9	1.3	1.0
	0.1%	0.7	0.7	0.6	0.6	0.7	1.2	0.6	0.3	1.3	0.7	0.7	0.1	1.5	0.3	1.0	0.5	0.3

Note: The actual size closest to the nominal size of all tests contained in Tables 4.1 and 4.2 is underlined.

Table 4.2: Monte Carlo estimated sizes for Pearson-type tests of overidentifying moment conditions: asset-pricing model (10 000 replications)

n	Size	P1		P2		P3 (L=8)						P3 (L=16)					
		ei	el	ei	el	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)
100	20.0%	26.7	28.6	30.4	28.3	25.2	22.4	20.2	25.8	22.9	21.8	25.1	22.5	20.2	25.4	22.9	21.2
	10.0%	17.0	19.3	20.4	17.6	14.3	15.4	9.3	14.0	14.7	8.3	14.2	15.6	9.7	13.7	15.0	8.1
	5.0%	11.8	14.0	14.6	11.2	8.4	10.9	3.8	7.3	10.2	2.1	8.6	11.1	4.6	7.3	10.4	2.3
	2.5%	8.9	10.8	10.9	7.4	5.0	7.9	1.1	3.8	6.9	0.3	5.3	8.2	1.7	3.7	7.5	0.4
	1.0%	6.4	8.1	7.9	4.3	2.4	4.7	0.1	1.4	4.0	0.0	2.7	5.2	0.3	1.6	4.5	0.0
	0.5%	5.0	7.0	6.2	2.9	1.4	2.7	0.0	0.6	2.3	0.0	1.6	3.5	0.1	0.8	2.8	0.0
	0.1%	2.8	4.6	4.3	1.1	0.3	0.6	0.0	0.1	0.5	0.0	0.6	1.1	0.0	0.1	0.9	0.0
200	20.0%	25.5	27.2	28.5	26.8	25.4	21.6	21.5	25.5	23.3	22.6	25.0	21.5	21.2	25.3	22.8	22.3
	10.0%	15.4	17.3	18.1	15.7	14.3	14.2	10.2	14.1	13.3	10.5	14.1	14.3	10.1	13.7	13.3	9.9
	5.0%	10.0	11.9	12.0	9.4	8.1	10.0	4.8	7.6	8.8	4.3	8.1	10.1	4.8	7.3	8.8	3.9
	2.5%	7.0	8.5	8.7	5.8	4.7	7.1	2.4	4.3	6.0	1.3	4.7	7.2	2.5	4.0	5.9	1.1
	1.0%	4.7	6.0	5.7	2.9	2.5	4.7	0.8	1.9	3.6	0.3	2.5	4.8	0.9	1.6	3.7	0.3
	0.5%	3.5	4.8	4.3	1.9	1.5	3.3	0.3	1.0	2.6	0.1	1.5	3.4	0.4	0.8	2.4	0.1
	0.1%	1.9	3.0	2.7	0.7	0.5	1.3	0.0	0.3	1.0	0.0	0.5	1.5	0.0	0.3	1.0	0.0
500	20.0%	23.6	25.1	26.0	24.7	24.2	21.0	21.6	24.3	22.8	22.2	24.0	21.0	21.3	24.0	22.5	22.0
	10.0%	13.5	15.0	15.6	13.5	13.5	12.4	10.4	13.3	12.6	11.0	13.4	12.3	10.4	12.8	12.3	10.6
	5.0%	8.3	9.7	9.9	7.9	7.7	8.1	5.1	7.7	7.2	5.6	7.5	8.1	4.9	7.4	7.0	5.2
	2.5%	5.3	6.6	6.8	4.6	4.5	5.6	2.5	4.6	4.5	2.6	4.4	5.5	2.5	4.1	4.4	2.4
	1.0%	3.3	4.2	4.2	2.4	2.4	3.6	1.0	2.3	2.7	1.0	2.3	3.6	1.0	2.0	2.6	0.7
	0.5%	2.3	3.2	3.1	1.4	1.4	2.6	0.4	1.4	1.8	0.4	1.3	2.6	0.5	1.1	1.7	0.3
	0.1%	1.0	1.8	1.7	0.5	0.4	1.2	0.1	0.5	0.7	0.1	0.4	1.1	0.1	0.4	0.6	0.0
1000	20.0%	22.1	23.2	23.8	22.6	22.9	20.8	21.1	22.6	22.0	21.2	22.9	20.7	20.9	22.5	21.7	21.1
	10.0%	12.3	13.6	13.6	12.1	12.6	11.5	10.2	12.3	11.9	10.5	12.4	11.5	10.1	12.1	11.7	10.2
	5.0%	7.2	8.3	8.4	7.1	7.1	7.1	5.2	7.0	6.5	5.6	7.0	7.0	5.2	6.9	6.3	5.3
	2.5%	4.6	5.4	5.4	3.9	4.2	4.8	2.7	4.0	4.0	2.6	4.0	4.7	2.7	3.8	4.0	2.5
	1.0%	2.6	3.4	3.0	1.8	2.2	3.0	1.1	2.0	2.3	1.0	2.1	3.0	1.0	1.8	2.2	0.9
	0.5%	1.8	2.5	2.1	1.0	1.2	2.1	0.6	1.2	1.5	0.5	1.2	2.0	0.5	1.0	1.4	0.4
	0.1%	0.7	1.3	1.0	0.3	0.4	1.0	0.1	0.4	0.5	0.1	0.4	1.0	0.1	0.3	0.5	0.1

Note: The actual size closest to the nominal size of all tests contained in Tables 4.1 and 4.2 is underlined.

Table 4.3: Monte Carlo estimated sizes for J, W and DM tests of overidentifying moment conditions: chi-squared moments model (10 000 replications)

n	Size	J									W						DM	
		2s(n)	ri(n)	cu(n)	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)	ei	el
100	20.0%	34.6	34.6	34.6	34.7	38.0	38.0	35.0	36.5	37.2	37.7	34.1	35.1	33.8	36.5	31.2	35.7	36.3
	10.0%	27.0	26.9	26.9	27.2	28.0	30.1	27.5	25.9	29.5	27.6	27.0	23.4	25.4	25.9	22.8	26.5	26.0
	5.0%	22.3	22.3	22.3	22.6	21.5	25.2	23.1	19.3	24.9	21.3	22.3	16.9	20.3	19.3	17.8	20.7	19.3
	2.5%	18.8	18.8	18.8	19.1	17.6	21.7	19.8	14.8	21.8	17.6	19.3	12.7	16.7	14.8	14.4	17.1	15.5
	1.0%	15.5	15.5	15.5	15.8	13.4	18.2	16.7	11.1	18.9	13.5	16.2	9.6	13.4	11.1	11.3	13.9	11.5
	0.5%	13.4	13.4	13.4	13.8	11.5	16.4	14.7	9.0	17.3	11.7	14.2	8.0	11.8	9.0	9.5	12.0	9.9
	0.1%	9.8	9.8	9.8	10.2	8.1	13.0	11.3	5.9	14.3	8.6	10.8	5.4	8.9	5.9	7.3	8.9	6.8
200	20.0%	29.0	29.0	29.0	29.0	30.9	31.9	29.1	29.6	31.2	30.5	28.5	28.2	26.0	29.6	23.1	29.6	30.0
	10.0%	20.9	20.9	20.9	21.0	21.1	23.8	21.1	19.5	22.9	20.8	20.8	16.8	17.4	19.5	14.7	20.2	19.8
	5.0%	16.4	16.4	16.4	16.5	14.6	19.0	16.8	12.9	18.4	14.4	16.5	10.5	12.8	12.9	10.8	15.1	13.6
	2.5%	13.8	13.8	13.8	13.8	11.0	15.8	14.1	8.9	15.7	10.9	13.9	6.9	9.8	8.9	8.5	11.7	9.7
	1.0%	10.4	10.4	10.4	10.6	7.6	12.7	11.0	5.8	13.1	7.6	10.6	4.4	7.8	5.8	6.3	8.7	6.5
	0.5%	9.1	9.0	9.0	9.2	6.1	10.8	9.5	4.1	11.3	6.1	9.2	3.1	6.5	4.1	5.3	7.2	5.1
	0.1%	6.3	6.3	6.3	6.5	3.5	8.4	6.9	2.1	9.0	3.5	6.6	1.7	4.7	2.1	3.8	4.7	2.9
500	20.0%	25.4	25.4	25.4	25.4	26.3	27.9	25.5	25.3	27.3	26.1	25.2	24.0	21.7	25.3	19.1	26.4	26.1
	10.0%	16.4	16.4	16.4	16.5	15.6	18.8	16.5	14.4	18.3	15.4	16.3	12.5	13.1	14.4	11.3	15.7	14.8
	5.0%	11.5	11.5	11.5	11.5	9.8	13.3	11.6	8.7	13.0	9.7	11.4	6.9	9.0	8.7	7.8	10.2	9.1
	2.5%	8.6	8.6	8.6	8.6	6.6	10.3	8.6	5.3	10.2	6.5	8.5	3.7	6.7	5.3	5.8	7.1	5.9
	1.0%	6.3	6.3	6.3	6.4	3.7	7.5	6.4	2.7	7.7	3.6	6.3	1.7	4.9	2.7	4.0	5.0	3.3
	0.5%	5.2	5.2	5.2	5.2	2.6	6.2	5.3	1.8	6.5	2.6	5.2	1.0	4.0	1.8	3.2	3.6	2.3
	0.1%	3.0	3.0	3.0	3.1	1.2	4.2	3.2	0.7	4.7	1.2	3.1	0.3	2.7	0.7	2.0	2.0	1.0
1000	20.0%	23.2	23.2	23.2	23.2	24.4	25.1	23.3	23.6	24.6	24.2	23.1	22.5	20.8	23.6	18.8	23.8	23.5
	10.0%	14.0	14.0	14.0	14.1	13.9	16.0	14.1	12.9	15.7	13.7	14.0	11.6	12.0	12.9	10.7	13.9	13.2
	5.0%	9.1	9.1	9.1	9.1	8.4	10.8	9.1	7.4	10.3	8.3	9.1	6.1	8.0	7.4	7.0	8.5	7.9
	2.5%	6.5	6.5	6.5	6.5	5.0	7.9	6.5	4.1	7.7	4.8	6.4	2.9	5.8	4.1	5.2	5.6	4.5
	1.0%	4.3	4.3	4.3	4.3	2.7	5.3	4.3	2.0	5.5	2.7	4.2	1.1	4.2	2.0	3.6	3.2	2.2
	0.5%	3.2	3.2	3.2	3.2	1.7	4.1	3.2	1.2	4.3	1.7	3.2	0.6	3.4	1.2	2.8	2.1	1.3
	0.1%	1.7	1.7	1.7	1.7	0.7	2.2	1.7	0.4	2.5	0.7	1.7	0.1	2.1	0.4	1.6	0.9	0.4

Note: The actual size closest to the nominal size of all tests contained in Tables 4.3 and 4.4 is underlined.

Table 4.4: Monte Carlo estimated sizes for Pearson-type tests of overidentifying moment conditions: chi-squared moments model (10 000 replications)

n	Size	P1		P2		P3 (L=8)						P3 (L=16)					
		ei	el	ei	el	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)	ei(n)	ei(s)	ei(r)	el(n)	el(s)	el(r)
100	20.0%	35.6	37.2	38.4	36.5	32.8	33.7	27.9	30.3	32.7	22.1	34.3	34.9	29.8	34.3	33.7	29.9
	10.0%	27.9	29.5	28.5	25.9	21.6	28.0	17.1	16.7	25.6	9.2	24.4	29.7	20.6	21.8	28.4	11.5
	5.0%	23.5	24.9	22.1	19.3	15.3	23.4	10.5	9.8	20.6	4.3	18.0	25.6	14.6	13.5	24.1	4.6
	2.5%	20.0	21.8	18.2	14.8	11.0	19.6	6.3	6.1	16.6	2.1	14.3	22.2	9.9	8.6	20.7	2.3
	1.0%	17.1	18.9	14.2	11.1	7.6	15.3	3.1	3.7	12.4	0.9	10.4	18.6	5.3	5.1	17.3	1.1
	0.5%	15.2	17.3	12.2	9.0	5.7	12.6	1.8	2.7	9.6	0.5	8.4	16.3	3.0	3.6	15.0	0.6
	0.1%	12.0	14.3	9.0	5.9	3.1	6.1	0.1	1.4	2.1	0.0	4.8	12.1	0.5	2.0	10.9	0.0
200	20.0%	29.8	31.3	31.3	29.6	28.0	27.7	23.5	26.1	27.6	20.4	29.0	28.6	24.9	28.7	27.8	25.9
	10.0%	21.5	22.9	21.8	19.5	17.6	21.8	13.2	14.1	19.1	8.2	19.2	23.2	15.8	17.5	21.1	11.4
	5.0%	17.2	18.4	15.5	12.9	11.1	17.8	7.6	7.6	14.8	3.5	13.4	19.4	10.4	10.9	17.4	4.6
	2.5%	14.4	15.7	11.7	8.9	7.4	15.0	4.4	4.5	11.5	1.6	9.6	16.6	7.1	6.7	14.8	1.5
	1.0%	11.5	13.1	8.4	5.8	4.6	11.7	2.1	2.4	8.4	0.6	6.6	13.9	4.1	3.3	11.8	0.4
	0.5%	9.8	11.3	6.8	4.1	3.1	9.7	1.3	1.5	6.7	0.3	5.2	11.8	2.6	2.1	9.9	0.1
	0.1%	7.4	9.0	4.1	2.1	1.5	6.7	0.3	0.5	3.9	0.1	2.8	8.9	1.0	0.8	7.2	0.0
500	20.0%	26.0	27.3	26.8	25.3	25.4	23.8	22.6	23.4	25.0	20.5	26.0	24.0	23.2	24.8	24.9	22.2
	10.0%	17.2	18.3	16.1	14.4	14.1	17.0	11.1	12.6	14.7	9.2	15.2	17.7	11.8	13.8	15.8	10.6
	5.0%	12.0	13.0	10.5	8.7	8.5	12.7	5.8	7.0	9.4	4.2	9.5	13.8	6.9	8.1	11.0	5.0
	2.5%	9.1	10.2	7.2	5.3	5.3	9.9	3.2	4.0	6.8	2.1	6.2	11.0	4.4	4.8	8.4	2.0
	1.0%	6.6	7.7	4.5	2.7	2.8	7.3	1.3	1.8	4.7	0.9	3.8	8.4	2.3	2.2	6.1	0.7
	0.5%	5.6	6.5	3.1	1.8	1.8	6.0	0.7	1.3	3.5	0.5	2.5	6.9	1.4	1.3	5.1	0.5
	0.1%	3.6	4.7	1.6	0.7	0.6	3.9	0.2	0.5	1.8	0.1	1.2	4.9	0.4	0.5	3.1	0.1
1000	20.0%	23.6	24.6	24.7	23.6	23.5	21.8	21.6	22.3	23.1	20.3	23.7	22.0	21.9	23.1	23.1	21.2
	10.0%	14.6	15.7	14.4	12.9	13.1	13.9	10.7	12.1	13.0	9.8	13.6	14.4	11.2	12.6	13.5	10.3
	5.0%	9.5	10.3	9.0	7.4	7.7	9.8	5.4	6.7	7.7	4.9	8.2	10.5	6.0	7.3	8.5	5.2
	2.5%	6.8	7.7	5.5	4.1	4.3	7.5	2.4	3.9	5.0	2.7	4.9	8.1	3.2	4.1	6.0	2.5
	1.0%	4.6	5.5	3.1	2.0	2.1	5.4	1.0	2.1	2.7	1.3	2.5	6.0	1.5	2.0	3.8	1.0
	0.5%	3.5	4.3	2.1	1.2	1.2	4.1	0.6	1.4	1.9	0.7	1.5	4.7	0.9	1.3	2.8	0.5
	0.1%	1.9	2.5	1.0	0.4	0.4	2.2	0.1	0.5	0.8	0.2	0.5	2.7	0.4	0.4	1.3	0.1

Note: The actual size closest to the nominal size of all tests contained in Tables 4.3 and 4.4 is underlined.

choice of an estimator for that matrix must depend on the level of significance that the practitioner chooses to use: in both models, the utilization of the $el\,(r)$ method is preferable for the largest nominal sizes, whereas the $el\,(s)$ method performs better for smaller significance levels.

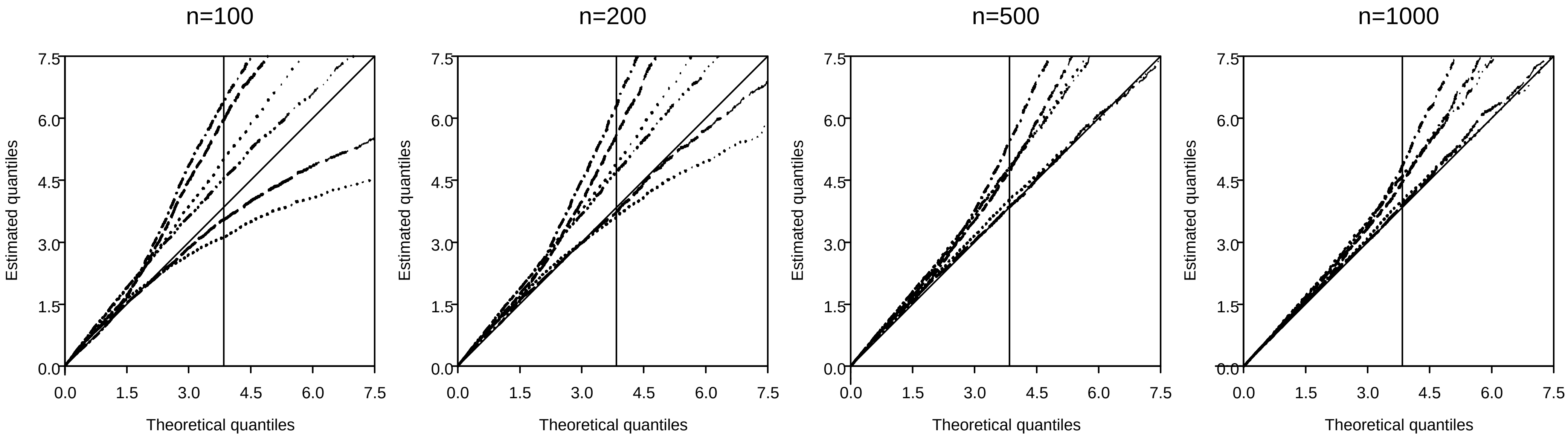
The estimated sizes for the Pearson-type statistics are reported in tables 4.2 and 4.4. The P_1 and P_2 tests perform very modestly, being substantially oversized in all cases. Their size behaviour does not differ much from that described for the other tests.[3] However, the P_3 statistic shows a very promising performance. Whichever the number of classes considered, the general effects of evaluating this test at different estimators are similar in all cases. Analogously to the W test, the least number of rejections of the null hypothesis occurs when robust estimation of V is employed. This is not surprising since the matrix V appears in their expressions in a similar form. However, while this was always beneficial for the W test, the P_3 test becomes sometimes quite undersized, particularly for the smallest nominal sizes and sample sizes considered.

Figure 4.1 displays QQ-plots comparing the six versions of the P_3 test for the $L = 8$ case. Vertical coordinates are Monte Carlo estimates of quantiles of the finite sample distribution of those statistics and horizontal coordinates are quantiles of a chi-square variable with one degree of freedom. The vertical solid line marks the asymptotic critical value for a nominal size of 0.05. Clearly, the best performances are obtained by $P_3^{ei(r)}$ and $P_3^{el(r)}$. Note how for $n \geq 500$ (first model) or $n = 1000$ (second model) the estimated quantiles of these tests are very close to the asymptotic ones while the other versions of P_3 are still significantly oversized. Notice also how, for small sample sizes, all three EL versions of the P_3 test tend to reject the null hypothesis significantly less than the corresponding EI variants.

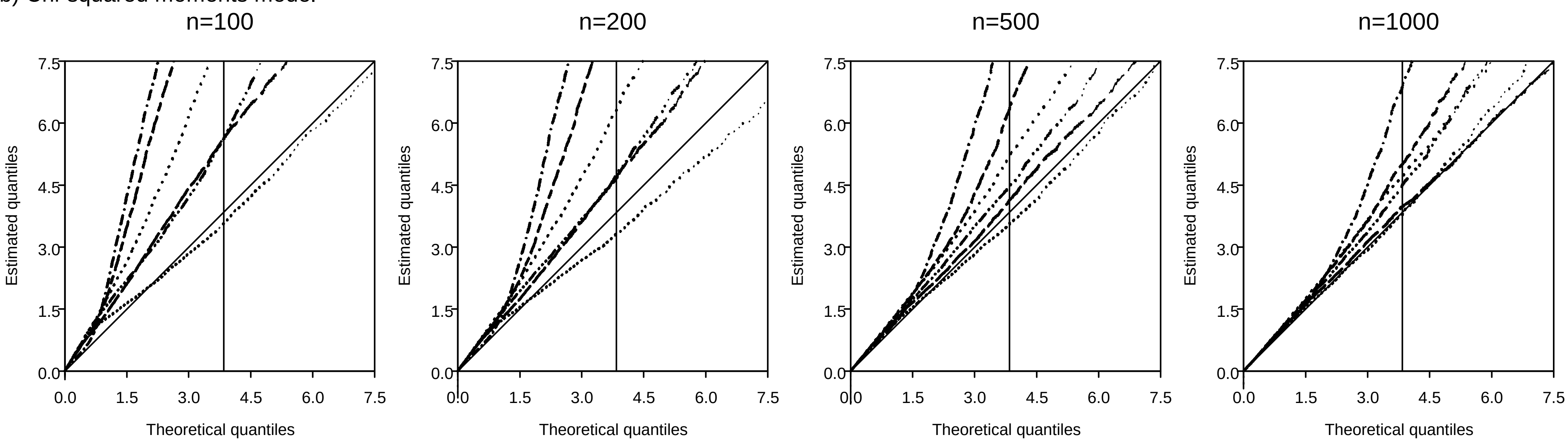
[3]The sizes estimated for the EL version of the P_2 test are numerically equal to those calculated for the $J_{el(s)}$ and $W_{el(s)}$ statistics which is due to the particular form assumed by the EL probabilities: $\hat{p}_i = n^{-1}\left[1 + \hat{\phi}' g\left(y_i, \hat{\theta}\right)\right]^{-1}$, $i = 1, ..., n$, see (2.54). For example, as $\hat{\phi}' g\left(y_i, \hat{\theta}\right) = n\hat{p}_i - 1$ and V is estimated by $\sum_{i=1}^{n} \hat{p}_i g\left(y_i, \hat{\theta}\right) g\left(y_i, \hat{\theta}\right)'$, we have $W_n = n\hat{\phi}' \hat{V}_n \hat{\phi} = \sum_{i=1}^{n} \frac{\hat{\phi}' g\left(y_i, \hat{\theta}\right) g\left(y_i, \hat{\theta}\right)' \hat{\phi}}{1 + \hat{\phi}' g\left(y_i, \hat{\theta}\right)} = P_2$.

Figure 4.1: QQ-plots of P3 tests of overidentifying moment conditions (L=8; 10 000 replications)

a) Asset-pricing model



b) Chi-squared moments model



Notes: P3ei(n) (dotted line), P3ei(s) (dot-dashed line), P3ei(r) (dashed line), P3el(n) (three-dot-dashed line), P3el(s) (two-dashed line), P3el(r) (frequent-dotted line).

The performance of the P_3 test does not seem to depend significantly on L on small samples. This is particularly evident for the asset-pricing model case. For the chi-squared moments model the differences between the $L = 8$ and $L = 16$ cases are more important being, however, attenuated as the sample size increases. Figure 4.2 illustrates this situation, displaying QQ-plots for the $P_3^{ei(r)}$ test for the two distinct values of L simulated.

Figure 4.3 compares the robust forms of the W and P_3 tests (for $L = 8$), both evaluated at EI and EL estimators. Recall that the $W_{ei(r)}$ statistic registered the best behaviour of all tests analyzed in the previous sub-section. From Figure 4.3 we see that the P_3 test clearly performs better for both models, its actual quantiles being in most cases closer to the asymptotic ones. Furthermore, while the P_3 test is relatively indifferent to the use of EI or EL estimation, at least for larger sample sizes, in the case of the Wald test EL estimation does not work well, even for $n = 1000$.

4.5.3 Alternative bootstrap J_{2s} tests

The Monte Carlo experiments in the previous sub-section confirmed that, at least for these sorts of models, first-order asymptotic theory does not provide a good approximation to the distribution of the J_{2s} statistic, the most applied test of overidentifying moment conditions. The alternative tests considered performed substantially better, especially the P_3 test proposed in this chapter. In this section we investigate other alternatives, namely the ability of bootstrap techniques to improve the size properties of the J_{2s} test. Due to the computation time involved, only the 10%, 5% and 1% bootstrap critical values were calculated and the $n = 100$ case analyzed.

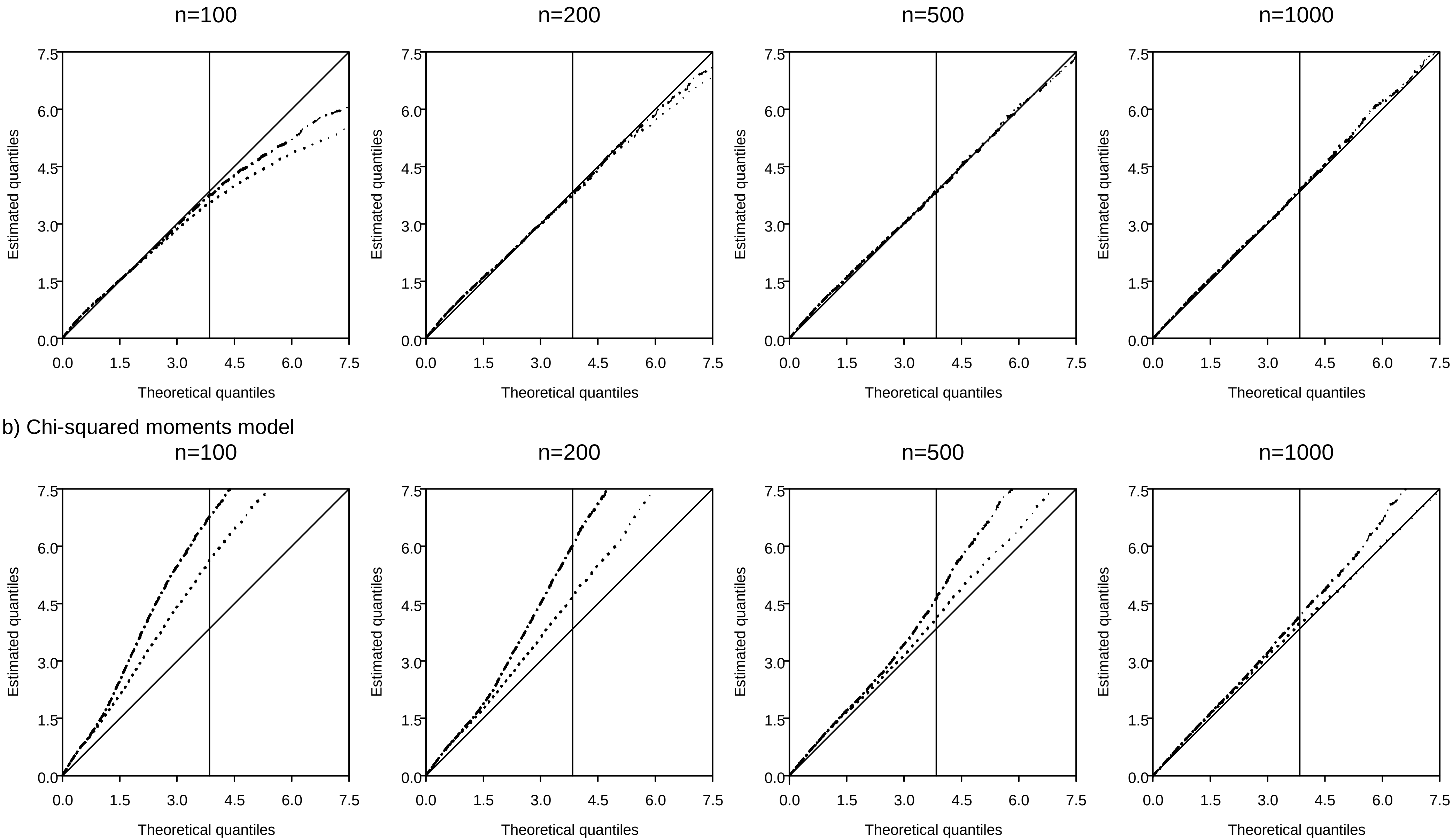
The general procedures implemented to obtain the bootstrap critical values can be summarized as follows:

1. Calculate the two-step GMM estimator using the original data and utilize it to evaluate the J test;

2. Generate $B = 100$ bootstrap samples of size $n = 100$ by sampling the original

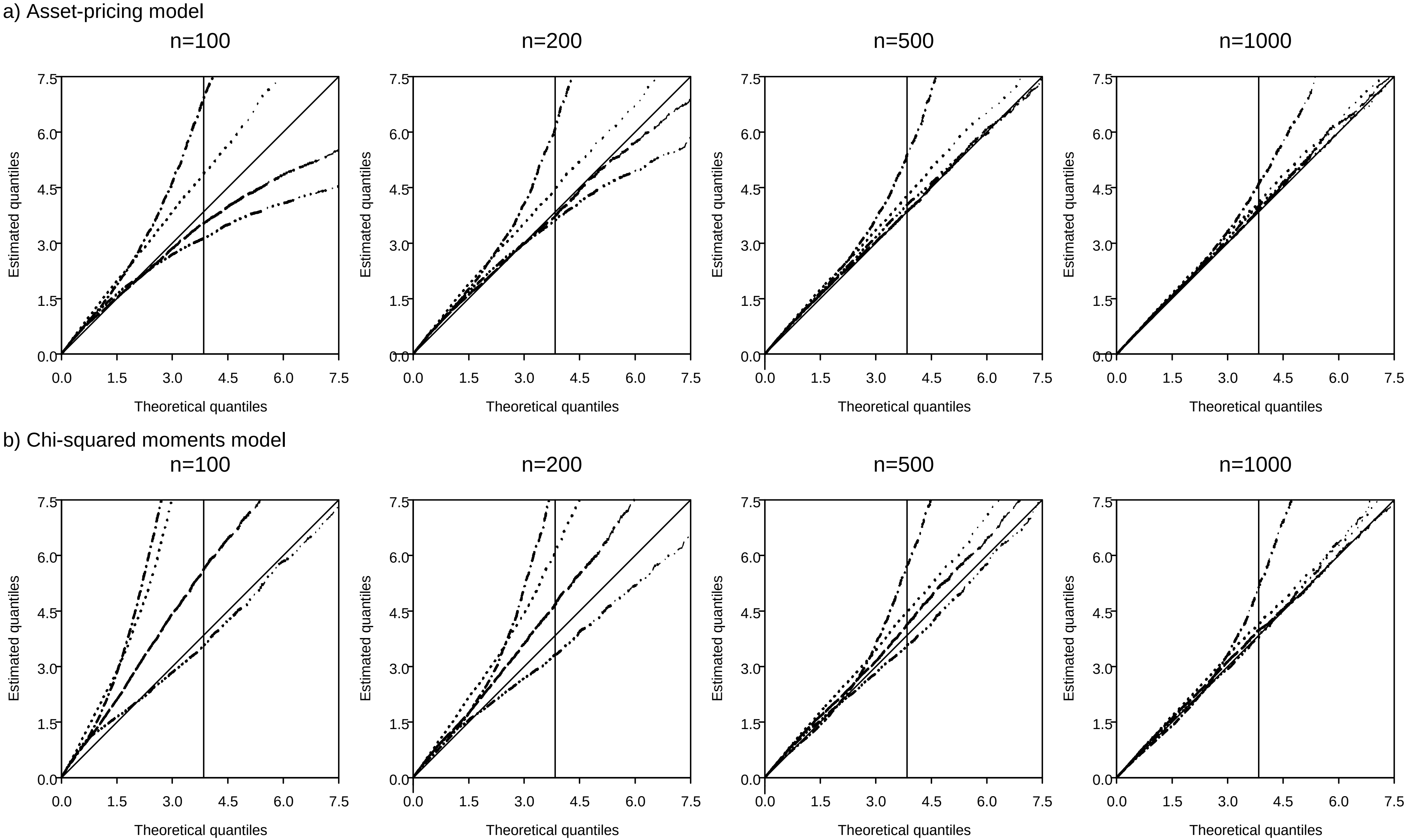
Figure 4.2: QQ-plots of ei(r) P3 tests of overidentifying moment conditions (10 000 replications)

a) Asset-pricing model



Notes: L=8 (dotted line), L=16 (dot-dashed line).

Figure 4.3: QQ-plots of robust forms of Wald and P3 (L=8) tests of overidentifying moment conditions (10 000 replications)



Notes: Wei(r) (dotted line), Wel(r) (dot-dashed line), P3ei(r) (dashed line), P3el(r) (three-dot-dashed line).

data randomly with replacement according to the chosen resampling distribution function F^*;

3. For each bootstrap sample compute the corresponding two-step GMM estimator and use it to obtain bootstrap versions of the J test;

4. Calculate the $1 - \alpha$ quantiles of the empirical distribution of the bootstrap versions of the tests in order to obtain bootstrap critical values;

5. Determine whether the null hypotheses of the tests are rejected using the bootstrap critical values.

By repeating these steps 10000 times we estimated the levels of the tests using bootstrap critical values. We analyzed five different techniques to obtain the bootstrap critical values, namely the NP, RNP, FSEL, EL and REL bootstrap methods, discussed in chapter 3.

The results of the experiments are shown in Table 4.5. Whichever bootstrap method is used, the empirical levels of this test are less than when asymptotic critical values were used. However, their behaviour is still not satisfactory. On the one hand, the J_{2s} test never rejects the null hypotheses when it is based on NP bootstrap critical values. As discussed in chapter 3, this failure results from the fact that, instead of imposing the null hypotheses (4.1), the empirical distribution corresponds to an alternative hypothesis where the moment conditions (4.1) do not hold. As Brown, Newey and May (1997, p.8) pointed out, "the overidentification test should tend to reject when the estimated sample moment conditions are far from zero, but it is exactly those cases where bootstrapping from the empirical distribution should yield large critical values, because they correspond to cases where the moments are far from their null hypothesis value of zero". Ziliak (1997) presents further evidence of the severe size distortions that result from the application of the NP bootstrap to obtain critical values for the J_{2s} test.

On the other hand, the other four bootstrap methods provide a nice improvement over the size behaviour of the J_{2s} test, especially for the asset-pricing model.

Table 4.5: Monte Carlo estimated sizes for the J_{2s} test of overidentifying moment conditions using bootstrap critical values (10 000 replications; 100 observations; 100 bootstrap samples)

Size	Asymp.	Bootstrap method				
		NP	RNP	FSEL	EL	REL
Asset-pricing model						
10%	17.5	0.0	13.5	14.5	14.8	11.3
5%	12.2	0.0	8.4	10.5	10.5	6.6
1%	6.9	0.0	4.2	6.2	5.7	2.7
Chi-squared moments model						
10%	27.0	0.0	23.2	24.3	24.6	24.5
5%	22.3	0.0	19.0	19.4	19.9	19.5
1%	15.5	0.0	13.6	12.4	13.8	12.8

Indeed, in this case, the utilization of bootstrap critical values allows the differences between the empirical and the nominal levels of this test to be reduced by 46-54% (RNP bootstrap), 12-40% (FSEL bootstrap), 20-36% (EL bootstrap) and 71-83% (REL bootstrap). For the chi-squared moments model the reductions are much more modest: 13-22% (RNP bootstrap), 16-21% (FSEL bootstrap), 12-14% (EL bootstrap) and 16-19% (REL bootstrap). However, even for the first model, the size distortions of the J_{2s} test are not completely eliminated. Moreover, as we saw in Tables 4.1-4.4, there are some tests (namely some versions of the P_3 tests, in both cases, and the robust forms of the Wald test, for the second model) with better size properties, particularly for the 1% level. As these tests are based on asymptotic critical values, being, therefore, less time consuming, there seems to be little point in using the bootstrap J_{2s} test simulated in this sub-section.

4.6 Summary

In this chapter we developed new Pearson-type statistics suitable for testing overidentifying moment conditions and parametric restrictions. One of those statistics, the P_3 test, performed very well in two Monte Carlo simulation studies concerning tests of overidentifying moment conditions. Its size behaviour, when based on robust estimation of the matrix V, seems to be superior to that of both alternative tests based on asymptotic critical values and the J_{2s} test based on bootstrap critical values. Moreover, the P_3 statistic does not seem to be sensitive to the number of classes into which the sample space is divided.

Chapter 5

Non-nested hypothesis tests

5.1 Introduction

This chapter is concerned with tests for non-nested hypothesis of models which are specified solely in terms of moment conditions. To the best of our knowledge, there are relatively few papers which address testing non-nested hypothesis in a moment condition framework. This issue has been investigated by Singleton (1985), Ghysels and Hall (1990b) and Smith (1992), who detail various tests based on efficient two-step GMM estimation. Although all of these tests may also be evaluated at GEL estimators with no alteration to their first order asymptotic properties, they do not utilize all the information provided by the GEL method. Cox-type non-nested tests [Cox (1961, 1962)] requiring evaluation at GEL estimators were, therefore, suggested in Smith (1997).

In this chapter we propose a number of new tests that integrate and complement the works of those authors. On the one hand, we derive generalized statistics that include most of the existing tests as particular cases. On the other hand, most of the tests that we suggest require evaluation at GEL estimators and are based on the encompassing principle of Mizon and Richard (1986). Thus, they will constitute an important alternative method for the assessment of moment condition models against specific non-nested alternatives.

According to the encompassing principle, the validity of a given model against a rival formulation may be tested by examining whether or not the former model can predict the relevant behaviour of the latter. We consider two different approaches to this question. First, we derive parametric encompassing tests, which are based on the usual approach of contrasting two consistent estimators, under the null hypothesis, of the pseudo true values of the parameters of the alternative model. Our second approach permits the construction of simpler tests, requiring only a single estimation of the alternative model. These tests, which we term generalized moment encompassing tests, involve the comparison of two consistent estimators, under the null hypothesis, of a statistic which may represent a particular feature of interest of the competing alternative model. Accordingly, a wide class of encompassing tests is defined. A particular variant in this class, based on the contrast between two consistent estimators of the moment indicators of the alternative model, provides a simple method of implementing Ghysels and Hall's (1990b) idea for constructing a moment-based test in the GMM framework. Moreover, unlike Ghysels and Hall's (1990b) test, ours does not require the introduction of auxiliary (and, possibly, erroneous) assumptions in addition to those given by the moment conditions. Some of Smith's (1997) Cox-type tests may also be viewed as members of this class.

This chapter is organized as follows. Section 5.2 introduces some notation, briefly outlines the competing hypothesis and gives general forms for Cox-type and encompassing-type non-nested test statistics. The Cox-type tests of Singleton (1985) and Smith (1992, 1997) are reviewed in section 5.3. Section 5.4 discusses Smith's (1992) GMM parametric encompassing statistic together with a new statistic based on GEL estimators. The new class of moment encompassing statistics is presented in section 5.5. The finite sample properties of some of these tests are investigated in a Monte Carlo simulation study in section 5.6. Section 5.7 concludes.

5.2 Non-nested hypothesis and tests

This and the following sections of this chapter are concerned with non-nested tests for the comparison of rival models based on differing moment conditions. The moment indicators associated with the competing moment condition models may be different in functional form and according to included conditioning variables. The main emphasis in these sections is non-nested tests based on the GEL approach described in the preceding chapters; however, tests based on GMM are also considered.

5.2.1 Non-nested hypothesis

A notation similar to that utilized in the previous chapters is employed to characterize the alternative models, an additional subscript (g or q) being used in some cases to distinguish between them. Thus, denote the model embodied in the moment conditions

$$E_g\left[g\left(y, \theta_0\right)\right] = 0 \tag{5.1}$$

by H_g, where $g\left(\cdot\right)$ is an s_g-vector of moment indicators known up to the k_g-element parameter vector θ, $s_g \geq k_g$, and $E_g\left[\cdot\right]$ denotes expectation taken with respect to the unknown distribution of y under H_g. Consider a rival model H_q based on the s_q-vector of moment indicators $q\left(y, \beta\right)$, where β is a k_q-vector of unknown parameters and $s_q \geq k_q$. The corresponding moment conditions defining H_q are

$$E_q\left[q\left(y_i, \beta_0\right)\right] = 0, \tag{5.2}$$

where $\beta_0 \in B$, with the parameter space B compact, and $E_q\left[\cdot\right]$ denotes expectation taken with respect to the unknown distribution of y under H_q. Throughout this chapter H_g is always considered as the null hypothesis that we aim to test against H_q. All tests discussed below are based on this assumption but, as usual with this kind of tests, just interchanging the roles of the hypothesis, we can find appropriate tests for assessing H_q against H_g.

Let $V_{gn}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} g(y_i,\theta)\, g(y_i,\theta)'$. Following section 2.3.3, the two-step H_g-efficient GMM estimator $\hat{\theta}$ for θ_0 minimizes $g_n(\theta)' \left[V_{gn}\left(\hat{\theta}^1\right)\right]^{-1} g_n(\theta)$, where $V_{gn}\left(\hat{\theta}^1\right)$ is a positive semi-definite H_g-consistent estimator for $V_g \equiv E_g\left[g(y,\theta_0)\, g(y,\theta_0)'\right]$ and $\hat{\theta}^1$ denotes a preliminary H_g-consistent estimator for θ_0. Similarly, define $q_n(\beta) \equiv \frac{1}{n} \sum_{i=1}^{n} q(y_i,\beta)$ and $V_{qn}(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} q(y_i,\beta)\, q(y_i,\beta)'$. Then, the two-step H_q-efficient GMM estimator $\hat{\beta}$ for β_0 minimizes $q_n(\beta)' \left[V_{qn}\left(\hat{\beta}^1\right)\right]^{-1} q_n(\beta)$, where $V_{qn}\left(\hat{\beta}^1\right)$ is, under H_q, a positive semi-definite consistent estimator for $V_q \equiv E_q\left[q(y,\beta_0)\, q(y,\beta_0)'\right]$, the limiting covariance matrix of the random vector $q_n(\beta_0)$, and $\hat{\beta}^1$ denotes a preliminary H_q-consistent estimator for β_0.

With regard to GEL estimation, the parameters of the H_g model are obtained by optimizing

$$R_g(\theta,\phi) = \sum_{i=1}^{n} h\left[\phi' g(y,\theta)\right], \tag{5.3}$$

while the GEL criterion appropriate for H_q is

$$R_q(\beta,\mu) = \sum_{i=1}^{n} h\left[\mu' q(y,\beta)\right], \tag{5.4}$$

where ϕ and μ are s_g- and s_q-vectors of auxiliary parameters, respectively; see section 2.5.3. In order to create convenient and simple forms for the non-nested tests discussed in later sections, the carrier function $h(\cdot)$ is chosen identically in both cases.

Of particular importance for the construction of non-nested tests of H_g against H_q is the asymptotic behaviour of H_q-estimators and associated statistics under H_g. For GEL, let μ_* and β_* denote the saddle point of $E_g\{h[\mu' q(y,\beta)]\}$, which is also the probability limit of the estimated normalized GEL criterion $\frac{1}{n} R_q\left(\hat{\beta},\hat{\mu}\right)$ under H_g.[1] Hence, $\hat{\mu} \xrightarrow{p} \mu_*$ and $\hat{\beta} \xrightarrow{p} \beta_*$, where $\xrightarrow{p}$ denotes convergence in probability, and μ_* and β_* are the *pseudo-true values* of the GEL estimators $\hat{\mu}$ and $\hat{\beta}$ under H_g. To avoid the possibility of observational equivalence between the H_g- and H_q-GEL criteria under

[1]That is, $E_g\left\{\nabla h\left[\mu_*' q(y,\beta_*)\right] q(y,\beta_*)\right\} = 0$ and $E_g\left\{\nabla h\left[\mu_*' q(y,\beta_*)\right] \frac{\partial q(y,\beta_*)'}{\partial \beta'} \mu_*\right\} = 0$; cf. (2.64) and (2.65).

H_g and, hence, to allow non-degenerate comparisons to be made, it is assumed that $E_g\{h[\mu_*' q(y,\beta_*)]\} < h(0)$, which ensures that $\mu_* \neq 0$. For GMM, $\hat{\beta} \xrightarrow{p} \beta_*$ under H_g, with the GMM pseudo-true value β_* solving $Q_g(\beta_*)'[V_{qg}(\beta_{**})]^{-1} E_g[q(y,\beta_*)] = 0$, where $Q_g(\beta) \equiv E_g\left[\frac{\partial q(y,\beta_*)}{\partial \beta'}\right]$, $V_{qg}(\beta) \equiv E_g[q(y,\beta_*)q(y,\beta_*)']$ and $\hat{\beta}^1 \xrightarrow{p} \beta_{**}$, the pseudo-true value of $\hat{\beta}^1$ under H_g.

5.2.2 Generalized non-nested tests

As will be seen in the next sub-sections and as observed by Smith (1997), non-nested test statistics for moment condition models are expressible, at least in an asymptotic sense, as linear combinations of the estimated sample H_g-moment vector $g_n\left(\hat{\theta}\right)$ as this vector represents the sole information feasible and available for inference purposes. Therefore, following Singleton (1985), let $\hat{c}_n$ denote a $(s_g \times s_m)$ random matrix that converges, under H_g, to a nonstochastic, non-zero vector c_g, $\hat{c}_n \xrightarrow{p} c_g$, and carries information concerning the alternative H_q. Assume that $M_G' c_g \neq 0$, where $M_G \equiv I_{s_g} - G\left(G'V_g^{-1}G\right)^{-1} G'V_g^{-1}$. Consider the statistic $\hat{c}_n' g_n\left(\hat{\theta}\right)$, where $\hat{\theta}$ denotes either a two-step H_g-efficient GMM estimator or a GEL estimator. Under H_g, noting from (4.30) that $\sqrt{n} g_n\left(\hat{\theta}\right) = M_G \sqrt{n} g_n(\theta_0) + O_p\left(n^{-\frac{1}{2}}\right)$, it follows that

$$\sqrt{n}\hat{c}_n' g_n\left(\hat{\theta}\right) \xrightarrow{d} N\left(0, c_g' M_G V_g M_G' c_g\right). \tag{5.5}$$

Cox-type non-nested statistics are univariate in construction, $s_m = 1$, and, therefore, a general form of Cox-type statistic to test H_g against H_q is

$$GC_n = \left(\hat{c}_n' \hat{M}_G \hat{V}_{gn} \hat{M}_G' \hat{c}_n\right)^{-\frac{1}{2}} \sqrt{n}\hat{c}_n' g_n\left(\hat{\theta}\right) \xrightarrow{d} N(0,1) \tag{5.6}$$

under H_g, where $\hat{V}_{gn} \equiv V_{gn}\left(\hat{\theta}\right)$ and $\hat{M}_G = I_{s_g} - \hat{G}_n\left(\hat{G}_n'\hat{V}_{gn}^{-1}\hat{G}_n\right)^{-1}\hat{G}_n'\hat{V}_{gn}^{-1}$ with $\hat{G}_n$ and $\hat{V}_{gn}$ H_g-consistent estimators for G and V_g. A test of H_g may then be based on a two-sided test constructed from (5.6) using critical values from the standard normal distribution.

Encompassing-type non-nested statistics possess a multivariate basis and, thus, a general form for an encompassing-type statistic to test H_g against H_q is

$$GE_n = ng_n\left(\hat{\theta}\right)' \hat{c}_n \hat{\Psi}_g^- \hat{c}_n' g_n\left(\hat{\theta}\right) \xrightarrow{d} \chi^2_{rk(\Psi_g)}, \tag{5.7}$$

under H_g, where $\hat{\Psi}_g^-$ denotes a H_g-consistent estimator for a generalized inverse of $\Psi_g \equiv c_g' M_G V_g M_G' c_g$. A test of H_g may then be based on (5.7) using critical values from the chi-square distribution with $rk\left(\Psi_g\right)$ degrees of freedom.

5.3 Non-nested Cox-type tests

5.3.1 GMM non-nested Cox-type tests

For the case of equal numbers of moment conditions under both H_g and H_q, that is, $s_g = s_q$, Singleton (1985) proposed a GC_n statistic with $c_g = V_g^{-1} E_g\left[q\left(y_i, \beta_*\right)\right]$. He showed that this choice produces an asymptotically locally most powerful test statistic against the sequence of local alternatives $H_{gn} : E_g\left[g\left(y, \theta_0\right)\right] = \upsilon\left(\zeta_n\right) E_g\left[q\left(y, \beta_*\right)\right]$, where $\zeta_n = \zeta_0 + n^{-1/2}\eta$, $\eta \neq 0$, $\upsilon\left(\zeta_0\right) = 0$, $\upsilon\left(\zeta_n\right) \neq 0$ and $\nabla\upsilon\left(\zeta_0\right) \neq 0$; cf. Singleton [1985, eq. (29), pp. 403-404]. He suggested estimating c_g by $\hat{c}_n = V_{gn}^{-1}\left[q_n\left(\hat{\beta}\right) - g_n\left(\hat{\theta}\right)\right]$ which has the merit of possessing a non-zero probability limit under both H_g and H_q. Another possible choice, analyzed in our Monte Carlo study in section 5.6, is $\hat{c}_n = \hat{V}_{gn}^{-1} q_n\left(\hat{\beta}\right)$, since $g_n\left(\hat{\theta}\right)$ converges in probability to zero under H_g. The main drawback of Singleton's (1985) test is the requirement of the existence of the same number of moment conditions in both the competing models. Conversely, its computation is very simple and quick.

To deal with situations in which $s_g \neq s_q$, Smith (1992) contrasts H_g-consistent estimators of the probability limit of the H_q-GMM criterion function evaluated at the corresponding pseudo-true value; *viz.* $E_g\left[q\left(y_i, \beta_*\right)\right]' \left[V_{qg}\left(\beta_{**}\right)\right]^{-1} E_g\left[q\left(y, \beta_*\right)\right]$. Smith's (1992) test statistic reduces to choosing $c_g = A_g' \left[V_{qg}\left(\beta_*\right)\right]^{-1} E_g\left[q\left(y_i, \beta_*\right)\right]$ in (5.6), where A_g is some finite and non-null $\left(s_q \times s_g\right)$ matrix with $rk\left(A_g\right) = \min\left(s_g, s_q\right)$, and $\hat{c} = \hat{A}'_{gn}\hat{V}_{qn}^{-1}q_n\left(\hat{\beta}\right)$, where $\hat{A}_{gn}$ is a H_g-consistent estimator for A_g and $\hat{V}_{qn} \equiv V_{qn}\left(\hat{\beta}\right)$.[2] Another possible choice, analyzed in our Monte Carlo study in section 5.6, is $\hat{c}_n = \hat{A}'_{gn}\hat{V}_{qn}^{-1}\left[q_n\left(\hat{\beta}\right) - \hat{A}_{gn}g_n\left(\hat{\theta}\right)\right]$, which is similar in spirit to Singleton's (1985) suggestion.

5.3.2 GEL non-nested Cox-type tests

Smith (1997) proposed some Cox-type tests based on the H_q-GEL criterion (5.4) using a contrast between consistent estimators for its probability limit $E_g\left\{h\left[\mu'_*q\left(y,\beta_*\right)\right]\right\}$ under H_g. The normalized optimized criterion $\frac{1}{n}R_q\left(\hat{\beta},\hat{\mu}\right)$ provides one such estimator. A second estimator is obtained from optimization of the reweighted H_q-GEL criterion

$$R_q^*\left(\beta,\mu\right) = \sum_{i=1}^{n}\hat{p}_i^g h\left[\mu'q\left(y_i,\beta\right)\right], \tag{5.8}$$

where $\hat{p}_i^g \equiv \frac{\nabla h\left[\hat{\phi}'g\left(y,\hat{\theta}\right)\right]}{\sum_{i=1}^{n}\nabla h\left[\hat{\phi}'g\left(y,\hat{\theta}\right)\right]}$, $i = 1,...,n$, denotes the H_g-implied probability measures used throughout this thesis; cf. section 2.5.4. Denote the corresponding saddle point estimators for β and μ by $\tilde{\beta}$ and $\tilde{\mu}$, respectively. Because, under H_g, $\hat{p}_i^g = \frac{1}{n}\left[1 + O_p\left(n^{-\frac{1}{2}}\right)\right]$, $i = 1,...,n$, see (4.9), $\tilde{\beta}$ and $\tilde{\mu}$ are also consistent estimators for β_* and μ_*, respectively, rendering $R_q^*\left(\tilde{\beta},\tilde{\mu}\right)$ as a consistent estimator for $E_g\left\{h\left[\mu'_*q\left(y,\beta_*\right)\right]\right\}$. Under H_g, the normalized contrast of optimized GEL criteria

$$\sqrt{n}\left[\frac{1}{n}R_q\left(\hat{\beta},\hat{\mu}\right) - R_q^*\left(\tilde{\beta},\tilde{\mu}\right)\right] \xrightarrow{d} N\left(0,\xi'_gM'_GV_g^{-1}M_G\xi_g\right) \tag{5.9}$$

if $M_G\xi_g \neq 0$, where $\xi_g \equiv E_g\left\{g\left(y_i,\theta_0\right)h\left[\mu'_*q\left(y_i,\beta_*\right)\right]\right\}$; see Smith (1997). Hence, under H_g, the GEL non-nested Cox-type statistic for H_g against H_q is given by

$$C_n = \left(\hat{\xi}'_n\hat{M}'_G\hat{V}_{gn}^{-1}\hat{M}_G\hat{\xi}_n\right)^{-\frac{1}{2}}\sqrt{n}\left[\frac{1}{n}R_q\left(\hat{\beta},\hat{\mu}\right) - R_q^*\left(\tilde{\beta},\tilde{\mu}\right)\right] \xrightarrow{d} N\left(0,1\right), \tag{5.10}$$

[2] A possible choice is $A_g = E_g\left[q\left(y,\beta_*\right)g\left(y,\theta_0\right)'\right]V_g^{-1}$, which solves the minimization problem $\min_{\beta,A}E_g\left\{\left[q\left(y,\beta\right) - Ag\left(y,\theta\right)\right]'\left[V_q\left(\beta\right)\right]^{-1}\left[q\left(y,\beta\right) - Ag\left(y,\theta\right)\right]\right\}$, so that the alternative H_q is "closest" to H_g. Thus, $\hat{A}_{gn} = \frac{1}{n}\sum_{i=1}^{n}\left[q\left(y,\hat{\beta}\right)g\left(y,\hat{\theta}\right)'\right]\hat{V}_g^{-1}$; cf. section 5.6.1 and Smith (1992).

where $\hat{\xi}_n$ is a H_g-consistent estimator for ξ_g. For example, $\hat{\xi}_n = \frac{1}{n}\sum_{i=1}^n g\left(y_i, \hat{\theta}\right) h\left[\hat{\mu}' q\left(y_i, \hat{\beta}\right)\right]$, $\hat{\xi}_n = \sum_{i=1}^n \hat{p}_i^g g\left(y_i, \hat{\theta}\right) h\left[\hat{\mu}' q\left(y_i, \hat{\beta}\right)\right]$, $\hat{\xi}_n = \frac{1}{n}\sum_{i=1}^n g\left(y_i, \hat{\theta}\right) h\left[\tilde{\mu}' q\left(y_i, \tilde{\beta}\right)\right]$ or $\hat{\xi}_n = \sum_{i=1}^n \hat{p}_i^g g\left(y_i, \hat{\theta}\right) h\left[\tilde{\mu}' q\left(y_i, \tilde{\beta}\right)\right]$.

The limit distribution (5.10) of C_n is obtained *via* a first-order Taylor expansion of the contrast about $(\theta_0, 0)$ and $\left(\hat{\beta}, \hat{\mu}\right)$. This expansion suggests two further statistics which are asymptotically equivalent tests to C_n under H_g: a linearized Cox-type statistic

$$LC_n = -\frac{\nabla^2 h(0)}{\nabla h(0)}\left(\hat{\xi}_n' \hat{M}_G' \hat{V}_{gn}^{-1} \hat{M}_G \hat{\xi}_n\right)^{-\frac{1}{2}} \hat{\xi}_n' \sqrt{n}\hat{\phi}, \tag{5.11}$$

and a simplified Cox-type statistic

$$SC_n = \left(\hat{\xi}_n' \hat{M}_G' \hat{V}_{gn}^{-1} \hat{M}_G \hat{\xi}_n\right)^{-\frac{1}{2}} \sqrt{n}\left[\frac{1}{n} R_q\left(\hat{\beta}, \hat{\mu}\right) - R_q^*\left(\hat{\beta}, \hat{\mu}\right)\right], \tag{5.12}$$

both of which require one less optimization than C_n; see Smith (1997). The form of LC_n indicates that, asymptotically, these GEL statistics correspond to choosing $c_g = V_g^{-1}\xi_g$ in GC_n of (5.6); recall from (2.66) and (4.30) that

$$\sqrt{n}\frac{\nabla^2 h(0)}{\nabla h(0)}\hat{\phi} = -V_g^{-1}\sqrt{n} g_n\left(\hat{\theta}\right) + O_p\left(n^{-\frac{1}{2}}\right). \tag{5.13}$$

5.4 Parametric encompassing tests

5.4.1 GMM parametric encompassing tests

Smith (1992) proposes a GMM parametric encompassing (PE) test based on the contrast of two H_g-consistent estimators for the pseudo true value β_*. One is the standard efficient two-step GMM estimator $\hat{\beta}$, the other is obtained from a Taylor's series expansion of the first-order conditions defining $\hat{\beta}$ around β_*, being given by

$$\hat{\beta}_* = \hat{\beta} - \left(\hat{Q}_n' \hat{V}_{qn}^{-1} \hat{Q}_n\right)^{-1} \hat{Q}_n' \hat{V}_{qn}^{-1} \hat{A}_{gn} g_n\left(\hat{\theta}\right), \tag{5.14}$$

where $\hat{Q}_n \equiv \sum_{i=1}^{n} \frac{\partial q\left(y_i,\hat{\beta}\right)}{\partial \beta'}$ is a consistent estimator for $Q_g \equiv Q_g\left(\beta_*\right)$. From (5.14), the resultant GMM PE test statistic is

$$E_n = n\left(\hat{\beta}-\hat{\beta}_*\right)'\hat{\Omega}_g^-\left(\hat{\beta}-\hat{\beta}_*\right), \tag{5.15}$$

where $\hat{\Omega}_g^-$ denotes a H_g-consistent estimator for a g-inverse of $\Omega_g = \left(Q_g'V_{qg}^{-1}Q_g\right)^{-1}$ $Q_g'V_{qg}^{-1}A_gM_GV_gM_G'A_g'V_{qg}^{-1}Q_g\left(Q_g'V_{qg}^{-1}Q_g\right)^{-1}$, $V_{qg} \equiv V_{qg}\left(\beta_*\right)$. Under H_g, this statistic has a limiting chi-squared distribution with $rk\left(\Omega\right) \leq \min\left(k_q, s_g - k_g\right)$ degrees of freedom. See Smith (1992).

Using (5.14), the statistic E_n may also be written as

$$E_n = ng_n\left(\hat{\theta}\right)'\hat{A}_{gn}'\hat{V}_{qn}^{-1}\hat{Q}_n\hat{\Psi}_g^-\hat{Q}_n'\hat{V}_{qn}^{-1}\hat{A}_{gn}g_n\left(\hat{\theta}\right), \tag{5.16}$$

where $\hat{\Psi}_g^-$ denotes a H_g-consistent estimator for a g-inverse of $\Psi_g = Q_g'V_{qg}^{-1}A_gM_GV_gM_G'$ $A_g'V_{qg}^{-1}Q_g$. From (5.16), it is clear that E_n corresponds to choosing $c_g = A_g'V_{qg}^{-1}Q_g$ in GE_n of (5.7).

5.4.2 GEL parametric encompassing tests

In this sub-section we derive a GEL PE test statistic based on the normalized contrast between $\hat{\mu}$ and $\tilde{\mu}$ and $\hat{\beta}$ and $\tilde{\beta}$:

$$\sqrt{n}\begin{pmatrix}\hat{\beta}-\tilde{\beta}\\ \hat{\mu}-\tilde{\mu}\end{pmatrix}. \tag{5.17}$$

To evaluate the limiting distribution of (5.17), we need to examine the first order conditions defining $\left(\hat{\beta},\hat{\mu}\right)$ and $\left(\tilde{\beta},\tilde{\mu}\right)$. The H_q-GEL criterion $R_q\left(\beta,\mu\right)$ and the reweighted GEL criterion $R_q^*\left(\beta,\mu\right)$ have first-order conditions

$$\sum_{i=1}^{n}\nabla h\left[\hat{\mu}'q\left(y_i,\hat{\beta}\right)\right]\begin{bmatrix}\frac{\partial q\left(y_i,\hat{\beta}\right)'}{\partial\beta'}\hat{\mu}\\ q\left(y_i,\hat{\beta}\right)\end{bmatrix} = 0 \tag{5.18}$$

and

$$\sum_{i=1}^{n} \hat{p}_i^g \nabla h \left[\hat{\mu}' q\left(y_i, \hat{\beta}\right)\right] \left[\begin{array}{c} \frac{\partial q\left(y_i, \tilde{\beta}\right)'}{\partial \beta'} \tilde{\mu} \\ q\left(y_i, \tilde{\beta}\right) \end{array} \right] = 0, \tag{5.19}$$

respectively. Expanding both sets of moment conditions about (β_*, μ_*) yields

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \nabla h \left[\mu_*' q\left(y_i, \beta_*\right)\right] \left[\begin{array}{c} \frac{\partial q(y_i, \beta_*)'}{\partial \beta'} \mu_* \\ q\left(y_i, \beta_*\right) \end{array} \right] + K_g \sqrt{n} \left[\begin{array}{c} \hat{\beta} - \beta_* \\ \hat{\mu} - \mu_* \end{array} \right] + O_p\left(n^{-\frac{1}{2}}\right) \tag{5.20}$$

and

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \sqrt{n} \sum_{i=1}^{n} \hat{p}_i^g \nabla h \left[\mu_*' q\left(y_i, \beta_*\right)\right] \left[\begin{array}{c} \frac{\partial q(y_i, \beta_*)'}{\partial \beta'} \mu_* \\ q\left(y_i, \beta_*\right) \end{array} \right] + K_g \sqrt{n} \left[\begin{array}{c} \tilde{\beta} - \beta_* \\ \tilde{\mu} - \mu_* \end{array} \right] + O_p\left(n^{-\frac{1}{2}}\right), \tag{5.21}$$

where

$$\begin{aligned} K_g \equiv\ & E_g \left\{ \nabla h \left[\mu_*' q\left(y_i, \beta_*\right)\right] \left[\begin{array}{cc} \sum_{j=1}^{s_q} \frac{\partial^2 q_j(y_i, \beta_*)}{\partial \beta \partial \beta'} \mu_{*j} & \frac{\partial q(y_i, \beta_*)'}{\partial \beta'} \\ \frac{\partial q(y_i, \beta_*)}{\partial \beta'} & 0 \end{array} \right] \right\} \\ & + E_g \left\{ \nabla^2 h \left[\mu_*' q\left(y_i, \beta_*\right)\right] \left[\begin{array}{c} \frac{\partial q(y_i, \beta_*)'}{\partial \beta} \mu_* \\ q\left(y_i, \beta_*\right) \end{array} \right] \left[\begin{array}{cc} \mu_*' \frac{\partial q(y_i, \beta_*)}{\partial \beta'} & q\left(y_i, \beta_*\right)' \end{array} \right] \right\} \end{aligned} \tag{5.22}$$

Subtracting (5.21) from (5.20) produces

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \hat{\beta} - \tilde{\beta} \\ \hat{\mu} - \tilde{\mu} \end{pmatrix} &= \sqrt{n} K_g^{-1} \sum_{i=1}^{n} \left(\hat{p}_i^g - \frac{1}{n}\right) \nabla h \left[\mu_*' q\left(y_i, \beta_*\right)\right] \left[\begin{array}{c} \frac{\partial q(y_i, \beta_*)'}{\partial \beta} \mu_* \\ q\left(y_i, \beta_*\right) \end{array} \right] + O_p\left(n^{-\frac{1}{2}}\right) \\ &= K_g^{-1} \sum_{i=1}^{n} \nabla h \left[\mu_*' q\left(y_i, \beta_*\right)\right] \left[\begin{array}{c} \frac{\partial q(y_i, \beta_*)'}{\partial \beta'} \mu_* \\ q\left(y_i, \beta_*\right) \end{array} \right] \sqrt{n} \left(\hat{p}_i^g - \frac{1}{n}\right) \\ &\quad + O_p\left(n^{-\frac{1}{2}}\right). \end{aligned} \tag{5.23}$$

Now, substituting $\frac{\nabla^2 h(0)}{\nabla h(0)} \frac{1}{n} g\left(y_i, \theta_0\right)' \sqrt{n} \hat{\phi} + O_p\left(n^{-\frac{3}{2}}\right)$ for $\sqrt{n}\left(\hat{p}_i^g - \frac{1}{n}\right)$ in (5.23), see

(4.9), it follows that, under H_g,

$$\sqrt{n}\begin{pmatrix}\hat{\beta}-\tilde{\beta}\\ \hat{\mu}-\tilde{\mu}\end{pmatrix}=\frac{\nabla^2 h(0)}{\nabla h(0)}K_g^{-1}W_g\sqrt{n}\hat{\phi}+O_p\left(n^{-\frac{1}{2}}\right), \tag{5.24}$$

where

$$W_g=E_g\left\{\nabla h\left[\mu_*'q(y_i,\beta_*)\right]\begin{bmatrix}\frac{\partial q(y_i,\beta_*)'}{\partial\beta'}\mu_*\\ q(y_i,\beta_*)\end{bmatrix}g(y_i,\theta_0)'\right\}. \tag{5.25}$$

Hence, the GEL PE test statistic is given by

$$PE_n=n\begin{pmatrix}\hat{\beta}-\tilde{\beta}\\ \hat{\mu}-\tilde{\mu}\end{pmatrix}'\hat{K}_{gn}\hat{\Psi}_g^{-}\hat{K}_{gn}\begin{pmatrix}\hat{\beta}-\tilde{\beta}\\ \hat{\mu}-\tilde{\mu}\end{pmatrix}, \tag{5.26}$$

where $\hat{K}_{gn}$ denotes a H_g-consistent estimators for K_g and $\hat{\Psi}_g^{-}$ is a H_g-consistent estimator for a g-inverse of $\Psi_g\equiv W_gM_G'V_g^{-1}M_GW_g'$ which is assumed non-null. The statistic PE_n has a limiting chi-square distribution under H_g with degrees of freedom equal to $rk(\Psi_g)$ whose critical values provide a basis for a test of H_g against H_q. Comparing (5.24) and (5.13), this test corresponds asymptotically to the choice $c_g=V_g^{-1}W_g'$ in the GE_n statistic (5.7).

A linearized statistic which is asymptotically equivalent to PE_n and avoids the necessity of providing the estimator $\hat{K}_{gn}$ and the GEL estimators $\tilde{\beta}$ and $\tilde{\mu}$ is obtained by noting from (5.23) that

$$\begin{aligned}K_g\sqrt{n}\begin{pmatrix}\hat{\beta}-\tilde{\beta}\\ \hat{\mu}-\tilde{\mu}\end{pmatrix}&=\sqrt{n}\sum_{i=1}^{n}\left(\hat{p}_i^g-\frac{1}{n}\right)\nabla h\left[\hat{\mu}'q\left(y_i,\hat{\beta}\right)\right]\begin{bmatrix}\frac{\partial q\left(y_i,\hat{\beta}\right)'}{\partial\beta'}\hat{\mu}\\ q\left(y_i,\hat{\beta}\right)\end{bmatrix}+O_p\left(n^{-\frac{1}{2}}\right)\\ &=\sqrt{n}\sum_{i=1}^{n}\hat{p}_i^g\nabla h\left[\hat{\mu}'q\left(y_i,\hat{\beta}\right)\right]\begin{bmatrix}\frac{\partial q\left(y_i,\hat{\beta}\right)'}{\partial\beta'}\hat{\mu}\\ q\left(y_i,\hat{\beta}\right)\end{bmatrix}+O_p\left(n^{-\frac{1}{2}}\right)\end{aligned} \tag{5.27}$$

[see also the first-order conditions (5.18)], which may be regarded as a re-weighted H_q-score and has a limiting normal distribution with variance matrix Ψ_g. Other encompassing statistics may be based on sub-vectors of (5.27), for example,

$$\sqrt{n}\sum_{i=1}^{n}\left(\hat{p}_i^g-\frac{1}{n}\right)\nabla h\left[\hat{\mu}'q\left(y_i,\hat{\beta}\right)\right]q\left(y_i,\hat{\beta}\right)=\sqrt{n}\sum_{i=1}^{n}\hat{p}_i^g\nabla h\left[\hat{\mu}'q\left(y_i,\hat{\beta}\right)\right]q\left(y_i,\hat{\beta}\right)+O_p\left(n^{-\frac{1}{2}}\right). \tag{5.28}$$

5.5 Moment encompassing tests

5.5.1 GMM moment encompassing tests

Ghysels and Hall (1990b) suggested a moment encompassing (ME) test for H_g against H_q using a contrast of two H_g-consistent estimators for $E_g\left[q\left(y_i,\beta_*\right)\right]$. Their test statistic is based on the difference between $q_n\left(\hat{\beta}\right)$ and $\frac{1}{n}\sum_{i=1}^{n}E_g\left[q\left(y_i,\beta_*\right)\right]$. The implementation of this statistic requires the specification of the data generation process of the maintained model in order to simulate $\frac{1}{n}\sum_{i=1}^{n}E_g\left[q\left(y_i,\beta_*\right)\right]$, which runs counter the spirit of GMM estimation and inference. Consequently, Ghysels and Hall's (1990b) statistic may reject H_g not due to the falsity of H_g but rather because the additional assumptions made might not hold in the population. Moreover, their statistic is computationally very intensive [see Ghysels and Hall (1990b), pp. 288-289].

5.5.2 GEL generalized moment encompassing tests

In this sub-section we outline a general class of moment-based test statistics. Again, the basis for GEL ME tests arises from noting that $n\left(\hat{p}_i^g-\frac{1}{n}\right)=O_p\left(n^{-\frac{1}{2}}\right)$ under H_g, $i=1,...,n$. Consider an s_m-vector of moment indicators $m_q\left(y_i,\theta,\phi,\beta,\mu\right)$ which, typically, but not necessarily, is obtained from the H_q-moment condition model. Simple H_g-consistent estimators for $E_g\left[m_q\left(y_i,\theta_0,0,\beta_*,\mu_*\right)\right]$ are provided by

$$\hat{m}_{qn}=\frac{1}{n}\sum_{i=1}^{n}m_q\left(y_i,\hat{\theta},\hat{\phi},\hat{\beta},\hat{\mu}\right) \tag{5.29}$$

and

$$\hat{m}^*_{qn} = \sum_{i=1}^{n} \hat{p}^g_i m_q \left(y_i, \hat{\theta}, \hat{\phi}, \hat{\beta}, \hat{\mu} \right), \tag{5.30}$$

the latter of which circumvents the necessity in Ghysels and Hall's (1990b) statistic of specifying the distribution of the random variable Y under H_g; see sub-section 5.5.3 below. The consequent contrast underlying GEL ME statistics is

$$\sqrt{n} \left(\hat{m}_{qn} - \hat{m}^*_{qn} \right) = -\sqrt{n} \sum_{i=1}^{n} m_q \left(y_i, \hat{\theta}, \hat{\phi}, \hat{\beta}, \hat{\mu} \right) \left(\hat{p}^g_i - \frac{1}{n} \right). \tag{5.31}$$

If the null hypothesis H_g is correct, the limiting distribution of the contrast (5.31) should be centred at zero.

The limiting distribution of (5.31) is straightforward to derive. From (4.9), it follows that

$$\sqrt{n} \left(\hat{m}_{qn} - \hat{m}^*_{qn} \right) = -\frac{\nabla^2 h(0)}{\nabla h(0)} \frac{1}{n} \sum_{i=1}^{n} m_q \left(y_i, \hat{\theta}, \hat{\phi}, \hat{\beta}, \hat{\mu} \right) g(y_i, \theta_0)' \sqrt{n}\hat{\phi} + O_p \left(n^{-\frac{1}{2}} \right). \tag{5.32}$$

A further expansion of $m_q \left(y_i, \hat{\theta}, \hat{\phi}, \hat{\beta}, \hat{\mu} \right)$ about $(\theta_0, 0, \beta_*, \mu_*)$ yields, under H_g,

$$\begin{aligned} \sqrt{n} \left(\hat{m}_{qn} - \hat{m}^*_{qn} \right) &= -\frac{\nabla^2 h(0)}{\nabla h(0)} \frac{1}{n} \sum_{i=1}^{n} m_q (\theta_0, 0, \beta_*, \mu_*) g(y_i, \theta_0)' \sqrt{n}\hat{\phi} + O_p \left(n^{-\frac{1}{2}} \right) \\ &= -\frac{\nabla^2 h(0)}{\nabla h(0)} \Gamma_g \sqrt{n}\hat{\phi} + O_p \left(n^{-\frac{1}{2}} \right), \end{aligned} \tag{5.33}$$

where $\Gamma_g \equiv E_g \left[m_q (y_i, \theta_0, 0, \beta_*, \mu_*) g(y_i, \theta_0)' \right]$ is a $(s_m \times s_g)$ matrix and it is assumed that $rk(\Gamma_g) = s_m$ and $M_G \Gamma'_g \neq 0$. Therefore, under H_g, recalling from (2.68) that $\sqrt{n}\hat{\phi} \xrightarrow{d} N \left\{ 0, \left[\frac{\nabla h(0)}{\nabla^2 h(0)} \right]^2 M'_G V_g^{-1} M_G \right\}$, it follows that

$$\sqrt{n} \left(\hat{m}_{qn} - \hat{m}^*_{qn} \right) \xrightarrow{d} N(0, \Psi_g), \tag{5.34}$$

where $\Psi_g \equiv \Gamma_g M'_G V_g^{-1} M_G \Gamma'_g$, and a general form for GEL ME statistics is then given by

$$GME_n = n\left(\hat{m}_{qn} - \hat{m}^*_{qn}\right)' \hat{\Psi}^-_g \left(\hat{m}_{qn} - \hat{m}^*_{qn}\right), \tag{5.35}$$

where $\hat{\Psi}^-_g$ denotes an H_g-consistent estimator for a g-inverse of Ψ_g; for example, Γ_g may be consistently estimated by $\hat{\Gamma}_{gn} = \frac{1}{n}\sum_{i=1}^n m_q\left(y_i, \hat{\theta}, \hat{\phi}, \hat{\beta}, \hat{\mu}\right) g\left(y_i, \hat{\theta}\right)'$ or $\hat{\Gamma}_{gn} = \sum_{i=1}^n \hat{p}^g_i m_q\left(y_i, \hat{\theta}, \hat{\phi}, \hat{\beta}, \hat{\mu}\right) g\left(y_i, \hat{\theta}\right)'$. Under H_g, the generalized ME statistic GME_n has a limiting chi-squared distribution with $rk\left(\Psi_g\right)$ degrees of freedom. A test of H_g may then be based on GME_n of (5.35) using critical values from the chi-square distribution with $rk\left(\Psi_g\right)$ degrees of freedom.

From (5.33), a first order H_g-asymptotically equivalent form is given by the linearized statistic:

$$LGME_n = n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2 \hat{\phi}'\hat{\Gamma}'_n \hat{\Psi}^-_g \hat{\Gamma}_n \hat{\phi}. \tag{5.36}$$

The form of this statistic shows that, asymptotically, these GEL statistics correspond to the choice $c_g = V_g^{-1}\Gamma'_g$ in the GE_n statistic (5.7); see also (5.13).

The GME_n and $LGME_n$ statistics (5.35) and (5.36) may be used to generate more familiar statistics. For example, choosing $m_q\left(y_i, \theta, \phi, \beta, \mu\right) = g\left(y_i, \theta\right)$ and, thus, $s_m = s_g$, results in $\hat{m}_{qn} = g_n\left(\hat{\theta}\right)$ and $\hat{m}^*_{qn} = 0$. Hence, GME_n of (5.35) reduces to Hansen's (1982) J statistic for overidentifying moment restrictions as $\Gamma_g = V_g$, $\hat{\Gamma}_{gn} = \hat{V}_{gn}$ and V_g^{-1} is a g-inverse for $\Gamma_g M'_G V_g^{-1} M_G \Gamma'_g = M_G V_g M'_G$. In this sense, therefore, the generalized ME statistic may be regarded as a generalization of Hansen's (1982) J statistic. An interesting form for $LGME_n$ arises when $s_m = s_g$ and, thus, Γ_g is non-singular. In this case,

$$\begin{aligned} LGME_n &= n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2 \hat{\phi}'\hat{\Gamma}'_n \left(\hat{\Gamma}_n \hat{M}'_{Gn} \hat{V}^{-1}_{gn} \hat{M}_{Gn} \hat{\Gamma}'_n\right)^- \hat{\Gamma}_n \hat{\phi} \\ &= n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2 \hat{\phi}'\left(\hat{M}'_{Gn} \hat{V}^{-1}_{gn} \hat{M}_{Gn}\right)^- \hat{\phi} \\ &= n\left[\frac{\nabla^2 h(0)}{\nabla h(0)}\right]^2 \hat{\phi}'\hat{V}_{gn}\hat{\phi}, \end{aligned} \tag{5.37}$$

which is a GEL Wald-type test for overidentifying moment conditions; see section

2.5.6.

5.5.3 GEL moment encompassing tests

In the non-nested test framework, the choice of $m_q(y_i, \theta, \phi, \beta, \mu)$ may be determined to maximize power against H_q in a particular direction or against a particular feature of the H_q competing specification. For example, the statistics LC_n and SC_n of (5.11) and (5.12) are obtained by choosing $m_q(y_i, \theta, \phi, \beta, \mu) = h[\mu' q(y_i, \beta)]$ and, hence, $s_m = 1$ and $\Gamma_g = \xi_g'$.

The above difficulties experienced by Ghysels and Hall (1990b) occasioned by the estimator $\frac{1}{n}\sum_{i=1}^{n} E_g[q(y_i, \beta_*)]$ may simply be avoided in the GEL framework using the GME_n and $LGME_n$ statistics (5.35) and (5.36). Analogously to Ghysels and Hall's (1990b) ME statistic, choosing $m_q(y_i, \theta, \phi, \beta, \mu) = q(y_i, \beta)$ yields for $\hat{m}_{qn}$ and $\hat{m}_{qn}^*$ respectively

$$q_n\left(\hat{\beta}\right) = \frac{1}{n}\sum_{i=1}^{n} q\left(y_i, \hat{\beta}\right) \tag{5.38}$$

and

$$q_n^*\left(\hat{\beta}\right) = \sum_{i=1}^{n} \hat{p}_i^g q\left(y_i, \hat{\beta}\right), \tag{5.39}$$

and the resultant GEL ME statistic

$$ME_n = n\left[q_n\left(\hat{\beta}\right) - q_n^*\left(\hat{\beta}\right)\right]' \hat{\Psi}_g^- \left[q_n\left(\hat{\beta}\right) - q_n^*\left(\hat{\beta}\right)\right], \tag{5.40}$$

where $\Gamma_g \equiv E_g\left[q(y_i, \beta_*) g(y_i, \theta_0)'\right]$ and may be estimated by $\hat{\Gamma}_{gn} = \frac{1}{n}\sum_{i=1}^{n} q\left(y_i, \hat{\beta}\right) g\left(y_i, \hat{\theta}\right)'$. Under H_g, ME_n has a limiting chi-square distribution with $rk(\Psi_g)$ degrees of freedom. Note that this statistic does not involve an estimator for μ_* unlike PE_n of (5.26). Moreover, estimation of μ_* may be avoided altogether as any H_g consistent, not necessarily GEL, estimator for β_* may be substituted in (5.40). A linearized ME statistic LME_n may also be constructed according to (5.36).

5.6 Simulation evidence

This section explores the finite sample size and power properties of some of the non-nested test statistics discussed above in a linear instrumental variable (IV) model context using Monte Carlo methods.

The tests considered in these simulation experiments fall into two groups: those associated with GMM and GEL. In the first group, we analyze Singleton's (1985) test using two different estimators for $E_g[q(y_i, \beta_*)]$ given by $q_n\left(\hat{\beta}\right)$ and $\left[q_n\left(\hat{\beta}\right) - g_n\left(\hat{\theta}\right)\right]$, labelled S and AS respectively; Smith's (1992) Cox-type test based on $q_n\left(\hat{\beta}\right)$ and $\left[q_n\left(\hat{\beta}\right) - \hat{A}_{gn} g_n\left(\hat{\theta}\right)\right]$, labelled C and AC respectively; and Smith's (1992) encompassing test (5.16) tests, labelled E; see sections 5.3.1 and 5.4.1. For comparison purposes, we also consider Hansen's (1982) J test of overidentifying moment conditions; see section 2.3.4. The group of GEL-based tests includes Smith's (1997) linearized Cox (LC) and simplified Cox (SC) tests of (5.11) and (5.12), respectively; the linearized form of the PE statistic (PE) of (5.26) obtained using (5.27); and the ME statistic (ME) of (5.40) and its linearized counterpart (LME); see sections 5.3.2, 5.4.2 and 5.5.3.

The H_g-consistent matrix estimators necessary for the computation of the GEL-based tests (namely, $\hat{G}_n$ and $\hat{V}_{gn}$, for all tests, $\hat{\xi}_n$, for the GEL Cox-type tests, $\hat{W}_{gn}$ for the PE test and $\hat{\Gamma}_{gn}$ for both ME and LME tests) were obtained *via* the $gel(s)$ and $gel(r)$ methods described in section 4.5.1. All tests in the GMM group were calculated using the methods $gel(n)$ and $gel(s)$. EL estimation is considered. The EL implied probabilities are given by $\hat{p}_i^g = \frac{1}{n}\left[1 + \hat{\phi}' g\left(y_i, \hat{\theta}\right)\right]^{-1}$ and $\hat{p}_i^q = \frac{1}{n}\left[1 + \hat{\mu}' q\left(y_i, \hat{\beta}\right)\right]^{-1}$, $i = 1, ..., n$. Hence, $SC = LC$ and $PE = ME = LME$, results which are specific to EL estimation and arise because of the particular form of the EL implied probabilities; see footnote 3 in page 119.

Two questions are of special interest: (a) is the H_g-asymptotic distribution a reliable guide to the finite sample distribution of the tests? (b) how do the tests perform comparatively in terms of power?

5.6.1 Linear instrumental variable models

Consider two competing regression models

$$\begin{aligned} H_g &: y = X_g\theta_0 + u_g, &(5.41)\\ H_q &: y = X_q\beta_0 + u_q, &(5.42)\end{aligned}$$

where $y = (y_1, ..., y_n)'$, $X_g = (x_{g1}, ..., x_{gn})'$ and $X_q = (x_{q1}, ..., x_{qn})'$ are an n-vector and $(n \times k_g)$ and $(n \times k_q)$ matrices of observations on a scalar dependent variable and k_g- and k_q-dimensioned regressor variables, respectively. In all experiments, H_g (5.41) is always the null hypothesis with H_q (5.42) the alternative hypothesis. It is assumed that, under H_g (resp. H_q), the regressor matrix X_g (resp. X_q) and the error term u_g (resp. u_q) are asymptotically correlated rendering least squares estimation of θ_0 (resp. β_0) inconsistent. Consequently, we assume the availability of n observations on s_g (resp. s_q) IVs $Z_g = (z_{g1}, ..., z_{gn})'$ [resp. $Z_q = (z_{q1}, ..., z_{qn})'$], where $s_g > k_g$ (resp. $s_q > k_q$), such that, under H_g, $E_g\left(\frac{1}{n}Z_g'u_g\right) = 0$ and, under H_q, $E_q\left(\frac{1}{n}Z_q'u_q\right) = 0$. The n observations comprising y, X_g, X_q, Z_g and Z_q are assumed to be independently distributed.

In the notation of this chapter, we have, for H_g, $g(y_i, \theta) \equiv z_{gi}\left(y - x_{gi}'\theta\right)$, $i = 1, ..., n$, $g_n\left(\hat{\theta}\right) \equiv \frac{1}{n}Z_g'\hat{u}_g$, $\hat{u}_g \equiv \hat{M}_G y$, $\hat{M}_G \equiv I_{s_g} - Z_g'X_g\left(X_g'Z_g\hat{V}_{gn}^{-1}Z_g'X_g\right)^{-1}X_g'Z_g\hat{V}_{gn}^{-1}$ and $\hat{G}_n \equiv -\frac{1}{n}Z_g'X_g$. For H_q, $q(y_i, \beta) \equiv z_{qi}\left(y - x_{qi}'\beta\right)$, $i = 1, ..., n$, $q_n\left(\hat{\theta}\right) \equiv \frac{1}{n}Z_q'\hat{u}_q$, $\hat{u}_g \equiv \hat{M}_Q y$, $\hat{M}_Q \equiv I_{s_q} - Z_q'X_q\left(X_q'Z_q\hat{V}_{qn}^{-1}Z_q'X_q\right)^{-1}X_q'Z_q\hat{V}_{qn}^{-1}$ and $\hat{Q}_n \equiv -\frac{1}{n}Z_q'X_q$. Utilizing efficient GMM estimation, explicit expressions can be found for the estimators:

$$\begin{aligned} H_g &: \hat{\theta} = \left(X_g'Z_g\hat{V}_{gn}Z_g'X_g\right)^{-1}X_g'Z_g\hat{V}_{gn}Z_g'y, &(5.43)\\ H_q &: \hat{\beta} = \left(X_q'Z_q\hat{V}_{qn}Z_q'X_q\right)^{-1}X_q'Z_q\hat{V}_{qn}Z_q'y. &(5.44)\end{aligned}$$

The implementation of the GMM non-nested tests C, AC, and E requires a choice of the matrix A_g and its associated estimator $\hat{A}_{gn}$; see section 5.3.1. Smith (1992) suggested two forms for $\hat{A}_{gn}$ in the IV context: $\hat{A}_{gn}^1 = \frac{1}{n}Z_q'Z_g\hat{V}_{gn}^{-1}$ and $\hat{A}_{gn}^2 =$

$Z_q'Z_g\left(Z_g'Z_g\right)^{-1}$. We only consider the latter choice. Indeed, the simulation design, described in more detail below, assumes homoskedastic errors under both H_g and H_q. Hence, if $\hat{V}_{gn}^{-1}$ were set as $n\frac{1}{\hat{\sigma}_{gn}^2}\left(Z_g'Z_g\right)^{-1}$, where $\hat{\sigma}_{gn}^2$ is the IV estimator for the variance of the elements of the error vector u_g, $\hat{A}_{gn}^2 = \hat{\sigma}_{gn}^2\hat{A}_{gn}^1$, so identical results for the C and E statistics would be obtained for both choices of $\hat{A}_{gn}$. However, the AC statistic is not invariant to scale transformations when based on $\hat{A}_{gn}^1$. Therefore, only the matrix $\hat{A}_{gn}^2$ is used in our experiments.

Numerous factors affect the performance of non-nested tests when applied to IV models. We concentrate on two main aspects: (a) when the H_g instruments are invalid under H_q (Design I); (b) when the forms of the H_g and H_q regressions differ (Design II). In both cases, we considered two sample sizes, $n = 200$ and $n = 400$. Each Monte Carlo experiment comprised 2000 replications.

5.6.2 Monte Carlo experiment I

In both null and alternative IV regression models, to aid the interpretability of the simulation results, the number of regressors comprising X_g and X_q was fixed to be unity, $k_g = k_q = 1$. To make the use of IVs necessary, the regressors of the null and the alternative models were generated according to the design

$$X_g = \epsilon + \tau\mu + \lambda u_g \tag{5.45}$$

$$X_q = \mu + \psi u_q, \tag{5.46}$$

where random n-vectors ϵ and μ and error vectors u_g and u_q are independent $N(0, I_n)$ vectors. The parameters λ, τ and ψ are the covariances between, respectively, X_g and u_g, X_g and X_q and X_q and u_q. These parameters allow the corresponding correlations

$$\rho_{xu_g} = \lambda\left(1 + \tau^2 + \lambda^2\right)^{-1/2}, \tag{5.47}$$

$$\rho_{xx} = \tau\left[\left(1 + \psi^2\right)\left(1 + \tau^2 + \lambda^2\right)\right]^{-1/2}, \tag{5.48}$$

and

$$\rho_{xu_q} = \psi \left(1 + \psi^2\right)^{-1/2}, \tag{5.49}$$

to be controlled. The dependent variable y was generated under the null hypotheses H_g (5.41) and under the alternative hypotheses H_q (5.42) with $\theta_0 = \beta_0 = 1$.

The H_g matrix of IVs Z_g, $s_g = 4$, was generated *via*

$$Z_{gj} = \varphi_1 \epsilon + \varsigma_j, \; j = 1, 2, \tag{5.50}$$

$$Z_{gj} = \varphi_2 \epsilon + \sigma u_q + \varsigma_j, \; j = 3, 4, \tag{5.51}$$

where $\varsigma_j \sim IN\left(0, I_n\right)$, $j = 1, ..., 4$, are independent of ϵ, μ, u_g and u_q. The parameter φ_1 in the first set of IVs (5.50) allow a degree of control over the correlation ρ_{xz_g} between Z_{gj}, $j = 1, 2$, and the regressor X_g:

$$\rho_{xz_{gj}} = \varphi_1 \left[\left(1 + \varphi_1^2\right)\left(1 + \tau^2 + \lambda^2\right)\right]^{-1/2}, \; j = 1, 2. \tag{5.52}$$

For the second set of IVs (5.51),

$$\rho_{xz_{gj}} = \varphi_2 \left[\left(1 + \varphi_2^2 + \sigma^2\right)\left(1 + \tau^2 + \lambda^2\right)\right]^{-1/2}, \; j = 3, 4, \tag{5.53}$$

with φ_2 set to equate $\rho_{xz_{gj}}$, $j = 1, 2$, and $\rho_{xz_{gj}}$, $j = 3, 4$.

The IVs Z_q, $s_q = 4$, of the alternative model H_q were generated to ensure their correlation with both sets of regressors X_g and X_q and no correlation with the error terms u_g and u_q:

$$Z_{qj} = \gamma \mu + \omega_j, \; j = 1, ..., 4, \tag{5.54}$$

where $\omega_j \sim IN\left(0, I_n\right)$, $j = 1, ..., 4$, are independent of ς_j, $j = 1, ..., 4$, ϵ, μ, u_g and u_q. The parameter γ, which represents the covariance between the IVs Z_q and the H_q regressor X_q, enables control of their correlation ρ_{xz_q} *via*:

$$\rho_{xz_q} = \gamma \left[\left(1 + \gamma^2\right)\left(1 + \psi^2\right)\right]^{-1/2}. \tag{5.55}$$

Although all IVs Z_{gj}, $j = 1, ..., 4$, from (5.50) and (5.51) are valid instruments under H_g, only the IVs Z_{gj}, $j = 1, 2$, are also valid under the alternative hypotheses H_q. As the structures of the two competing models H_g and H_q are rather similar, the IVs Z_{gj}, $j = 3, 4$, from (5.51), should be the main source of misspecification of the H_g model (5.41) when the alternative hypothesis H_q model (5.42) is correct. We anticipate, therefore, that the ability of the various tests to reject the false H_g model will decisively depend on the degree of misspecification of these IVs Z_{gj}, $j = 3, 4$, measured by their correlation with the error term u_q:

$$\rho_{z_g u_q} = \sigma \left(1 + \varphi_2^2 + \sigma^2\right)^{-1/2}. \tag{5.56}$$

Table 5.1 reports empirical sizes for each of the various non-nested statistics detailed above applied to test the IV models H_g (5.41) against H_q (5.42). The nominal size for all tests is 0.05 based on critical values taken from their H_g asymptotic distributions. To relate the behaviour of the tests to the quality of the instruments utilized, two distinct values of ρ_{xu_g} (0.3 and 0.6) and ρ_{xz_g} (0.25 and 0.5) were considered. To check the effect of different degrees of proximity between the regressors of the two competing models, we simulated experiments for two distinct values of ρ_{xx}: 0.2 and 0.4. We fixed $\psi = 1$, $\gamma = 1$ and $\sigma = 0$ in these experiments.

From Table 5.1, for tests in the GMM group, size behaviour for both sample sizes does not appear to be markedly affected by the correlation between the regressor X_g and IVs Z_g (ρ_{xz_g}), except for E and J at $n = 200$. These two statistics seem also sensitive to the feedback from u_g to X_g (ρ_{xu_g}). The C and AC tests had the best performances, with empirical sizes close to the nominal ones in all cases. Overall, the adjustment to the Cox statistic suggested in section 5.3.1 does not appear to affect size behaviour particularly. In contradistinction, the adjusted statistic AS is quite oversized in all cases, even when $n = 400$, while the S test seems a little undersized when the proximity between the competing models is higher ($\rho_{xx} = 0.4$). The AS statistic is also sensitive to the correlation between regressors across models and, in

Table 5.1: Monte Carlo estimated sizes (%) for a nominal size of 5% for non-nested hypothesis tests: design I (2000 replications)

n	ρ_{xug}	ρ_{xzg}	ρ_{xx}	S		AS		C		AC		E		J		SC/LC		PE/ME	
				gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	el(s)	el(r)	el(s)	el(r)
200	0.3	0.25	0.2	5.1	5.2	9.5	9.3	4.7	5.0	4.5	4.7	5.8	6.1	5.1	5.3	6.0	5.1	7.1	5.3
			0.4	3.8	3.7	8.2	7.8	4.5	5.1	4.4	4.9	5.6	6.0	5.1	5.2	6.0	5.4	7.1	5.5
		0.50	0.2	4.9	5.2	9.4	9.3	4.8	5.5	4.5	5.2	5.3	5.8	5.2	5.2	6.0	5.1	7.6	5.7
			0.4	3.9	4.3	7.7	7.8	4.6	5.3	4.5	5.0	5.6	5.8	5.4	5.2	6.0	5.3	7.8	5.8
	0.6	0.25	0.2	5.2	5.0	11.2	11.4	5.3	5.3	5.3	5.3	6.5	6.2	6.1	5.3	5.6	5.0	7.2	5.5
			0.4	4.2	4.0	10.1	10.4	5.3	5.4	5.0	5.3	6.4	5.8	6.4	5.2	5.9	5.0	7.5	5.6
		0.50	0.2	5.1	4.9	11.4	11.4	5.0	5.6	5.0	5.4	6.0	5.8	5.6	5.3	5.7	5.2	7.6	5.9
			0.4	4.0	4.2	10.4	10.6	5.8	5.5	5.2	5.1	5.6	6.0	5.8	5.6	6.2	5.3	7.6	5.9
400	0.3	0.25	0.2	4.5	4.7	10.7	10.7	4.4	4.8	4.4	4.9	4.4	4.6	4.7	4.7	5.3	4.6	5.3	4.2
			0.4	4.5	4.5	9.4	9.5	4.3	4.9	4.4	5.0	4.2	4.5	4.6	4.6	4.5	4.3	5.6	4.1
		0.50	0.2	4.7	4.5	10.4	10.5	4.1	4.4	4.5	4.4	4.3	4.4	4.7	4.8	5.2	4.8	5.3	4.3
			0.4	4.4	4.6	9.6	9.5	4.1	4.2	4.1	4.2	4.2	4.5	4.8	4.8	4.6	4.2	5.5	4.3
	0.6	0.25	0.2	4.6	4.5	8.8	8.7	4.9	4.7	4.7	5.0	4.5	4.5	5.0	4.8	5.5	5.0	5.3	4.1
			0.4	4.7	4.4	7.7	7.5	4.7	4.5	4.5	4.4	4.6	4.6	5.0	4.6	5.1	4.6	5.5	4.1
		0.50	0.2	4.5	4.5	8.7	8.8	4.5	4.5	4.3	4.6	4.5	4.4	4.8	4.8	5.5	5.2	5.3	4.3
			0.4	4.6	4.5	7.5	7.6	4.2	4.2	4.1	4.1	4.5	4.4	4.9	4.8	5.3	4.9	5.4	4.2

Note: the values underlined are significantly different from the nominal size at the 5% level (95% confidence interval limits: 4.045 and 5.955).

addition, to the correlation between u_g and X_g. Clearly, high feedbacks ($\rho_{xu_g} = 0.7$) from u_g to X_g have a very negative effect on its behaviour.

For the GEL group of tests, evaluated at EL estimators, there seems to be no major influence of the control variables, apart from the sample size n, on the behaviour of the tests. The most decisive influence appears *via* the estimation method for the variance matrix V_g. When the tests are based on $el(s)$, they are significantly oversized when $n = 200$, particularly the encompassing tests. If robust estimation of V_g is employed, $el(r)$, their size performance improves quite dramatically, now being fairly well approximated by the nominal size 0.05 which conforms to the results reported by Imbens, Spady and Johnson (1998) and in chapter 4. Preliminary experiments, not reported here, indicate that evaluation of GMM tests using $el(r)$ produces worse results than those obtained using $el(s)$, presumably because of the inverse manner in which V_g appears in these tests relative to GEL tests; see chapter 4. The beneficial effect of robust estimation of V_g for the PE/ME/LME tests is emphasized in Figure 5.1, which displays a QQ-plot for the case corresponding to the first row of Table 5.1. As it can be clearly seen, the robust forms of these tests are uniformly better.

Table 5.2 reports the empirical (size-corrected) powers of the above statistics which are the percentage of times the statistics exceeded the 0.05 nominal critical values obtained from their H_g empirical distribution. Two different values for ρ_{xu_q} (0.3 and 0.6) and ρ_{xz_q} (0.25 and 0.50) are simulated with different degrees of proximity between the competing models ($\rho_{xx} = 0.2$ and 0.4). The correlations ρ_{xu_g} and ρ_{xz_g} are both fixed at 0.5. Table 5.2 has an additional column indicating the value of $\rho_{z_g u_q}$ (0.1 and 0.2) in each experiment. As expected, for the reasons given above, the crucial determinant of power for all tests is the value of $\rho_{z_g u_q}$ with the other correlations having minor effects. Although the AS statistic appears most powerful, owing to the rather excessive sizes displayed in Table 5.1 it cannot be viewed as providing a reliable test of H_g against H_q. The performances of the J and $PE/ME/LME$ tests appear uniformly superior to the S, C, AC, E, and SC/LC tests. The Cox-type test AC performs only slightly better than the unadjusted C.

Figure 5.1: QQ-plots for PE/ME/LME non-nested hypothesis tests

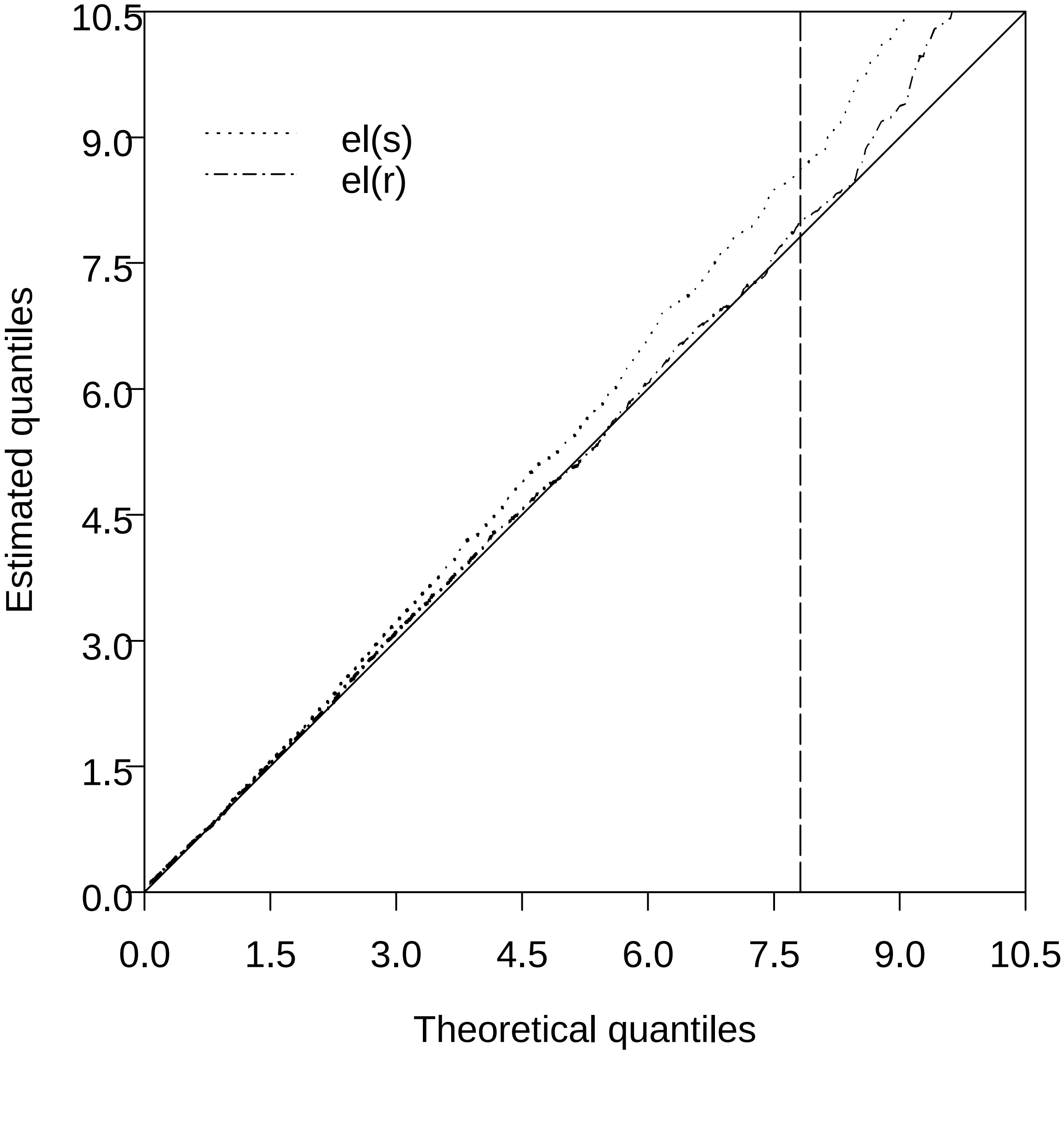


Table 5.2: Monte Carlo estimated (size-corrected) powers (%) for a nominal size of 5% for non-nested hypothesis: design I (2000 replications)

n	ρ_{xuq}	ρ_{xzq}	ρ_{xx}	ρ_{zguq}	S		AS		C		AC		E		J		SC/LC		PE/ME	
					gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	el(s)	el(r)	el(s)	el(r)
200	0.3	0.25	0.2	0.1	13.8	14.0	24.5	24.0	14.8	14.7	16.2	16.1	13.9	13.8	19.6	19.5	14.2	14.1	19.8	18.9
				0.2	35.8	36.5	73.7	75.4	39.4	36.6	40.4	39.5	36.0	35.9	69.4	69.9	35.3	33.4	63.9	58.6
			0.4	0.1	12.7	12.6	20.4	20.5	13.4	13.4	14.8	14.7	12.6	12.9	16.9	17.9	13.0	12.5	17.7	17.1
				0.2	31.6	31.2	66.7	66.4	35.1	33.1	36.6	36.2	32.4	31.5	61.4	62.8	31.4	28.9	56.5	51.5
		0.50	0.2	0.1	13.9	13.6	24.6	24.2	14.3	14.9	16.0	17.1	17.2	17.3	19.6	19.5	14.4	13.6	19.8	18.9
				0.2	35.5	35.6	73.7	75.5	38.7	35.9	40.6	40.6	36.7	36.1	69.4	69.9	37.3	34.9	63.9	58.6
			0.4	0.1	12.5	12.1	20.8	20.7	13.7	13.5	14.8	16.1	16.3	16.2	16.9	17.9	12.4	12.2	17.7	17.1
				0.2	31.1	31.1	66.3	66.2	35.1	33.0	36.4	36.1	33.6	32.9	61.4	62.8	32.2	30.2	56.5	51.5
	0.6	0.25	0.2	0.1	17.6	17.6	34.1	35.0	16.9	17.0	19.9	20.1	16.3	16.4	26.0	26.6	16.9	16.0	25.0	23.5
				0.2	43.5	44.5	89.0	88.6	44.7	43.9	50.9	48.2	43.7	43.8	84.3	84.9	42.2	38.9	78.0	71.6
			0.4	0.1	14.2	14.7	27.4	26.0	15.0	14.0	18.0	17.2	14.4	14.7	20.3	21.2	14.3	13.5	21.7	19.9
				0.2	35.4	36.4	76.7	77.7	39.5	36.4	43.3	42.8	37.7	36.3	70.3	71.0	36.3	34.1	65.6	59.3
		0.50	0.2	0.1	17.8	17.8	34.0	34.1	16.2	17.2	20.0	20.1	19.3	18.3	26.0	26.6	16.7	15.9	25.0	23.5
				0.2	43.7	44.1	88.7	88.8	44.3	44.3	50.9	50.3	43.2	41.2	84.3	84.9	43.6	40.6	78.0	71.6
			0.4	0.1	14.1	14.1	27.3	25.8	14.4	15.0	18.2	17.7	16.7	16.2	20.3	21.2	14.7	13.6	21.7	19.9
				0.2	36.2	35.9	77.2	78.3	39.0	37.1	44.5	45.0	36.8	35.1	70.3	71.0	37.4	36.1	65.6	59.3
400	0.3	0.25	0.2	0.1	26.8	26.8	44.0	43.6	26.2	25.3	29.1	28.0	29.1	28.2	42.2	42.5	23.9	22.7	42.1	40.9
				0.2	56.5	55.8	96.6	96.6	58.7	57.5	60.1	58.9	56.8	55.3	97.1	97.1	56.5	53.9	95.2	93.0
			0.4	0.1	24.6	24.7	36.4	37.5	23.4	23.3	26.3	25.8	27.8	25.5	37.7	37.8	21.1	20.8	37.4	36.8
				0.2	51.9	52.0	93.1	93.5	54.1	53.6	56.6	55.5	53.3	52.2	93.1	93.3	52.0	51.0	91.4	89.6
		0.50	0.2	0.1	26.8	26.1	44.7	44.1	26.9	24.5	29.1	28.6	29.5	27.9	42.2	42.5	23.6	23.1	42.1	40.9
				0.2	56.6	56.4	96.6	96.7	58.5	56.9	58.8	58.7	55.3	53.9	97.1	97.1	57.3	54.6	95.2	93.0
			0.4	0.1	23.7	24.2	37.8	37.4	22.4	22.2	26.3	26.5	27.7	26.5	37.7	37.8	20.6	20.5	37.4	36.8
				0.2	51.9	52.5	93.0	93.5	52.9	52.4	55.6	54.5	52.1	51.1	93.1	93.3	53.4	51.7	91.4	89.6
	0.6	0.25	0.2	0.1	32.8	32.7	60.4	59.8	33.3	32.8	35.0	34.9	33.1	31.7	55.4	55.3	31.6	31.0	53.7	51.8
				0.2	62.7	62.7	99.6	99.6	64.5	63.6	68.2	67.7	62.4	60.4	99.4	99.5	63.7	60.9	98.8	97.2
			0.4	0.1	27.4	27.2	47.1	47.5	26.8	26.7	30.1	29.8	29.0	27.9	44.1	44.8	25.0	24.7	44.1	42.8
				0.2	57.2	56.8	97.2	97.1	58.0	58.1	62.7	60.7	57.7	56.6	97.0	97.0	57.1	54.7	95.6	93.4
		0.50	0.2	0.1	32.6	32.5	57.9	59.2	32.7	32.4	35.4	36.0	32.4	33.0	55.4	55.3	31.8	30.7	53.7	51.8
				0.2	62.7	62.6	99.4	99.5	63.6	63.9	68.5	67.9	61.3	62.2	99.4	99.5	64.0	60.9	98.8	97.2
			0.4	0.1	27.8	27.5	47.1	46.5	26.3	25.8	30.2	29.7	29.2	28.7	44.1	44.8	25.3	25.0	44.1	42.8
				0.2	57.4	57.2	97.3	97.2	58.5	57.2	61.9	61.2	56.2	56.8	97.0	97.0	56.9	54.5	95.6	93.4

5.6.3 Monte Carlo experiment II

In these second set of experiments the number of regressors comprising X_q was increased to $k_q = 2$ and all the instruments of the null model H_g are now also valid under the alternative hypothesis H_q, that is, $\rho_{z_g u_q} = 0$. Hence, a difference in regression functions is now the main possible source of misspecification.

The regressors of both models were generated in a similar way to Design I. Thus, X_g and X_q are given by

$$X_g = \epsilon + \tau\mu_1 + \lambda u_g, \tag{5.57}$$

$$X_{qj} = \mu_j + \psi u_q,\ j = 1, 2, \tag{5.58}$$

where ϵ, μ_1, μ_2, u_g and u_q are independent $N(0, I_n)$ vectors. The H_g matrix of IVs Z_g, $s_g = 4$, are generated as

$$Z_{gj} = \varphi_1\epsilon + \mu_1 + \mu_2 + \varsigma_j,\ j = 1, 2, \tag{5.59}$$

$$Z_{gj} = \varphi_2\epsilon + \mu_1 + \varsigma_j,\ j = 3, 4, \tag{5.60}$$

and the H_q matrix of IVs Z_q, $s_q = 4$, as

$$Z_{qj} = \gamma\mu_1 + \omega_j,\ j = 1, 2, \tag{5.61}$$

$$Z_{qj} = \gamma\mu_2 + \omega_j,\ j = 3, 4, \tag{5.62}$$

where ς_j and ω_j, $j = 1, ..., 4$, are independent $N(0, I_n)$ vectors generated independently of ϵ, μ_1, μ_2, u_g and u_q.

The parameters λ, τ, ψ, φ_1, φ_2 and γ control correlations in a similar manner to the previous sub-section with the above formulas still appropriate with the exception of those concerning the correlation between the regressor and the instruments in the H_g model, which are now given by

$$\rho_{xz_{gj}} = (\varphi_1 + \tau)\left[\left(3 + \varphi_1^2\right)\left(1 + \tau^2 + \lambda^2\right)\right]^{-1/2}, j = 1, 2 \tag{5.63}$$

and

$$\rho_{xz_{gj}} = (\varphi_2 + \tau)\left[\left(2+\varphi_2^2\right)\left(1+\tau^2+\lambda^2\right)\right]^{-1/2}, \; j = 3, 4. \tag{5.64}$$

Table 5.3 reports the empirical sizes obtained in this second set of experiments. We considered the same control variables of the first experimental design with ranges: ρ_{xu_g} (0.3 and 0.6), ρ_{xz_g} (0.3 and 0.6) and ρ_{xx} (0.25 and 0.5). We also fixed $\rho_{xu_q} = 0.5 = \rho_{xz_q} = 0.5$. As can be seen from Table 5.3, most of the conclusions achieved in the previous sub-section are still valid. Thus, the superior performance of the robust forms of the GEL-based tests relative to their standard versions is once again clear. The over-rejection of the AS test is now even more apparent as well as its sensitiveness to the correlation between X_g and u_g. The poor performance of the E test when that correlation is higher becomes evident as well as the influence of ρ_{xz_g}. The C and $el\,(s)\;J$ tests continue to exhibit empirical sizes close to the nominal ones. In contradiction, the size characteristics of AC deteriorated significantly over those in Table 5.3. Like the E test, the size behaviour of the AC test is worse for higher values of ρ_{xu_g} and improves with larger correlations between the regressor X_g and the instruments Z_g.

Table 5.4 reports (size-corrected) powers. As before, we fixed $\rho_{xu_g} = \rho_{xz_g} = 0.5$. The power of all tests seem to be negatively related to estimation accuracy under H_q, ρ_{xu_q}, and to model proximity, ρ_{xx}. The most important difference relative to the previous study is that the power of the AC test is now more than twice that of the unadjusted version C in almost all cases and is the most powerful test in this second set of experiments. Unfortunately, as seen in Table 5.3, this good power performance came at the expense of a poor size behaviour. The AS, E, J and $PE/ME/LME$ tests also perform well, with rather moderate power for the S, C and SC/LC tests.

Table 5.3: Monte Carlo estimated sizes (%) for a nominal size of 5% for non-nested hypothesis tests: design II (2000 replications)

n	ρ_{xug}	ρ_{xzg}	ρ_{xx}	S		AS		C		AC		E		J		SC/LC		PE/ME	
				gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	el(s)	el(r)	el(s)	el(r)
200	0.3	0.3	0.25	5.0	5.3	14.5	14.3	4.7	5.1	8.0	8.0	4.8	5.5	4.8	4.7	6.1	5.2	7.4	5.6
			0.50	5.3	5.5	14.0	13.6	4.7	5.7	7.7	7.6	5.2	6.1	4.6	4.5	6.0	5.3	7.2	5.4
		0.6	0.25	4.9	5.5	13.9	13.8	4.9	5.6	6.7	7.3	4.4	5.6	4.6	4.7	5.9	5.0	7.8	6.1
			0.50	5.3	5.5	13.4	13.2	4.5	5.4	6.5	7.1	5.0	5.5	4.6	4.8	6.1	5.1	8.2	6.2
	0.6	0.3	0.25	5.3	5.3	16.1	16.6	5.1	5.1	10.1	7.2	7.9	8.1	5.8	4.7	6.1	5.5	7.7	5.7
			0.50	5.3	5.4	16.2	16.2	4.8	5.4	9.1	6.7	8.8	9.5	5.9	4.6	6.3	5.4	7.6	5.6
		0.6	0.25	5.2	5.4	15.8	16.4	4.7	5.6	7.0	6.6	5.5	6.4	4.8	4.7	5.7	5.1	8.1	6.2
			0.50	5.7	5.6	15.7	16.2	4.5	5.6	6.9	6.9	6.7	7.3	5.0	5.1	6.3	5.4	8.3	6.4
400	0.3	0.3	0.25	4.2	4.1	12.4	12.2	4.0	4.4	6.9	6.5	5.0	5.9	4.8	4.6	5.3	4.9	4.9	3.9
			0.50	4.3	4.0	11.1	10.9	4.5	4.6	6.4	6.6	5.1	5.8	4.5	4.5	5.3	4.9	4.8	3.9
		0.6	0.25	4.2	4.1	12.3	12.0	4.3	4.3	6.1	6.1	4.7	5.8	4.4	4.5	5.4	4.9	5.3	4.2
			0.50	4.6	4.0	11.4	11.3	4.5	4.4	6.1	6.4	4.6	5.0	4.3	4.5	5.3	4.7	5.4	4.2
	0.6	0.3	0.25	4.3	4.3	14.2	14.0	4.4	4.3	7.9	6.1	7.1	7.3	5.2	4.6	5.6	5.0	5.0	4.1
			0.50	4.5	3.7	13.2	13.1	4.6	4.4	7.6	6.1	7.9	8.4	4.9	4.5	5.3	4.7	5.0	4.0
		0.6	0.25	4.1	4.0	14.5	14.3	4.2	4.3	6.2	5.9	5.4	6.1	4.5	4.5	5.6	5.2	5.2	4.2
			0.50	4.3	4.1	13.8	13.4	4.2	4.2	6.7	6.2	5.8	6.9	4.5	4.5	5.2	5.0	5.5	4.6

Note: the values underlined are significantly different from the nominal size at the 5% level (95% confidence interval limits: 4.045 and 5.955).

Table 5.4: Monte Carlo estimated (size-corrected) powers (%) for a nominal size of 5% for non-nested hypothesis tests: design II (2000 replications)

n	ρ_{xuq}	ρ_{xzq}	ρ_{xx}	S		AS		C		AC		E		J		SC/LC		PE/ME	
				gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	gmm	el(s)	el(s)	el(r)	el(s)	el(r)
200	0.3	0.3	0.25	40.0	37.5	68.5	67.5	31.2	31.5	74.6	71.0	70.5	62.8	68.2	67.9	33.8	32.1	58.2	52.9
			0.50	38.3	36.3	63.0	59.9	29.0	29.8	70.9	65.4	67.7	65.0	64.0	63.1	32.4	30.4	54.2	48.8
		0.6	0.25	39.8	37.0	68.9	67.7	32.4	37.1	73.7	73.2	70.4	68.5	68.2	67.9	34.2	32.4	58.2	52.9
			0.50	37.7	35.3	62.2	59.2	30.9	33.3	69.1	67.4	69.9	64.9	64.0	63.1	33.6	31.6	54.2	48.8
	0.6	0.3	0.25	25.2	24.2	56.1	54.9	23.4	22.0	54.3	53.8	45.9	43.9	41.3	41.5	23.0	21.4	36.6	35.2
			0.50	22.4	21.7	35.6	34.4	22.9	19.5	40.4	38.5	38.6	36.1	34.5	35.2	19.8	18.6	30.9	28.3
		0.6	0.25	25.1	18.6	56.6	56.9	23.2	25.4	51.1	50.2	46.7	44.7	41.3	41.5	23.8	22.8	36.6	35.2
			0.50	22.0	15.0	35.5	35.9	23.1	23.3	39.2	39.8	37.2	33.5	34.5	35.2	19.6	19.0	30.9	28.3
400	0.3	0.3	0.25	58.4	58.1	94.2	94.2	44.7	49.0	95.5	95.1	95.4	78.6	93.7	93.7	52.3	50.4	90.9	89.0
			0.50	56.2	55.8	91.1	91.3	42.2	47.8	93.1	92.5	94.1	77.9	91.4	91.2	50.8	48.9	88.7	86.5
		0.6	0.25	58.3	57.4	94.3	94.2	45.1	57.2	96.4	95.9	94.8	35.1	93.7	93.7	51.9	49.8	90.9	89.0
			0.50	56.6	55.1	91.3	91.2	42.5	54.2	93.7	94.4	94.5	59.5	91.4	91.2	51.4	49.0	88.7	86.5
	0.6	0.3	0.25	43.7	44.0	84.8	84.9	41.4	35.0	82.1	83.1	76.3	34.8	74.4	74.3	37.9	36.7	69.4	68.2
			0.50	40.3	39.8	64.5	64.0	34.2	31.9	68.4	69.2	69.5	35.0	63.7	64.0	32.3	31.2	59.1	57.7
		0.6	0.25	43.7	36.1	84.3	85.6	40.4	41.8	78.7	79.4	71.3	7.5	74.4	74.3	38.3	37.2	69.4	68.2
			0.50	40.2	34.5	66.1	67.5	35.1	39.0	68.9	70.1	68.5	20.9	63.7	64.0	32.8	32.0	59.1	57.7

5.7 Conclusion

This chapter has proposed a number of new non-nested test statistics for evaluating competing models specified solely in terms of moment conditions. A simple modification of Smith's (1992) GMM-based Cox-type test, which is similar in construction to that of Singleton (1985), is given, which appears to lead, in some cases, to significantly improved power properties but with concomitant poorer size characteristics. Furthermore, and this is the main contribution of this chapter, a parametric and a class of moment encompassing tests are also suggested within the GEL framework. Simulation experiments for competing linear instrumental variable models indicate that GEL-based encompassing tests using a robust estimator for the variance matrix of the moment indicators are particularly efficacious.

The generalized moment encompassing tests can be implemented in a number of different ways, accordingly to the statistic chosen to represent a specific characteristic of the rival model. Here, we compared two of the possible choices for that statistic, one based on the moment indicators, which give rise to the moment encompassing test, and the other based on the objective function, which produces Smith's (1997) SC and LC tests. Naturally, different choices would produce tests with different finite sample properties, so a future avenue for research is to explore the possibility of constructing tests with better small sample properties by changing the feature of the competing model that is contrasted.

Chapter 6

Conclusion

6.1 Main findings

The study of GMM and alternative estimation methods for moment condition models is, nowadays, one of the most popular research topics in theoretical econometrics. In this thesis we focused mainly on the analysis of GMM and GEL estimators and related statistics, achieving three major contributions to this subject.

First, through the realization of two extensive Monte Carlo simulation studies, we examined the small sample bias of two classes of alternative estimators that are theoretically appropriate for estimating models defined solely in terms of moment conditions. The first class includes the first-order asymptotically equivalent GMM, CU-GMM and GEL estimators, while the second contains six distinct bootstrap GMM estimators, three of which were developed in this thesis. The three bootstrap techniques that we propose use the GEL implied probabilities to construct the bootstrap samples, which are, thus, generated in a more efficient way than in the three bootstrap methods previously suggested by other authors. Our simulation results, involving covariance structure and instrumental variable models, popular applications of GMM, show clearly that there are much better methods to estimate moment condition models than conventional GMM estimation. Indeed, this estimation method produced the worst results in almost all cases. In contrast, the PHGEL bootstrap behaved in a very promising manner, being the method with less mean bias in most cases. The RGEL bootstrap, also derived in this thesis, despite behaving more modestly in the second set of experiments, produced also better results than the remaining bootstrap methods. Considering only the non-bootstrap methods, the EL estimator had the best performance. For this class of estimators, we found that Newey and Smith's (2000) results seem to be a good guide for their small sample behaviour.

Our second major investigation concerned the development of Pearson-type test statistics suitable for testing both overidentifying moment conditions and parametric restrictions in models estimated by GEL methods. We derived two classes of Pearson-type statistics, both based on the comparison of two consistent estimators, under the corresponding null hypothesis in assessment, of the unknown distribution of the data. The first class includes tests that are very similar in form to the classical Pearson χ^2 statistics. The other requires the partition of the sample space in several sets, the contrast between the empirical and the GEL implied probabilities (or two GEL implied probabilities) estimated for each set forming the basis for the test. The two Monte Carlo simulation studies realized, concerning tests of overidentifying moment conditions, revealed a very promising performance of the latter Pearson-type statistic relative to both bootstrap versions of the J test and alternative tests. The best results were obtained when robust estimation of the variance matrix of the moment indicators was employed.

A number of new non-nested hypothesis tests that integrate and complement the work of other authors constitute our last major contribution to the econometrics of moment condition models. On the one hand, we derived generalized statistics that include most of the existing tests as particular cases. On the other hand, we developed GEL parametric and moment encompassing tests that enlarge substantially the number of tests available to the practitioner to assess non-nested moment condition models. One of our suggestions provides a simple method of implementing Ghysels and Hall's (1990) idea for constructing a moment-based test in the GMM framework, without requiring the introduction of auxiliary assumptions in addition to those given by the moment conditions. Simulation experiments for competing non-nested linear instrumental variable models indicate that GEL-based encompassing tests using a robust estimator for the variance matrix of the moment indicators are particularly efficacious.

6.2 Future Research

The findings from this thesis provide some avenues for future research in the econometric analysis of moment condition models. This is especially true for the investigation undertaken in chapter 3, a natural extension of it being the examination of the ability of bootstrap methods to eliminate the finite sample bias of CU-GMM and GEL estimators. Although theoretically simple, such extension will require a great amount of computer time and power. A more interesting topic of investigation is perhaps the analysis of alternative methods for obtaining bias-corrected GMM and GEL estimators. We are already investigating, in a joint paper with R. J. Smith and A. D. Chesher, the small sample properties of such corrected estimators when based on Newey and Smith's (2000) asymptotic bias functions. We consider two approaches. One uses those expressions evaluated at the corresponding estimator to obtain an estimate of its bias; by directly subtracting this estimate from the standard estimator we are able to calculate a bias-corrected estimator. The second approach, based on the work of Firth (1993), utilizes Newey and Smith's (2000) expressions to correct the first-order conditions defining the estimator, which can be or not previously evaluated at it; solving these modified first-order conditions other bias-corrected estimators are obtained.

All major contributions of this thesis are based in some way on the utilization of the GEL implied probabilities to estimate some features of the data. In fact, we showed how to employ them to construct the three new bootstrap techniques, the Pearson-type statistics and the parametric and moment encompassing non-nested tests. However, more applications of these probabilities are certainly possible. For example, in a parametric context, it should be relatively straightforward to assess distributional assumptions using a GEL Kolmogorov-Smirnov-type statistic based on the comparison of the GEL and the assumed cumulative distribution functions.

Finally, the extension of all methods and statistics concerning GEL estimation for a time-series framework is also an interesting and important avenue for future research. Indeed, only Kitamura and Stutzer (1997) and Smith (1997, 2001) have dealt with this issue, proposing the smoothing of the observations before the optimization. However, the performance in practice of such GEL estimators remains to be examined.

Bibliography

Abowd, J. M. and Card, D. (1987), "Intertemporal labor supply and long-term employment contracts", *American Economic Review*, 77, pp. 50-68.

Abowd, J. M. and Card, D. (1989), "On the covariance structure of earnings and hours changes", *Econometrica*, 57, pp. 411-445.

Altonji, J. G. and Segal, L. M. (1996), "Small-sample bias in GMM estimation of covariance structures", *Journal of Business & Economic Statistics*, 14(3), pp. 353-365.

Amemiya, T. (1974), "The nonlinear two-stage least-squares estimator", *Journal of Econometrics*, 2, pp. 105-110.

Amemiya, T. (1977), "The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model", *Econometrica*, 45(4), pp. 955-968.

Andersen, T. G. and Sorensen, B. E. (1996), "GMM estimation of a stochastic volatility model: a Monte Carlo study", *Journal of Business & Economic Statistics*, 14(3), pp. 328-352.

Andrews, D. W. K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation", *Econometrica*, 59(3), pp. 817-858.

Back, K. and Brown, D. P. (1993), "Implied probabilities in GMM estimators", *Econometrica*, 61(4), pp. 971-975.

Behrman, J., Rosenzweig, M. and Taubman, P. (1994), "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment", *Journal of Political Economy*, 102, pp. 1131-1174.

Bera, A. K. and Bilias, Y. (2000), "MM, ME, ML, EL, EF and GMM approaches to estimation: a synthesis", presented at the EC2 meeting, Dublin.

Blomquist, S. and Dahlberg, M. (1999), "Small sample properties of LIML and Jackknife IV estimators: experiments with weak instruments", *Journal of Applied Econometrics*, 14, pp. 69-88.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", *Journal of the American Statistical Association*, 90(430), pp. 443-450.

Brown, B. W., Newey, W. K. and May, S. (1997), "Efficient bootstrapping for GMM", mimeo.

Carroll, R. J. (1982), "Adapting for heteroscedasticity in linear models", *Annals of Statistics*, 10(4), pp. 1224-1233.

Chamberlain, G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions", *Journal of Econometrics*, 34, pp. 305-334.

Chernick, M. R. (1999), *Bootstrap methods: a practitioner's guide*, Wiley.

Chesher, A. and Smith, R. J. (1997), "Likelihood ratio specification tests", *Econometrica*, 65(3), pp. 627-646.

Corcoran, S. A. (1998), "Bartlett adjustment of empirical discrepancy statistics", *Biometrika*, 85(4), pp. 967-.

Cox, D. R. (1961), "Tests of separate families of hypotheses", in Neyman, J. (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pp. 105-123.

Cox, D. R. (1962), "Further results on tests of separate families of hypotheses", *Journal of the Royal Statistical Association, Series B*, 24(2), pp. 406-424.

Cressie, N. and Read, T. R. C. (1984), "Multinomial goodness-of-fit tests", *Journal of the Royal Statistical Society, Series B*, 46(3), pp. 440-464.

Davidson, R. and MacKinnon, J. G. (1993), *Estimation and inference in econometrics*, Oxford University Press.

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap methods and their application*, Cambridge University Press.

Den Haan, W. J. and Levin, A. T. (1997), "A practitioner's guide to robust covariance matrix estimation", in Maddala, G. S. and Rao, C. R. (eds.), *Handbook of Statistics, Vol. 15*, Elsevier Science Publishers.

Donald, S. G. and Newey, W. K. (2000), "A Jackknife interpretation of the continuous updating estimator", *Economics Letters*, 67, pp. 239-243.

Efron, B. (1979), "Bootstap methods: another look at the Jackknife", *Annals of Statistics*, 7(1), pp. 1-26.

Efron, B. (1990), "More efficient bootstrap computations", *Journal of the American Statistical Association*, 85(409), pp. 79-89.

Efron, B. and Tibshirani, R. J. (1993), *An introduction to the bootstrap*, Chapman & Hall.

Eichenbaum, M. S., Hansen, L. P. and Singleton, K. J. (1988), "A time series analysis of representative agent models of consumption and leisure choice under uncertainty", *Quarterly Journal of Economics*, 103, pp. 51-78.

Ferson, W. E. and Foerster, S. R. (1994), "Finite sample properties of the generalised method of moments in tests of conditional asset pricing models", *Journal of Financial Economics*, 36(1), pp. 29-55.

Firth, D. (1993), "Bias reduction of maximum likelihood estimators", *Biometrika*, 80(1), pp. 27-38.

Ghysels, E. and Hall, A. (1990a), "A test for structural stability of Euler conditions parameters estimated via the generalised method of moments estimator", *International Economic Review*, 31(2), pp. 355-364.

Ghysels, E. and Hall, A. (1990b): "Testing non-nested Euler conditions with quadrature-based methods of approximation", *Journal of Econometrics*, 46, 273-308.

Griliches, Z. (1979), "Sibling models and data in economics: beginnings of a survey", *Journal of Political Economy*, 87, pp. S37-S64.

Hahn, J. (1996), "A note on bootstrapping generalised method of moments estimators", *Econometric Theory*, 12(1), pp. 187-197.

Hall, A. (1993), "Some aspects of generalized method of moments estimation", in Maddala, G. S., Rao, C. R. and Vinod, H. D. (eds.), *Handbook of Statistics, Vol. 11*, Elsevier Science Publishers, pp. 393-417.

Hall, P. (1992), *The bootstrap and Edgeworth expansion*, Springer-Verlag.

Hall, P. and Horowitz, J. L. (1996), "Bootstrap critical values for tests based on generalised-method-of-moments estimators", *Econometrica*, 64(4), pp. 891-916.

Hall, R. E. and Mishkin, F. (1982), "The sensitivity of consumption to transitive income: estimates from panel data on households", *Econometrica*, 50, pp. 461-481.

Hansen, L. P. (1982), "Large sample properties of generalised method of moments estimators", *Econometrica*, 50(4), pp. 1029-1054.

Hansen, L. P. (1985), "A method for calculating bounds on the asymptotic covariance matrices of generalised method of moments estimators", *Journal of Econometrics*, 30, pp. 203-238.

Hansen, L. P., Heaton, J. C. and Ogaki, M. (1988), "Efficiency bounds implied by multiperiod conditional moment restrictions", *Journal of the American Statistical Association*, 83(403), pp. 863-871.

Hansen, L. P., Heaton, J. and Yaron, A. (1996), "Finite-sample properties of some alternative GMM estimators", *Journal of Business & Economic Statistics*, 14(3), pp. 262-280.

Heaton, J. C. and Ogaki, M. (1991), "Efficiency bound calculations for a time series model, with conditional heteroskedasticity", *Economics Letters*, 35, pp. 167-171.

Horowitz, J. L. (1998), "Bootstrap methods for covariance structures", *The Journal of Human Resources*, 33(1), pp. 39-61.

Imbens, G. W. (1997), "One-step estimators for over-identified generalised method of moments models", *Review of Economic Studies*, 64, pp. 359-383.

Imbens, G. W., Spady, R. H. and Johnson, P. (1998), "Information theoretic approaches to inference in moment condition models", *Econometrica*, 66(2), pp. 333-357.

Jorgenson, D. W. and Laffont, J. J. (1974), "Efficient estimation of nonlinear simultaneous equations with additive disturbances", *Annals of Economic and Social Measurement*, 3/4, pp. 615-640.

Kitamura, Y. and Stutzer, M. (1997), "An information-theoretic alternative to generalised method of moments estimation", *Econometrica*, 65(4), pp. 861-874.

Kocherlakota, N. R. (1990), "On tests of representative consumer asset pricing models", *Journal of Monetary Economics*, 26, pp. 285-304.

Mariano, R. S. (1982), "Analytical small-sample distribution theory in econometrics: the simultaneous-equations case", *International Economic Review*, 23(3), pp. 503-533.

Mátyás, L. (1999), *Generalized method of moments estimation*, Cambridge University Press.

Mittelhammer, R. C., Judge, G. G. and Miller, D. J. (2000), *Econometric foundations*, Cambridge University Press.

Mizon, G. E. and Richard, J.-F. (1986), "The encompassing principle and its applications to testing non-nested hypotheses", *Econometrica*, 54(3), pp. 657-678.

Nelson, C. R. and Startz, R. (1990), "The distribution of the instrumental variables estimator and its t-statistic when the instrument is a poor one", *Journal of Business*, 63(1), pp. S125-S140.

Newey, W. K. (1985a), "Generalized method of moments specification testing", *Journal of Econometrics*, 29, pp. 226-256.

Newey, W. K. (1985b), "Maximum likelihood specification testing and conditional moment tests", *Econometrica*, 53(5), pp. 1047-1070.

Newey, W. K. (1993), "Efficient estimation of models with conditional moment restrictions", in Maddala, G. S., Rao, C. R. and Vinod, H. D. (eds.), *Handbook of Statistics, Vol. 11*, Elsevier Science Publishers, pp. 419-454.

Newey, W. K. and McFadden, D. (1994), "Large sample estimation and hypothesis testing", in Engle, R. F. and McFadden, D. L. (eds.), *Handbook of Econometrics, Vol. 4*, Elsevier Science Publishers, pp. 2111-2245.

Newey, W. K. and Smith, R. J. (2000), "Asymptotic bias and equivalence of GMM and GEL estimators", mimeo, University of Bristol.

Newey, W. K. and Smith, R. J. (2001), "Higher order properties of GMM and Generalized Empirical Likelihood estimators", mimeo, University of Bristol.

Newey, W. K. and West, K. D. (1987a): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55(3), pp. 703-708.

Newey, W. K. and West, K. D. (1987b), "Hypothesis testing with efficient method of moments estimation", International Economic Review, 28, pp. 777-787.

Newey, W. K. and West, K. D. (1994), "Automatic lag selection in covariance matrix estimation", *Review of Economic Studies*, 61, pp. 631-653.

Neyman, J. (1959), "Optimal asymptotic tests of composite statistical hypothesis", in Grenander, U. (ed.), *Probability and Statistics*, Wiley, pp. 213-234.

Ogaki, M. (1993), "Generalized method of moments: econometric applications", in Maddala, G. S., Rao, C. R. and Vinod, H. D. (eds.), *Handbook of Statistics, Vol. 11*, Elsevier Science Publishers, pp. 455-488.

Owen, A. B. (1988): "Empirical likelihood ratio confidence intervals for a single functional", Biometrika, 75(2), pp. 237-249.

Owen, A. B. (1990), "Empirical likelihood ratio confidence regions", *Annals of Statistics*, 18(1), pp. 90-120.

Owen, A. B. (1991), "Empirical likelihood for linear models", *Annals of Statistics*, 19(4), pp. 1725-1747.

Owen, A. B. (2001), *Empirical likelihood*, Chapman & Hall / CRC.

Qin, J. and Lawless, J. (1994), "Empirical likelihood and general estimating equations", *Annals of Statistics*, 22(1), pp. 300-325.

Qin, J. and Lawless, J. (1995), "Estimating equations, empirical likelihood and constraints on parameters", *Canadian Journal of Statistics*, 23, pp. 145-159.

Rao, C. R. and Mitra, S. K. (1971), *Generalized inverse of matrices and its applications*, Wiley.

Robinson, P. M. (1987), "Asymptotically efficient estimation in the presence of heterocedasticity of unknown form", *Econometrica*, 55(4), pp. 875-891.

Shao, J. and Tu, D. (1995), *The Jackknife and bootstrap*, Springer-Verlag.

Singleton, K. J. (1985): "Testing specifications of economic agents' intertemporal optimum problems in the presence of alternative models", *Journal of Econometrics*, 30, pp. 391-413.

Smith, R. J. (1987), "Alternative asymptotically optimal tests and their application to dynamic specification", *Review of Economic Studies*, 54, pp. 665-680.

Smith, R. J. (1992): "Non-nested tests for competing models estimated by generalised method of moments", *Econometrica*, 60(4), pp. 973-980.

Smith, R. J. (1997), "Alternative semi-parametric likelihood approaches to generalised method of moments estimation", *Economic Journal*, 107(441), pp. 503-519.

Smith, R. J. (2000), "Empirical likelihood estimation and inference", in Salmon, M. and Marriot, P. (eds.), *Applications of Differential Geometry to Econometrics*, Cambridge University Press.

Smith, R. J. (2001), "GEL criteria for moment condition models", mimeo, University of Bristol.

Stock, J. H. and Wright, J. H. (2000), "GMM with weak identification", *Econometrica*, 68(5), pp. 1055-1096.

Szroeter, J. (1983), "Generalized Wald methods for testing nonlinear implicit and overidentifying restrictions", Econometrica, 51(2), pp. 335-353.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics*, 30, pp. 415-443.

Tauchen, G. (1986a), "A note on the asymptotic lower bound for the covariance matrix of the GMM estimator of the parameters of agents' utility functions", *Economics Letters*, 20, pp. 151-155.

Tauchen, G. (1986b), "Statistical properties of generalised method of moments estimators of structural parameters obtained from financial market data", *Journal of Business & Economic Statistics*, 4(4), pp. 397-425.

Ziliak, J. P. (1997), "Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators", *Journal of Business & Economic Statistics*, 15(4), pp. 419-431.