Econometric Applications of Empirical Likelihood

Flor Angelica Gonzalez Solano

Thesis submitted for the PhD degree in Economics

University of York

Department of Economics

June 2005

Abstract

We provide some evidence of Empirical Likelihood's (EL) practical value in econometrics. We present EL as an alternative to GMM estimation and assess the finite-sample properties of their overidentification tests (size and power) through Monte Carlo simulations. We address the issue of the importance of the results to applied workers and use as laboratories to our experiments two settings with potential empirical applications: the Mean-Variance and Three-Moment CAPM and a dynamic panel model with individual effects. In cases in which we found important size distortions we introduced efficient bootstrap critical values. Prior research applied this bootstrapping technique to the GMM (GMM-bootstrap) and we present results for the EL (EL-bootstrap). We also include an empirical example on a United States panel cash-flow model. Even if our findings do not uniformly support the conclusion that one estimator dominates the other, we found evidence that EL and EL-bootstrap are good alternatives to GMM and GMM-bootstrap in some econometric applications.

Contents

List of Figures

List of Tables

Acknowledgments

Above all, I want to thank the Lord for having filled my reality with opportunities. I also want to thank my family, from the bottom of my heart, for their full support. To Eneas Caldino, who showed me how to love Econometrics. To Karim Abadir, who diligently read and commented on Chapter 3. To Manuel Arellano, whose recommendations and suggestions on earlier versions of Chapter 4 were extremely useful. To Leslie Godfrey, Joel Horowitz and Peter Phillips for many helpful comments. I would also want to express my gratitude to Catherine Heinemeyer, Peter Moles, Andrew Street and Duncan Whitehead for correcting my "Spanglish". To the National Council of Science and Technology in Mexico for their financial support. I am very grateful to the friends who helped me. Finally, a special thanks is owed to my supervisors, Patrick Marsh and Francesco Bravo, for their guidance.

Introduction

Econometricians seem to share a general desire to avoid making strong distributional assumptions about the stochastic characteristics of probability models when *a priori* knowledge is insufficient to support such assumptions. When there is insufficient information to specify the functional form of a parametric likelihood function, an Empirical Likelihood (EL) function is still available under appropriate sample conditions. This is based on a multinomial probability distribution function supported on the sample of data. The resulting EL estimator of model parameters has first-order asymptotic properties that are in many ways analogous to parametric methods. Furthermore, the EL function behaves much the same as an ordinary likelihood function in terms of the usual likelihood ratio statistics that can be used for inference purposes.

In this dissertation we shall focus on studying Owen's (1991, 1990, 1988) EL as an alternative to the General Method of Moments (GMM). Both methods share important characteristics; for example, the Empirical Likelihood Ratio (ELR) possesses an asymptotic variance that is the same as that for the efficient GMM, thus it is asymptotically efficient. Both tests are distribution-free and their general setting are moment-condition models. The EL overidentification test is similar to that of the GMM; they are asymptotically first-order equivalent and have the same interpretation.

The size (level), power and other properties of hypothesis tests, as well as the confidence level and other properties of confidence sets based on asymptotic theory are all approximate, for they are only valid asymptotically. A logical question to consider is "How accurate are the approximations?", the answer to which is directly related to how accurate the hypothesis tests and confidence region procedures are that are derived from the approximations. One might also ask, "How large does the sample size have to be to obtain an accurate approximation?". The preceding questions are of the utmost importance if one wishes to utilize asymptotic theory as a guide to finite-sample behaviour. Through Monte Carlo experiments we shall provide an indication of how useful asymptotic approximations are for a range of different problems. We will mostly focus on the finite-sample properties of tests of overidentifying restrictions based on EL and different versions of GMM.

Why study the EL as an alternative to GMM? It has been extensively documented that the asymptotic approximation for GMM confidence intervals and tests can be poor (see, among others, the 1996 special issue of the *Journal of Business and Economic Statistics* on GMM). For example, the asymptotic properties of the GMM test of overidentifying restrictions can be a poor guide to finite-sample behaviour in small data sets often encountered in empirical analyses. Therefore, it is important either to explore new procedures or to improve on the existing ones.

Several problems were encountered while writing this dissertation. The first problem is that solutions to EL problems cannot be written in closed form and must be numerically computed. This is necessarily computationally intensive and can lead to cumbersome calculations. Having to design new computer programs was only the beginning of a series of complications. Another problem that is worth noting here is that there are both advantages and disadvantages to using Monte Carlo experimentations as an approach to studying finite-sample behaviour when it is compared with a mathematical analysis of the statistic in question. In respect to the drawbacks of this approach, the results of simulations are imprecise and specific (to an unknown extent) to the particular parameter values being investigated and the distributional assumptions made. Therefore, to reduce imprecision a high number of replications for the particular question being addressed had to be carried out. To deal with specificity each experiment was repeated several times using a range of values for the sample size, number of time periods (in a panel data context) and parameter values. The latter was extremely time consuming and a long processing time was required for each model.

Prior research has investigated the performance of the EL overidentification test. This is valuable in the sense that contributions, similarities and limitations of the existing literature were identified and used as a baseline in the design of our research. The following facts are noteworthy:

1. Most studies are invariably based on Qin and Lawless (1994), Hall and Horowitz (1996) and chi-squared moments models. These studies make the same distributional

assumptions.

2. Most studies assess the finite-sample size properties of overidentification tests but not its power.

3. There is no simulation evidence for overidentification tests where EL is combined with other methods, *e.g.* the bootstrap[1].

4. There are relatively few empirical applications, as compared to the GMM, for the EL.

The dissertation is laid out as follows: Chapter 1 is devoted to the necessary background to follow most of the results provided in this work. Chapter 2 analyses some numerical properties of EL. First, we examine the computational aspects of EL. Then, we assess the adequacy of the asymptotic approximations of its estimators and test statistics. The laboratory settings of our experiments are the invariably exploited models of Qin and Lawless (1994), Hall and Horowitz (1996) and a chi-squared moments model. However, extensions (alternative distributions, parameter values and sample sizes) and contributions (a non-uniform bootstrap procedure based on EL probabilities and applied to EL) to existing literature are also discussed. Chapter 3 then examines the question of size and power of overidentification tests based on a financial framework: the Capital Asset Pricing Model (CAPM). Two versions of the CAPM are studied: Mean-Variance and Three-Moment CAPM. As far as we are aware this is the first time that the Three-Moment version is used as a laboratory setting to assess the finite-sample properties of overidentification tests. Chapter 4 extends EL to a dynamic

[1]Namba (2004) combines EL and the bootstrap to obtain critical values from the ELR for the case of estimating a population mean.

panel data context. First, simulation evidence is provided. Then, an empirical application on an AR(1) univariate panel data model with individual effects is carried out using a cash-flow series for 174 firms in the United States from 1981-1985. Finally, conclusions and appendices on the entire analysis are provided at the end of the dissertation.

Chapter 1

Background

1.1 Introduction

The aim of this Chapter is to introduce the EL estimation procedure and to provide the background material that is most useful for the topics covered in this thesis. The general concept of EL is given in Section 1.2. We present EL as discussed by Owen (1991, 1990, 1988) in Section 1.3. Section 1.4 introduces estimating equations into the analysis as in Qin and Lawless (1994). Section 1.5 presents the Maximum Empirical Likelihood (MEL) estimator and its first-order asymptotic properties. Section 1.6 reviews some relevant hypothesis tests and special emphasis is given to the overidentification test.[1] Section 1.7 defines the Cressie-Read statistics, identifies the EL and Kullback Leibler Information Criterion (KLIC) as members of this family and concentrates on the KLIC overidentifying restrictions test. Section 1.8 analyses two other overidentification tests. These are based on the two-step

[1]The terms overidentification test/statistic, J-test, moment restrictions test, and overidentifying restrictions test/statistic are used interchangeably.

and the continuously updated GMM estimators. An efficient bootstrapping technique is described in Section 1.9. This procedure is used to obtain bootstrapped critical values for the overidentification test statistics, both for the GMM and the EL. Monte Carlo simulation is introduced in Section 4.5. Finally, a software section is included.

1.2 Concept of Empirical Likelihood

According to Owen (1988), who originally developed the EL approach, the name "Empirical Likelihood" was adopted because the empirical distribution of the data plays a central role. It was not called nonparametric likelihood, so as not to assume that it would be the only way to extend nonparametric maximum likelihood to likelihood ratio functions.

The EL is a nonparametric method of statistical inference based on a data-driven likelihood ratio function. The EL approach yields an estimator, the MEL estimator, that has sampling properties similar to resampling methods such as the bootstrap and an inference basis analogous to that used with parametric methods (Owen, 2001). However, instead of the resampling process underlying the bootstrap, the EL method works by profiling a multinomial likelihood supported on the sample data. In situations involving independent and identically distributed (*i.i.d.*) random variables, EL has the advantage over some parametric methods in that it only makes mild assumptions about the existence of certain moments or estimating equations of the random variables.

1.3 Empirical Likelihood and the Mean Problem

Owen (1991, 1990, 1988) proposed EL for the mean and some other statistics, using results of Stein (1956) and extending earlier work of Thomas and Grunkemeier (1975).[2]

Let $x_1, x_2, ..., x_n$ be *i.i.d.* observations from a d-variate distribution F with mean μ and nonsingular covariance matrix. In order to recover an estimate of the probability distribution function from the observed sample $x_1, x_2, ..., x_n$; Stein (1956) approximates it with a multinomial distribution. Owen (1990) applies Stein's estimate and defines a nonparametric (multinomial type) likelihood function, the EL function

$$L(F) = \prod_{i=1}^{n} dF(x_i) = \prod_{i=1}^{n} p_i, \tag{1.1}$$

where $p_i = dF(x_i) = \Pr(X = x_i)$. Only distributions with an atom of probability on each x_i have nonzero likelihood, and (1.1) is maximized by the Empirical Distribution Function (EDF)

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(x_i < x), \tag{1.2}$$

where $I(x_i < x) = \begin{cases} 1 \text{ if } x_i < x \\ 0 \text{ otherwise} \end{cases}$.

Now, we define the ELR function

$$R(F) = L(F)/L(F_n) = \prod_{i=1}^{n} np_i. \tag{1.3}$$

$R(F)$ can be used to construct nonparametric confidence regions and tests for the mean μ of F. We define the profile ELR function (this is a likelihood function that has been

[2]Thomas and Grunkemeier (1975) are the first to use an ELR function to set confidence intervals.

partially maximized with respect to a subset of its parameters conditional on given values of the remaining parameters)

$$R_E(\mu) = \sup\left\{\prod_{i=1}^{n} np_i \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i x_i = \mu\right\}. \tag{1.4}$$

Owen (1990, 1988) noted that a unique value for the right hand side of $R_E(\mu)$ exists –the maximum of the ELR is unique– provided that μ is inside the convex hull of the points $x_1, ..., x_n$.

An explicit expression for $R_E(\mu)$ can be derived by a Lagrange multiplier argument. The maximum of $\prod_{i=1}^{n} np_i$ subject to the constraints $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} p_i x_i = \mu$ is attained when

$$p_i = p_i(\mu, t) = n^{-1}\{1 + t^\tau(x_i - \mu)\}^{-1}, \tag{1.5}$$

where the multiplier $t = t(\mu)$ is the $d \times 1$ vector solution to

$$\sum_{i=1}^{n}\{1 + t^\tau(x_i - \mu)\}^{-1}(x_i - \mu) = 0 \tag{1.6}$$

and t^τ is the transpose of t. Since $\prod_{i=1}^{n} p_i$ is maximized unconditionally by F_n, it follows that $R_E(\mu)$ is maximized with respect to μ at $\widehat{\mu} = \overline{x}$, and that

$$R_E(\mu) = \prod_{i=1}^{n}\{1 + t^\tau(x_i - \mu)\}^{-1}. \tag{1.7}$$

The empirical log-likelihood ratio statistic is

$$W_E(\mu) = -2\log R_E(\mu). \tag{1.8}$$

Under mild conditions –which include finite variance covariance matrix of full rank d and finite second and third moments– Owen (1990, 1988) proves that if $\mu = \mu_0$, then $W_E(\mu_0)$

converges in distribution to $\chi^2_{(d)}$ as $n \to \infty$; *i.e.* :

$$W_E(\mu_0) \to \chi^2_{(d)}. \tag{1.9}$$

The asymptotic result in (1.9) is very useful since it enables us to obtain confidence regions and to carry out hypothesis tests. Approximate $\alpha - level$ confidence regions for μ may therefore be obtained as the set of points μ such that $W_E(\mu) \leq c_\alpha$, where c_α is defined such that $\Pr\left(\chi^2_{(d)} \leq c_\alpha\right) = \alpha$. Hypothesis tests are reviewed in Section 1.6.

1.4 Empirical Likelihood and Estimating Equations

Qin and Lawless (1994) extend Owen's (1991, 1990, 1988) formulation by combining the concept of unbiased estimating functions and EL. They assume that $x_1, x_2 ...., x_n$ are *i.i.d.* random variables from an unknown distribution function F, that there is a q-dimensional parameter θ associated with F and that information about θ and F is available in the form of $r \geq q$ functionally independent unbiased estimating functions. The functions are written as

$$g_j(x, \theta), \quad j = 1, 2, ..., r,$$

such that

$$E_F\{g_j(x, \theta)\} = 0.$$

The notation E_F is used to emphasize that expectations are being taken with respect to F. In vector form we have

$$g(x, \theta) = (g_1(x, \theta), ..., g_r(x, \theta))^T,$$

$$E_F\{g(x,\theta)\}=0.$$

Note that when $r=q$, estimators of the parameters can be obtained as roots of the corresponding estimating equations.

To apply EL to this framework we maximize the logarithm of (1.1) subject to: $p_i \geq 0$, $\sum_i p_i = 1$ and $\sum_i p_i g(x_i,\theta) = 0$ via Lagrange multipliers. Let

$$\mathfrak{L}(p,\lambda,t)=\sum_i \ln(p_i)+\lambda\left(1-\sum_i p_i\right)-nt^{\tau}\sum_i p_i g(x_i,\theta),$$

where λ and $t=(t_1,t_2,...,t_r)^{\tau}$ are Lagrange multipliers.

The first order conditions (FOC) for p_i, λ and t are

$$\frac{1}{p_i}=\lambda+nt^{\tau}g(x_i,\theta), \tag{1.10}$$

$$\sum_i p_i = 1, \tag{1.11}$$

$$\sum_i p_i g(x_i,\theta)=0. \tag{1.12}$$

Multiplying (1.10) by p_i, summing over i and using (1.11) and (1.12) yields

$$\lambda = n,$$

$$p_i=p_i(\theta,t)=n^{-1}\{1+t^{\tau}g(x_i,\theta)\}^{-1}, \tag{1.13}$$

with the restriction from (1.12) that

$$\sum_i p_i g(x_i,\theta)=\sum_i n^{-1}\{1+t^{\tau}g(x_i,\theta)\}^{-1}g(x_i,\theta)=0. \tag{1.14}$$

Qin and Lawless (1994) show that a solution for t can be determined in terms of θ from (1.14) if:

(i) $0 \leq p_i \leq 1$, which implies that t and θ must satisfy $1 + t^T g(x_i, \theta) \geqslant 1/n$ for each i,

(ii) 0 is inside the convex hull of the $g(x_i, \theta)'s$.

By substituting the optimal Lagrange multiplier, $t(\theta)$, into the expression for the optimal p weights, $p_i(\theta, t)$ in (1.13), the empirical probabilities can be represented in terms of θ as

$$p_i(\theta, t(\theta)) = n^{-1}\{1 + t(\theta)^T g(x_i, \theta)\}^{-1}. \tag{1.15}$$

It follows that the (profile) EL function for θ takes the form

$$L_E(\theta) = \prod_{i=1}^{n}\left\{\left(\frac{1}{n}\right)\frac{1}{1 + t^T(\theta) g(x_i, \theta)}\right\}. \tag{1.16}$$

Since $p_i = n^{-1}$ in the absence of constraints, the empirical log-likelihood ratio is

$$l_E(\theta) = \sum_{i=1}^{n} \ln[1 + t^T(\theta) g(x_i, \theta)]. \tag{1.17}$$

1.5 Maximum Empirical Likelihood Estimator

The form of $\widetilde{\theta}_{EL}$, the MEL estimator for θ, is the solution to the minimization of (1.17).[3] Substituting $\widetilde{\theta}_{EL}$ into (1.15) we find

$$\widetilde{p}_{iEL} = p_{iEL}\left(\widetilde{\theta}_{EL}, t\left(\widetilde{\theta}_{EL}\right)\right) = n^{-1}\left\{1 + t\left(\widetilde{\theta}_{EL}\right)^T g\left(x_i, \widetilde{\theta}_{EL}\right)\right\}^{-1}, \tag{1.18}$$

and an estimator for the distribution function F is

$$\widetilde{F}_{nEL}(x) = \sum_i \widetilde{p}_{iEL} I(x_i < x). \tag{1.19}$$

[3]When $r > q$ computational issues arise as the best way to obtain $\widetilde{\theta}_{EL}$. We will discuss computational procedures in more detail in the following Section and in Chapter 2.

1.5.1 Asymptotic Properties

Qin and Lawless (1994) derive asymptotic properties for the MEL estimator. These are summarized in the following Theorems and Corollaries. For outlines of proofs refer to Qin and Lawless (1994).

Lemma 1 *Assume that* $E\left[g\left(x,\theta_0\right)g^{\tau}\left(x,\theta_0\right)\right]$ *is positive definite,* $\partial g\left(x,\theta\right)/\partial\theta$ *is continuous in a neighbourhood of the true value* θ_0, $\left\|\partial g\left(x,\theta\right)/\partial\theta\right\|$ *and* $\left\|g\left(x,\theta\right)\right\|^3$ *are bounded by some integrable function* $G(x)$ *in its neighbourhood, and the rank of* $E\left[\partial g\left(x,\theta\right)/\partial\theta\right]$ *is* q. *Then, as* $n\to\infty$, *with probability one* $l_E\left(\theta\right)$ *attains its minimum value at some point* $\tilde{\theta}_{EL}$ *in the interior of the ball* $\left\|\theta-\theta_0\right\|\leq n^{-1/3}$, *and* $\tilde{\theta}_{EL}$ *and* $\tilde{t}_{EL}=t\left(\tilde{\theta}_{EL}\right)$ *satisfy:*

$$Q_{1n}\left(\tilde{\theta}_{EL},\tilde{t}_{EL}\right)=0,\ Q_{2n}\left(\tilde{\theta}_{EL},\tilde{t}_{EL}\right)=0; \tag{1.20}$$

where

$$Q_{1n}\left(\theta,t\right)=\frac{1}{n}\sum_i\frac{1}{1+t^{\tau}g\left(x_i,\theta\right)}g\left(x_i,\theta\right), \tag{1.21}$$

$$Q_{2n}\left(\theta,t\right)=\frac{1}{n}\sum_i\frac{1}{1+t^{\tau}g\left(x_i,\theta\right)}\left(\frac{\partial g\left(x_i,\theta\right)}{\partial\theta}\right)^{\tau}t. \tag{1.22}$$

Note that in the EL procedure, the focus is on a vector of empirical estimating equations. The r dimensional vector of moment equations, $g\left(x,\theta_0\right)$, coupled with the adding up restriction $\sum_i p_i=1$ are sufficient to determine only $(r+1)$ of the $(n+q)$ unknowns, θ and p, leaving $[n+q-(r+1)]$ of the unknowns undetermined. The undetermined nature of the system of moment equations is resolved by the introduction of an estimation objective function, $\sum_i \ln p_i$, which optimized subject to the moment conditions determines the

$(n+q)$ unknowns. More specifically, from (1.21) an expression for each estimating equation is obtained, yielding r equations. From (1.22) an expression for each parameter is attained, yielding q equations. So there are $q+r$ equations, which is identical to the number of parameters to be estimated: q original ones and r Lagrange multipliers which were imposed for each estimating equation. We will refer to (1.20) as the FOC of the empirical log-likelihood function. Note that a closed-form exists for neither $\widetilde{\theta}_{EL}$ nor $\widetilde{t}_{EL}$. So we must either solve the FOC numerically or optimize the empirical log-likelihood function directly. We analyse two numerical optimization methods based on simultaneous and sequential algorithms in Chapter 2.[4]

The conditions stated in Lemma 1 are relatively mild assumptions. In order to define the limiting distribution of the MEL estimator, some conditions must be added. These are provided in Theorem 1, which formally presents the first-order asymptotic properties of the statistics.

Theorem 1 *In addition to the conditions of Lemma* (1)*, we assume that* $\frac{\partial^2 g(x,\theta)}{\partial\theta\partial\theta'}$ *is continuous in* θ *in a neighbourhood of the true value* θ_0*. If* $\left\|\frac{\partial^2 g(x,\theta)}{\partial\theta\partial\theta'}\right\|$ *can be bounded by some integrable function* $G(x)$ *in the neighbourhood, then:*

$$\sqrt{n}\left(\widetilde{\theta}_{EL}-\theta_0\right)\rightarrow N\left(0,V\right), \tag{1.23}$$

$$\sqrt{n}\left(\widetilde{t}_{EL}-0\right)\rightarrow N\left(0,U\right), \tag{1.24}$$

$$\sqrt{n}\left(\widetilde{F}_{n_{EL}}\left(x\right)-F\left(x\right)\right)\rightarrow N\left(0,W\left(x\right)\right), \tag{1.25}$$

[4]In Chapter 12 of Owen (2001), several optimization methods and algorithms are discussed. Qin and Lawless (1994) also present computational issues through various examples.

$$\sqrt{n}\left(\widetilde{p}_{i_{EL}} - p_i\right) I\left(x_i \leq x\right) \rightarrow N\left(0, \eta^2\right); \tag{1.26}$$

where

$$\widetilde{F}_{n_{EL}}(x) = \sum_i \widetilde{p}_{i_{EL}} 1\left(x_i < x\right), \tag{1.27}$$

$$\widetilde{p}_{i_{EL}} = \left(\frac{1}{n}\right) \frac{1}{1 + \widetilde{t}^{\tau} g\left(x_i, \widetilde{\theta}_{EL}\right)}, \tag{1.28}$$

$$V = \left[E\left(\frac{\partial g}{\partial \theta}\right)^{\tau} \left(E g g^{\tau}\right)^{-1} E\left(\frac{\partial g}{\partial \theta}\right)\right]^{-1}, \tag{1.29}$$

$$W(x) = F(x)\left(1 - F(x)\right) - B(x) U B^{\tau}(x), \tag{1.30}$$

$$B(x) = E\left\{g\left(x_i, \theta_0\right) I\left(x_i < x\right)\right\}, \tag{1.31}$$

$$U = \left[E\left(g g^{\tau}\right)\right]^{-1} \left\{I - E\left(\frac{\partial g}{\partial \theta}\right) V E\left(\frac{\partial g}{\partial \theta}\right)^{\tau} \left[E\left(g g^{\tau}\right)\right]^{-1}\right\}, \tag{1.32}$$

$$\eta^2 = \left[E\left(g g^{\tau}\right)\right]^{-1} - B^{\tau} U B, \tag{1.33}$$

and $\widetilde{\theta}_{EL}$ *and* $\widetilde{t}_{EL}$ *are asymptotically uncorrelated.*

The asymptotic variance, V, is consistently estimated by

$$\left[\left\{\sum_i \widetilde{p}_{i_{EL}} \frac{\partial g\left(x_i, \widetilde{\theta}_{EL}\right)}{\partial \theta}\right\}^{\tau} \left\{\sum_i \widetilde{p}_{i_{EL}} g\left(x_i, \widetilde{\theta}_{EL}\right) g^{\tau}\left(x_i, \widetilde{\theta}_{EL}\right)\right\}^{-1} \left\{\sum_i \widetilde{p}_{i_{EL}} \frac{\partial g\left(x_i, \widetilde{\theta}_{EL}\right)}{\partial \theta}\right\}\right]^{-1}.$$

We give some additional asymptotic properties of MEL estimators and statistics in the following corollaries to Theorem 1 (these corollaries are due to Qin and Lawless, 1994).

The MEL estimator is equivalent in asymptotic distribution to the most efficient estimator in the class of estimators defined through the solution to estimating equations formed by linear combinations of the unbiased estimating functions (see McCullagh and Nelder, 1989).

When the number of moment constraints, r, exceeds the number of unknown parameters, q, the EL approach defines the optimal combination of the moment equations, *i.e.* the asymptotic efficiency of the most efficient estimating equations estimator is duplicated. This result is formalized below.

Corollary 1 *$\widetilde{\theta}_{EL}$ based on $g_1(x,\theta)$,..., $g_r(x,\theta)$ is fully efficient in the sense that it has the same asymptotic variance as the optimal estimator obtained from the class of $(q \times 1)$ estimating equations that are linear combinations of $g_1(x,\theta),...,g_r(x,\theta)$.*

The following Corollary gives an important asymptotic property of the MEL estimator.

Corollary 2 *If the conditions given in Lemma 1 and Theorem 1 are fulfilled then when $r > q$, the asymptotic variance $V{=}V_r$ of $\sqrt{n}\left(\widetilde{\theta}_{EL} - \theta\right)$ cannot decrease if an estimating equation is dropped.*

1.6 Hypothesis Tests

In the EL context the usual likelihood ratio, Wald and Lagrange multiplier tests can be constructed. It is the ELR on which we focus in the current investigation (for details related to the Wald and Lagrange Multiplier tests refer to Owen, 2001).

1.6.1 Empirical Likelihood Ratio

The following Theorem, which was proved by Qin and Lawless (1994), allows us to use the ELR statistic for testing and/or obtaining confidence limits for parameters.

Theorem 2 *Under the assumptions of Theorem* 1, *the ELR statistic for testing* $H_0 : \theta = \theta_0$ *is*

$$W_E(\theta_0) = 2l_E(\theta_0) - 2l_E\left(\widetilde{\theta}_{EL}\right), \tag{1.34}$$

$$W_E(\theta_0) \to \chi^2_{(q)}$$

as $n \to \infty$ *when* H_0 *is true.* l_E *is the empirical log-likelihood function as given in (1.17).*

An asymptotic $\alpha - level$ test of H_0 is

$$reject\ H_0 : \theta = \theta_0 \ \ if \ \ W_E(\theta_0) \geq c_\alpha, \tag{1.35}$$

where c_α is defined such that $Pr\left(\chi^2_{(q)} \leq c_\alpha\right) = 1 - \alpha$. An asymptotic $100(1-\alpha)\%$ confidence region for θ can be obtained in the usual way applying the duality principle to the test procedure given in (1.35), resulting in the set of θ_0 values not rejected by the test.

1.6.2 Overidentification Test

Anderson and Rubin (1949) derived the likelihood ratio statistic for testing the overidentifiability conditions on a structural equation in a simultaneous equation system. The null hypothesis of Anderson and Rubin (1949) is that the estimated equation is correctly specified. Sargan (1958) noted that this test should have some power against false orthogonality conditions and proposed a misspecification test, the so-called Sargan test, which can also be interpreted as a test of overidentifying restrictions. The overidentifying restrictions test of Hansen (1982) tests the restrictions implied by the econometric model for the GMM. His test is an extension of the specification test proposed by Sargan (1958).

Overidentified models, in moments-based settings, imply that there need not be a parameter value θ such that the moment condition $E\left[g\left(x,\theta\right)\right] = 0$ holds. Thus the model, the overidentifying restrictions, are testable.

In practice, however, it is not clear which of the orthogonality conditions identify and which overidentify the model. Rejection of the null does not specify which of the restrictions are false (see Magdalinos and Symeonides, 1996). Note that two circumstances that lead to the null hypothesis being rejected is either because of an exclusion restriction or false orthogonality conditions.

The ELR's test for overidentifying moment conditions requires two values for the probabilities. One in which the overidentifying restrictions holds, $\widehat{p}_{iEL}$ in (1.18), and one in which these are removed from the optimization problem, $p_i = \frac{1}{n}$. After substituting $\widehat{p}_{iEL}$ and $p_i = \frac{1}{n}$ in the EL functions, a test of the r restrictions can be conducted based on the ELR statistic. Corollary 3 formalizes the test and provides the asymptotic distribution of the statistic (this Corollary corresponds to Corollary 4 in Qin and Lawless, 1994).

Corollary 3 *Under the conditions of Theorem* 1, *the statistic given by* W_j *is asymptotically* $\chi^2_{(r-q)}$ *if the estimating equations are unbiased, i.e.*

$$W_j = 2\sum_i \log\left[1 + \widetilde{t}^T_{EL}\; g\left(x_i, \widetilde{\theta}_{EL}\right)\right] \;\; \rightarrow \;\; \chi^2_{(r-q)}.$$

An asymptotic $\alpha - level$ test of the validity of the moment restrictions is then conducted as:

$$reject\ H_0 : E\left[g\left(x,\theta\right)\right] = 0 \ \ if \ \ W_j \geq c_\alpha, \tag{1.36}$$

where c_α is defined such that $\Pr\left(\chi^2_{(r-q)} \leq c_\alpha\right) = 1-\alpha$. An asymptotic $100\left(1-\alpha\right)\%$ confidence region can be obtained in the usual way applying the duality principle to the test procedure given in (1.36).

1.7 Cressie-Read Statistics

Maximizing the EL function and producing the MEL estimator is but one way to define an EL estimator that has good first-order asymptotic properties. Well known alternatives include

$$KLIC = \sum_{i=1}^{n} p_i \ln\left(np_i\right), \tag{1.37}$$

and the Hellinger distance

$$H = \sum_i \left(p_i^{1/2} - n^{-1/2}\right)^2.$$

EL, KLIC and H belong to the Cressie-Read family, which we analyse in some detail. The Cressie-Read power divergence statistic for a multinomial with O_i observations where E_i observations were expected takes the form

$$CR\left(\varphi\right) = \frac{2}{\varphi\left(\varphi+1\right)} \sum_{i=1}^{k} O_i \left[\left(\frac{O_i}{E_i}\right)^{\varphi} - 1\right], \tag{1.38}$$

where $-\infty < \varphi < \infty$ (for a description of power divergence discrepancies refer to Cressie and Read, 1984). The cases $\varphi \in \{-1, 0\}$ are handled by taking limits. To derive the EL and KLIC as members of this family consider that their setups have n distinct values observed once each. Thus $k = n$ and $O_i = 1$ and we write $E_i = np_i$. Then,

$$CR(\varphi) = \frac{2}{\varphi(\varphi+1)}\sum_i \left[(np_i)^{-\varphi} - 1\right].$$

After taking the required limits

$$CR(0) = -2\sum_i \log(np_i),$$

$$CR(-1) = -2\sum_i np_i \log(np_i).$$

The quantity $CR(0)$ is minus twice the empirical log-likelihood ratio, and $CR(-1)$ is equal to $2n \times KLIC$. Note that different values of φ give rise to new members of the Cressie-Read family. The value $\varphi = -2$ corresponds to the Euclidean log-likelihood. The value $\varphi = 1$ yields the Pearson's χ^2 and the Freeman-Tukey statistic follows with $\varphi = -1/2$.

All members of the Cressie-Read family originate empirical divergence analogues of the EL.

1.7.1 Kullback Leibler Information Criterion

The KLIC is often singled out because of its interpretation as a measure of entropy. We can obtain optimal probability weights, $\widehat{p}_{iKLIC}$, in an analogous way to the MEL context. The corresponding problem for the KLIC is to maximize

$$-\sum_i p_i \log(np_i) \quad \text{subject to} \quad p_i \geq 0, \quad \sum_i p_i = 1 \;\; and \;\; \sum_i p_i g(x_i, \theta) = 0. \tag{1.39}$$

Upon close examination it is evident that, except for the estimation objective function, the form of the problem and the basis for a solution closely mirror that of the MEL. We will refer to the maximum KLIC estimator as $\widehat{\theta}_{KLIC}$. Imbens (1997) demonstrates that under regularity conditions analogous to those assumed in the MEL context, $\widehat{\theta}_{KLIC}$ and $\widetilde{\theta}_{EL}$ have the same limiting distribution. Due to the fact that $\widehat{\theta}_{KLIC}$ is consistent and asymptotically normally distributed, asymptotic tests and confidence regions can be based on asymptotic normality of the estimator analogous to the MEL estimator case.

Overidentification Test

An asymptotic chi-square overidentification test can be based on a suitably scaled version of the KLIC. The test statistic is given by

$$J_{KLIC} = 2\,n \sum_i \widehat{p}_{iKLIC}\ \ln\left(n\ \widehat{p}_{iKLIC}\right).$$

If the null hypothesis is true, $H_0 = E\left[g\left(x, \theta\right)\right] = 0$, then

$$J_{KLIC} \longrightarrow \chi^2_{(r-q)}$$

(for a proof that all members of the Cressie-Read family have a χ^2 calibration see Baggerly, 1998). An asymptotic $\alpha - level$ test of the validity of the moment restrictions is then conducted as:

$$reject\ H_0 : E\left[g\left(x, \theta\right)\right] = 0\ \ if\ \ J_{KLIC} \geq c_\alpha, \tag{1.40}$$

where c_α is defined such that $\Pr\left(\chi^2_{(r-q)} \leq c_\alpha\right) = 1 - \alpha$.

1.8 General Method of Moments

GMM estimators are based on the moment restrictions $E\left[g\left(X, \theta_0\right)\right] = 0$. As for the EL, these moment restrictions are a partial implication of some model, although they may also embody all the available information.

For our purposes a GMM estimator will be obtained as the solution to

$$\min_{\theta} \left[Q_W(\theta) \right],$$

where

$$Q_W(\theta) = \left[\frac{1}{n} \sum_i g(x_i, \theta) \right]^{\tau} W^{-1} \left[\frac{1}{n} \sum_i g(x_i, \theta) \right]$$

for some positive definite matrix W (for the asymptotic theory and a detailed description of the GMM see Newey and McFadden, 1994; Hansen, 1982).

1.8.1 Overidentification Test

Two-Step Estimator

An efficient estimator can be based on minimizing $Q_W(\theta)$ for $W = W_0 \equiv E\left[g(x, \theta_0)\, g(x, \theta_0)^{\tau}\right]$. A feasible version of this efficient procedure is based on an initial consistent estimator $\overline{\theta}$ of θ_0 obtained by minimizing $Q_W(\theta)$ for an arbitrary choice of W such as the $dim(g)$ dimensional identity matrix. The inverse of the optimal weight matrix, W_0, is then estimated as

$$\widehat{W} = \frac{1}{n} \sum_{i=1}^{n} g\left(x_i, \overline{\theta}\right) g\left(x_i, \overline{\theta}\right)^{\tau}.$$

An efficient two-step GMM estimator, $\widehat{\theta}_{2GMM}$, is obtained by minimizing $Q_{\widehat{W}}(\theta)$. Under general regularity conditions including that the model is correctly specified, and that there is indeed a unique value θ_0 such that $E[g(x,\theta_0)] = 0$[5], then

$$\sqrt{N}\left(\widehat{\theta}_{2GMM} - \theta_0\right) \rightarrow N\left(0, \left(\Gamma^{\tau}\Delta^{-1}\Gamma\right)^{-1}\right),$$

where

$$\Delta = E\left[g(x,\theta_0)\, g(x,\theta_0)^{\tau}\right],$$
$$\Gamma = E\left(\frac{\partial g}{\partial \theta}^{\tau}(x,\theta_0)\right).$$

The statistic for testing restrictions in this context is the J-test. Under the hypothesis of correct specification

$$J_{2GMM} = n \times Q_{\widehat{W}}\left(\widehat{\theta}_{2GMM}\right) \rightarrow \chi^2_{(r-q)}.$$

It is important to stress that unless otherwise stated we set W as the identity matrix in the first-step of the GMM procedure.

Continuously Updated Estimator

The continuously updated GMM overidentification test is based on the continuously updated GMM estimator, $\widehat{\theta}_{CuGMM}$. Instead of taking the weighted average as given in each step of the GMM's estimation, the covariance matrix is continuously altered as $\overline{\theta}$ is changed in the minimization. The statistic for testing the moment conditions is

$$J_{CuGMM} = n \times Q_{W\left(\widehat{\theta}_{CuGMM}\right)}\left(\widehat{\theta}_{CuGMM}\right)$$

Under the null of correct specification

$$J_{CuGMM} \rightarrow \chi^2_{(r-q)}.$$

[5]The complete set of regularity conditions under which GMM estimators have good asymptotic properties is given in Theorem 2.6 and 3.4 of Newey and McFadden (1994).

1.9 Efficient Bootstrap

For the theory of the bootstrap upon which some of the procedures described below draw see, for example: Horowitz (1997), Shao and Tung (1995), Efron and Tibshirani (1993), Hall (1992), Beran and Ducharme (1991).

Brown and Newey (2001) define efficient bootstrap as the bootstrap based on resampling from the EL distribution, that imposes the moment conditions, rather than the empirical distribution.

It is fairly well known that in general the moment restrictions are not satisfied for the empirical distribution. One situation in which this happens is GMM estimation of an overidentified parameter when the EDF is considered (see Horowitz, 2000). To explain why assume that $\dim(g) > \dim(\theta)$ and that the distribution of x does not belong to a known parametric family so that the EDF of x is considered. The sample analog of $E[g(x,\theta)]$ is $E_F[g(x,\theta_0)] = \frac{1}{n}\sum_{i=1}^{n} g(x_i,\theta)$ and that of θ_0 is $\widehat{\theta}_{2GMM}$. In general, $E_F\left[g\left(x,\widehat{\theta}_{2GMM}\right)\right] \neq 0$ in an overidentified model, so bootstrap estimation based on the EDF of x implements a moment condition that does not hold in the population the bootstrap resamples. As a result, the bootstrap estimator of the distribution of the statistic for testing the overidentifying restrictions is inconsistent, because its limiting distribution has a discontinuity where the estimated moment conditions depart from zero (see Brown and Newey, 2001). Hall and Horowitz (1996) show that recentering the moment conditions yields asymptotic refinements for the rejection probabilities of tests of overidentifying restrictions.

Efficient bootstrap does not require recentering the moment conditions since the estimated moment conditions based on the EL distribution are already equal to zero. Brown and Newey (2001) show that the efficient bootstrap yields a larger sample improvement relative to Hall and Horowitz (1996) recentering technique.

The efficient bootstrap can be described in some simple steps:

1. Draw n observations $x_1, ..., x_n$; each satisfying the moment condition

   $E[g(X, \theta_0)] = 0.$

2. Calculate $\widetilde{p}_{iEL}$.[6]

3. Draw n *i.i.d.* observations $x_1^b, ..., x_n^b$ with replacement from the distribution with

   $\Pr(x = x_i) = \widetilde{p}_{iEL}.$

Efficient bootstrap differs from standard approaches in the use of $\widetilde{p}_{iEL}$ in step 2 rather than $1/n$ for the probability of the i^{th} observation. Figure 1.1 illustrates this fact. The MEL sample weights with which efficient bootstrap resamples are non-uniform. This can be seen from the curvature of the mass function, which reflects the impact of the moment equations.[7] It seems appealing to weight the bootstrap sampling procedure using the best estimate of these probabilities, $\widetilde{p}_{iEL}$, rather than an inefficient estimate of these probabilities, $1/n$, since the former incorporate all available information.

[6] As defined in (1.18).

[7] Here, we considered the first and second moments of a random variable, as in Qin and Lawless (1994). The observations in the sample were ordered and the mass assigned to each element is indicated by the height of the associated bar.

Note that a flat mass function is inherent to uniform probabilities, these are the unrestricted probabilities with which standard bootstrap procedures resample.

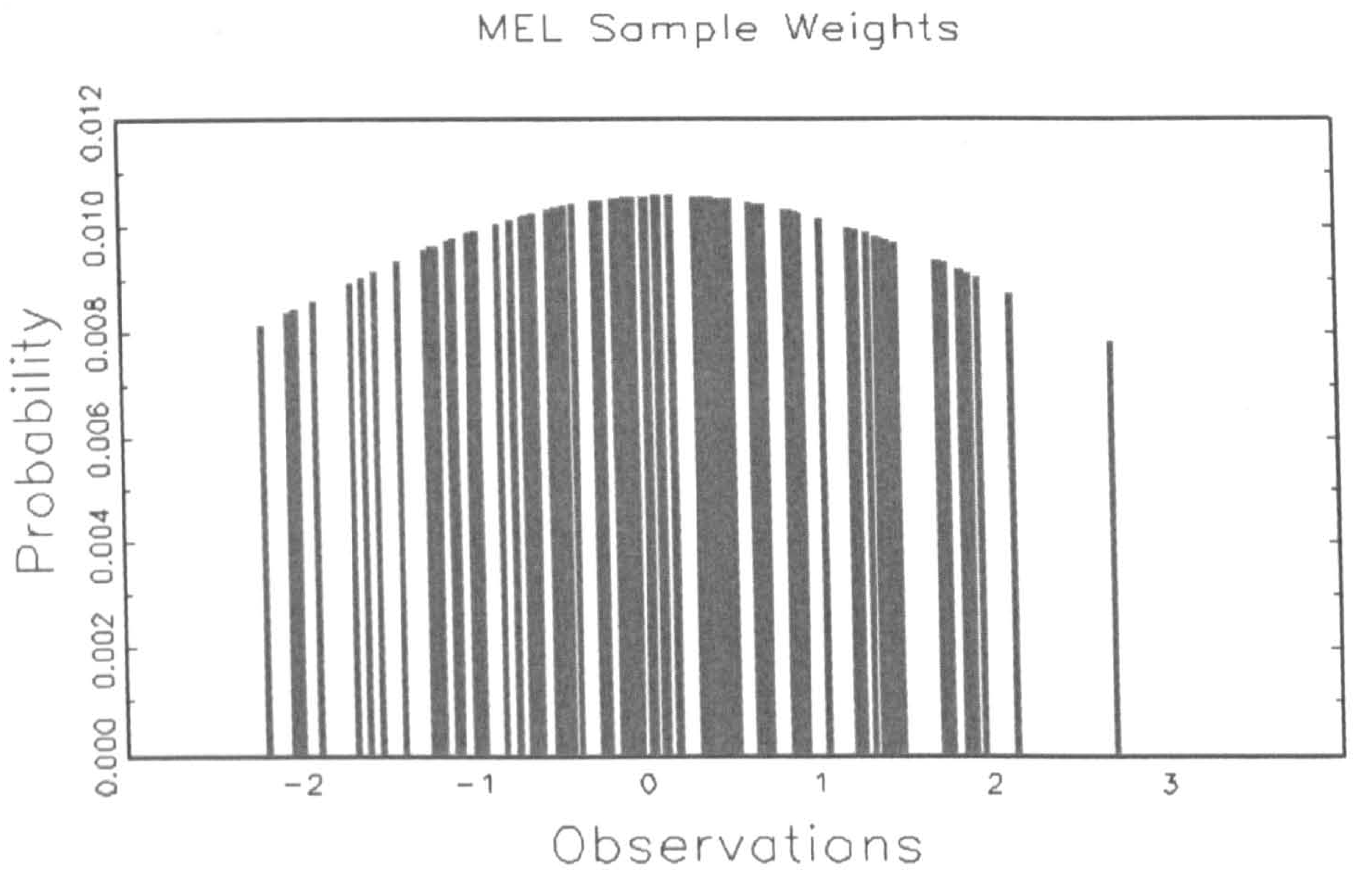

Figure 1.1: Efficient Bootstrap

Brown and Newey (2001) applied efficient bootstrap for the GMM. We apply it for the EL. We define the former as GMM-bootstrap and the latter as EL-bootstrap.

To compute the GMM-bootstrap and the EL-bootstrap overidentification tests the following steps are added to steps 1 to 3.

1.9.1 Overidentification Tests

GMM-bootstrap

4. Calculate the overidentification test statistic

$J_{2GMM}\left(x_1^b, ..., x_n^b, \widehat{\theta}_{2GMM}\right) = J_{2GMM}^b.$

5. Repeat steps 3 to 4 B times, where B is an integer, to obtain $J_{2GMM}^1, ..., J_{2GMM}^B$.

6. Let the estimator of the distribution of $J_{2GMM}(x_1, ..., x_n)$ be the discrete distribution with $\Pr\left(J_{2GMM}(x_1, ..., x_n, \theta_0) = J_{2GMM}^b\right) = 1/B$.

7. Let $\widehat{q}_\alpha^B$ be the $(1-\alpha)$ quantile of the empirical distribution from step 6.

8. A test that rejects if $J_{2GMM}(x_1, ..., x_n) > \widehat{q}_\alpha^B$ is a GMM-bootstrap overidentification test.

EL-bootstrap

4. Calculate the overidentification test statistic

   $W_j\left(x_1^b, ..., x_n^b, \widetilde{\theta}_{EL}\right) = W_j^b.$

5. Repeat steps 3 to 4 B times, where B is an integer, to obtain $W_j^1, ..., W_j^B$.

6. Let the estimator of the distribution of $W_j(x_1, ..., x_n)$ be the discrete distribution with $\Pr\left(W_j(x_1, ..., x_n, \theta_0) = W_j^b\right) = 1/B$.

7. Let $\widehat{q}_\alpha^B$ be the $(1-\alpha)$ quantile of the empirical distribution from step 6.

8. A test that rejects if $W_j(x_1, ..., x_n) > \widehat{q}_\alpha^B$ is an EL-bootstrap overidentification test.

1.10 Monte Carlo Experimentation

There are two disadvantages of Monte Carlo experimentation as an approach to studying finite-sample behaviour: (*i*) results can be imprecise and, (*ii*) results may be specific to the particular parameter values being investigated and the distributional assumptions made. We discuss (*i*) and (*ii*) below.

(*i*) As far as precision is concerned, we have performed what we consider to be a relatively high number of replications, $m = 5000$, to achieve the desired level of precision for the Monte Carlo estimators[8].

We often wish to estimate the probability for a given Data Generating Process (DGP) of the test rejecting at some critical value c^*. That is, we want to estimate $q = P(S \geq c^*)$, where the finite-sample distribution of S is unknown. The rejection frequency (RF), or proportion of the simulated test statistics that exceed c^*, is the maximum likelihood estimator of q. This can be seen by considering each replication as the outcome of independent Bernoulli trials with 'success' parameter q so that the number of rejections is Binomial (m, q) and RF has mean q and variance $q(1-q)/m$ (see Bowsher, 2000a).

To derive a 95% confidence interval on the estimate of q, denote X_i as the ith realisation of the Bernoulli random variables so that RF$=m^{-1}\sum_{i=1}^{m} X_i$. Then by the Lindeberg-Levy Central Limit Theorem, RF converges in distribution to a Normal random variable with mean q and variance $q(1-q)/m$. Using this asymptotic result, the length of the 95% confidence

[8]Most of the literature that we reviewed used values of the order m=1000 and m=5000.

interval when $RF = q$ is given by

$$\varrho(q) = 3.92\left(\frac{q(1-q)}{m}\right)^{1/2} \tag{1.41}$$

which is a function of the unknown value q.

The values of (1.41) for $m = 5000$ are respectively .0166, .0120 and .0055 for $q = .10$, .05, .01.

(*ii*) As for specificity, we have investigated several DGPs by assuming different distributions and sample sizes for each one of the models that we investigate.

Despite these limitations, Monte Carlo experimentation provides a valuable approach to investigate the finite-sample properties of statistics since many problems are analytically intractable. We illustrate with a detailed example how to implement Monte Carlo simulations. The aim of the following experiment is to give further insights into the asymptotic normality property of the MEL, given in Theorem 1. Consider the two moment problem presented by Qin and Lawless (1994); *i.e.*:

$$E(x-\theta) = 0 \tag{1.42}$$

$$E\left(x^2 - 2\theta^2 - 1\right) = 0. \tag{1.43}$$

For three sample sizes, n=$\{30, 100, 200\}$, GAUSS generates pseudo-random samples which satisfy (1.42) and (1.43). Assume $N\left(\theta, \theta^2+1\right)$ and let $\theta = 0$, namely $N \sim (0,1)$. To estimate the unobserved parameter θ, we can form the EL function based on moments

(1.42) and (1.43) by noting that

$$g(x,\theta) = \begin{pmatrix} x-\theta \\ x^2-2\theta^2-1 \end{pmatrix}. \tag{1.44}$$

Our optimization problem can be written in terms of the Lagrange multiplier as

$$\mathcal{L}(p,\lambda,t) = \sum_i \ln(p_i) + \lambda\left(1-\sum_i p_i\right) - nt_1\sum_i p_i(x_i-\theta) - nt_2\sum_i p_i\left(x_i^2-2\theta^2-1\right).$$

The FOC of this problem, given in (1.20), are:

$$\frac{1}{n}\sum_i\left(\frac{x_i-\theta}{1+t_1(x_i-\theta)+t_2\left(x_i^2-2\theta^2-1\right)}\right)=0, \tag{1.45}$$

$$\frac{1}{n}\sum_i\left(\frac{x_i^2-2\theta^2-1}{1+t_1(x_i-\theta)+t_2\left(x_i^2-2\theta^2-1\right)}\right)=0, \tag{1.46}$$

$$\frac{1}{n}\sum_i\left(\frac{-t_1-4t_2\theta}{1+t_1(x_i-\theta)+t_2\left(x_i^2-2\theta^2-1\right)}\right)=0. \tag{1.47}$$

There are three equations and three parameters $-\theta$, t_1, t_2- to be estimated. We use the GAUSS procedure EqSolve[9] to compute the values of θ and t that satisfy:

(i) (1.45), (1.46) and (1.47),

(ii) $0 \leq p_i \leq 1$, which implies that t and θ must satisfy $1+t^{\tau}g(x_i,\theta) \geqslant 1/n$ for each i,

(iii) 0 is inside the convex hull of the $g(x_i,\theta)'s$.

[9]EqSolve, a simultaneous algorithm, is analysed in Section 2.3.

These values are: $\widehat{\theta}_{EL}$ and $t\left(\widehat{\theta}_{EL}\right)$.

GAUSS uses the MEL weights, $\widetilde{p}_{iEL}$, to generate bootstrap samples with replacement.[10] The number of bootstrap samples is 1000.[11] We must note that in the bootstrap context, the "true" value of parameters equals the parameter value estimated in the original solution to the estimation problem. While the larger the data sample, the more probable is that these original parameter estimates will be in a small neighbourhood of the true parameter value (due to the consistency of the estimation procedure). Because of random sampling variation, the result from a bootstrap distribution may have a central tendency that may not coincide with the actual-true values of the parameters (see Horowitz, 2000). Figures 1.2, 1.3 and 1.4 give the plots of the EDF of the bootstrap samples and the normal cumulative distribution function. The two vertical lines correspond to the true parameter value, $\theta_0 = 0$, and to the MEL estimator, $\widetilde{\theta}_{EL}$. We would expect that the EDF of the bootstrap samples approach the normal CDF as the sample size, n, increases.

[10]This is the efficient bootstrap technique descibed in Section 1.9.

[11]Authors in the bootstrap literature often recommend 1000 bootstrap samples for most applications. We noted that the results based on 100 bootstrap samples were very similar to those based on 1000 bootstrap samples for the models of Qin and Lawless (1994) and Hall and Horowitz (1996).

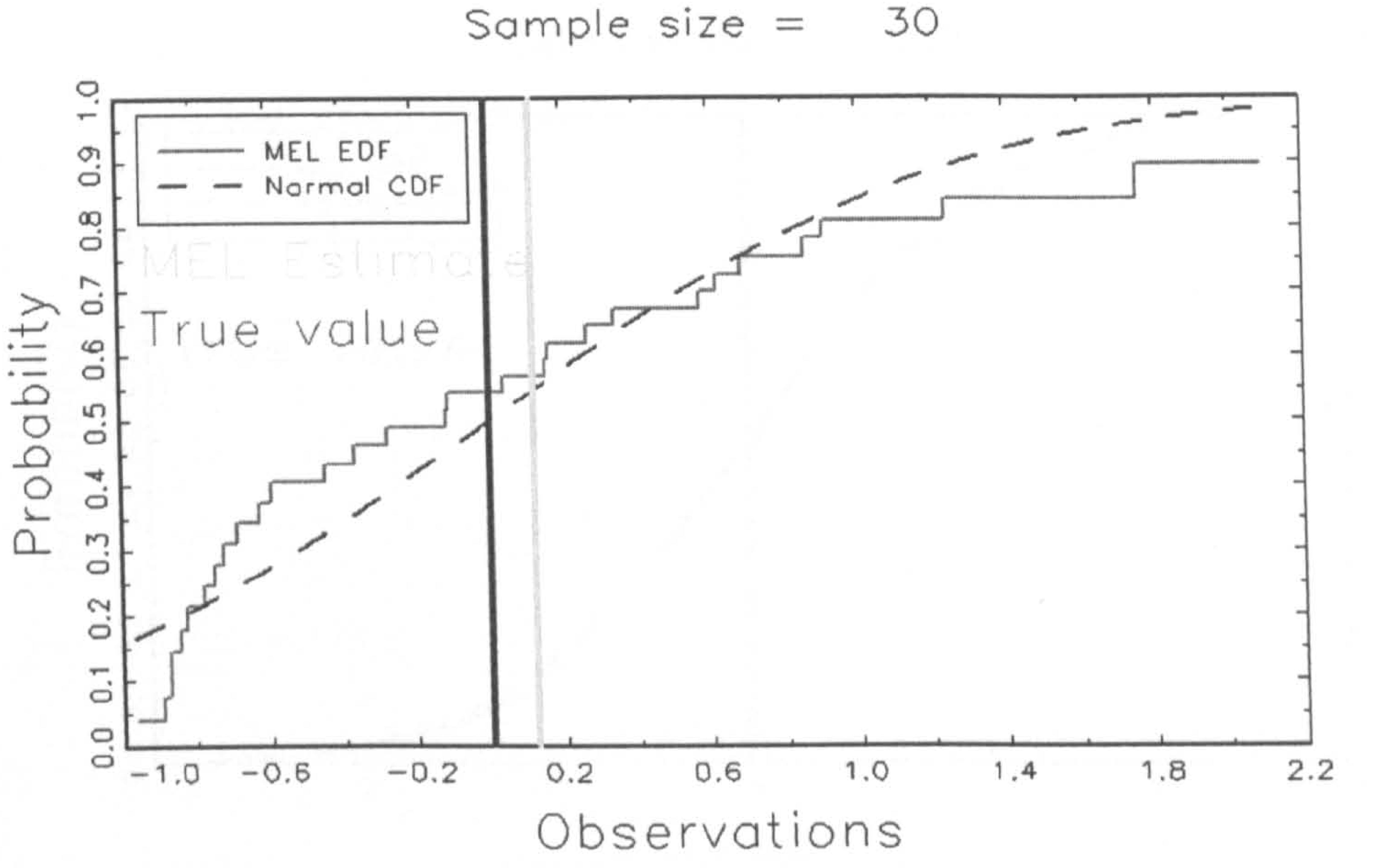

Figure 1.2: Asymptotic normality n=30

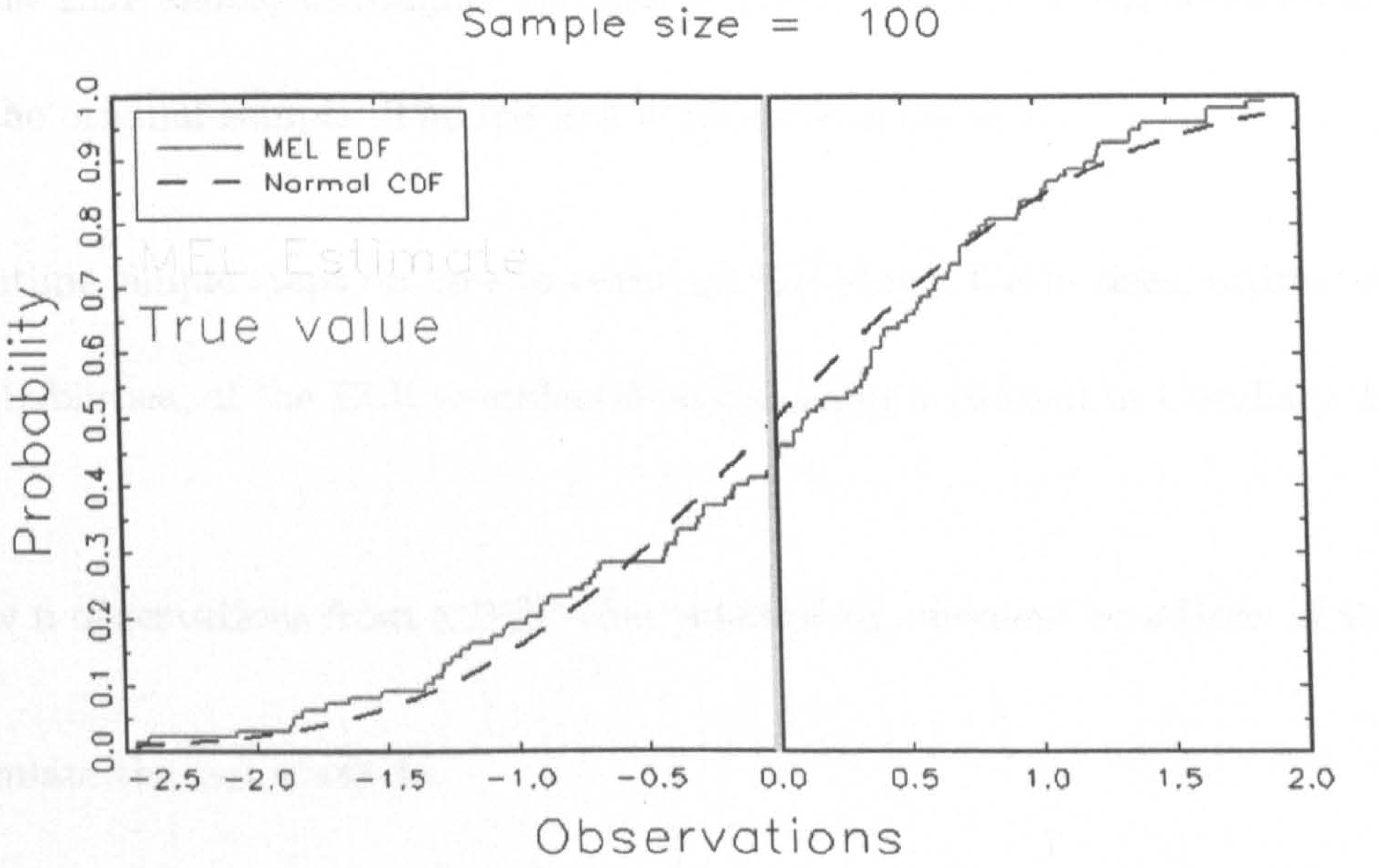

Figure 1.3: Asymptotic normality n=100

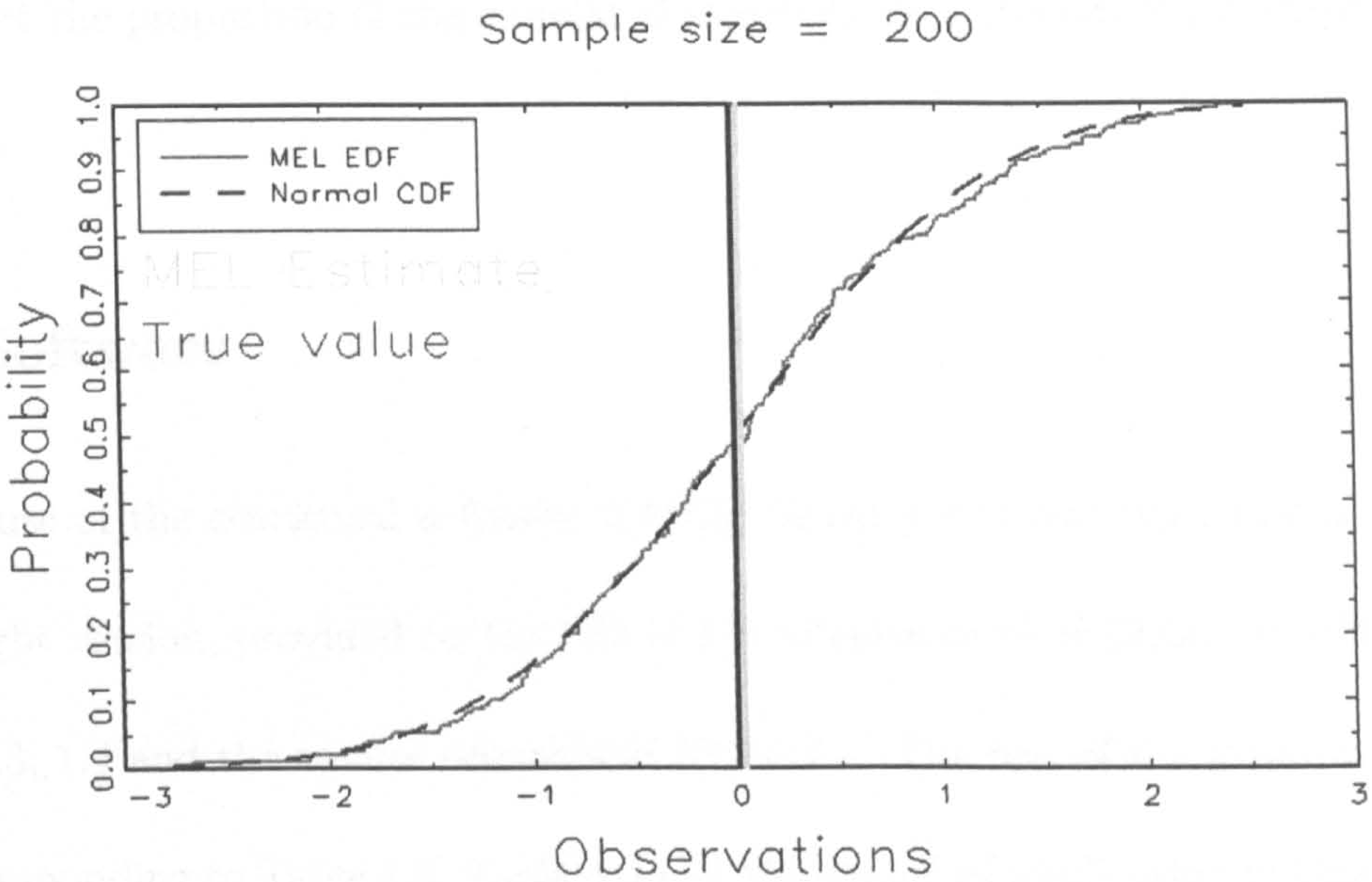

Figure 1.4: Asymptotic normality n=200

Note that the EDF should be roughly centered on the yellow line, which is the MEL estimate of θ from the original sample. The red line is the true value of θ.

We now outline simple steps on how to calculate the Monte Carlo sizes, estimates of Type I error probabilities, of the ELR overidentification statistic defined in Corollary 3.[12]

1. Draw n observations from a DGP that satisfies the moment equations of the model.

2. Calculate the test statistic.

3. Repeat steps 1 to 2 m times.

[12]Throughout the dissertation we use the terms Monte Carlo sizes, rejection probabilities/frequencies, and empirical sizes/levels interchangeably.

4. Report the proportion of the simulated statistics that exceeds the asymptotic critical value.

1.11 Software

We made use of the statistical software GAUSS to carry out our experiments. We used a GAUSS light version, provided on the CD of Mittelhammer *et al* (2000), to obtain Figures 1.1, 1.2, 1.3, 1.4 and the results reported in Table 2.1. The rest of the experiments, except those corresponding to Table 4.4, Table 4.5 and the empirical application in Chapter 4, were carried out with version 3.5. The already defined exceptions were calculated using GAUSS version 6.

Numerical analysts have devised several algorithms to simulate pseudo-random outcomes from the U(0,1) distribution. We focus on a particular algorithm known as the linear congruential generator.

The linear congruential rule generates a set of n values $y_1, y_2, ..., y_n$ in the (0,1) interval by forming the ratio of two integers I_t/w and reporting the fractional remainder y_t for $t = 1, ..., n$. The integer in the denominator, w, is known as the modulus. To form a sequence of pseudo-random numbers, I_t is changed for each number in the sequence. The numerator sequence begins from a starting value I_0, *i.e.* the seed. Subsequent integers are generated by the linear progression $I_t = \alpha + \beta I_{t-1}$ for $t = 1, ..., n$ and fixed integers α and β.

Owing to the finite numbers of integers that may be represented on a modern 32-bit computer, the linear congruential algorithm will eventually repeat the sequence for adequately large n. The period of an algorithm is an integer d such that $Y_t = Y_{t+d}$ for each t.

The GAUSS procedure rndu() generates pseudo-random outcomes for the $U(0,1)$ distribution. The seed is automatically reset when the GAUSS program is started and is updated each time the rndu() procedure is called. Under the default options, GAUSS uses the largest possible modulus for 32-bit computers ($w = 2^{31} - 1 = 2,147,483,647$), and the period d equals the modulus w (see Mittelhammer *et al*, 2000).

GAUSS provides commands and procedures that generate pseudo-random draws from many common parametric families. Most of the procedures are based on the probability integral transformation (see Mittelhammer, 1996). On the basis of Theorem 6.22 in Mittelhammer (1996), if X is a continuous random variable with distribution function $F(x)$, then $Y = F(X)$ is a $U(0,1)$ random variable. By the converse to the theorem (Mittelhammer, 1996, Theorem 6.23), the random variable $X_* = F^{-1}(Y)$ for $Y \sim U(0,1)$ has distribution function $F(x)$ if $F^{-1}(y)$ exists. Thus, given the inverse CDF, we can transform $U(0,1)$ outcomes to represent outcomes from distribution F. If the inverse CDF does not exist, then $x_* = \min_x \{x : F(x) \geq y\}$ for $U(0,1)$ outcome y represents an outcome from random variable X_* with distribution function F.[13] Additionally to rndu(), GAUSS provides the inverse standard normal CDF in the cdfni() procedure, the inverse Chi-square CDF in the cdfchii() procedure and the cdftci() procedure. The latter returns the inverse of the complement to the distribution function $G(x) = 1 - F(x)$ for the Student's (central) t-distribution.

[13] For further details see Mittelhammer *et al* (2000).

We use two algorithms throughout this thesis: a simultaneous and a sequential one. We adopt the algorithm which seems better suited to the application of interest. An algorithm could fail too frequently through non-convergence when applied to the particular problem or simply exceed a sensible time benchmark without yielding an output. "EqSolve", the simultaneous algorithm, is applied within the setting proposed by Qin and Lawless (1994), the Chi-Squared Moments Model and the Mean-Variance CAPM. The sequential algorithm, "constrained optimization", is used in the rest of the experiments. However, changes and additions were made to tailor the code to the particular task in hand. To deal with those estimation procedures that are new in the literature, EL-bootstrap as defined in Section 1.9.1, we constructed our own GAUSS programs. To generate chi-squared random variables whose higher moments were similar to their theoretical counterparts, within the Three-Moment CAPM, we used an acceptance-rejection algorithm.[14]

The general design of computer procedures follow closely those of Bruce Hansen[15] and the software of Mittelhammer *et al* (2000).

[14]We found that the third sample moments of a chi-squared random variable could differ from their theoretical counterparts by 70%. The differences between sample and theoretical fourth moments could be as large as 150%. These percentages correspond to a sample size n=100.

[15]Available from his webpage: www.ssc.wisc.edu/~bhansen/progs/progs_gmm.html.

Chapter 2

Numerical Properties

2.1 Introduction

The aim of this chapter is to analyse the computational aspects of EL and the adequacy of the asymptotic approximations of its estimators and test statistics.

As was discussed in Chapter 1, the solution for the MEL estimate of θ $-\widetilde{\theta}_{EL}-$ is generally not obtainable in closed form. This is because the Lagrange multiplier $-t(\theta)-$ of the EL function $-L_E(\theta)-$ is not a closed-form function of θ.[1] Thus numerical optimization techniques are most often required to obtain outcomes of $\widetilde{\theta}_{EL}$. Computer programs obtain MEL estimates of θ and the Lagrange multipliers, by simultaneous or sequential methods.

Under the simultaneous solution method, $t\left(\widetilde{\theta}_{EL}\right)$ and $\widetilde{\theta}_{EL}$ are simultaneously selected to minimize the first-order necessary conditions of the empirical log-likelihood function (see 1.20). MEL estimates may also be computed sequentially. That is, we fix a value of θ

[1] $L_E(\theta)$ was defined in (1.16).

and then find the optimal t under the MEL criterion. Then, the optimal value of θ can be determined conditional on the preceding value of t, and the sequential process continues until convergence is achieved.

To demonstrate the feasibility of both procedures we carry out a Monte Carlo simulation experiment and report the MEL solution values as well as the number of iterations and the estimation time required until convergence occurs.

A logical question to consider is: "How sensitive are our estimations to the initial values that we set in each algorithm?" One might also ask: "How large does the sample size, n, have to be to obtain an accurate approximation?" A practitioner may also want to know: "How long does it take to compute EL estimates?"
Even if there are no general answers to these questions, we provide some simulation evidence that addresses these issues.

We study sequential and simultaneous algorithms, the sensitivity of estimators to starting values and the estimation time within a model which incorporates information relating the first and second moments of a chi-squared random variable.

To evaluate the adequacy of the asymptotic approximations of EL estimators and test statistics:

(i) We examine the average confidence interval length (AVL) and the empirical coverage probability (ECV) of three methods of obtaining confidence intervals for θ.[2]

[2]We define AVL and ECV in Section 2.5.1.

(ii) We assess the finite-sample size properties of overidentification tests.

The first two methods of obtaining confidence intervals for θ are based on the EL. One obtains confidence intervals from the ELR statistic[3] and the $\chi^2_{(q)}$ approximation of Theorem 2. The other is based on the limiting distribution for $\widetilde{\theta}_{EL}$, given in Theorem 1, and the variance estimator following Theorem 1. The third method of obtaining confidence intervals for θ is based on the parametric likelihood ratio statistic with a $\chi^2_{(q)}$ as the approximating distribution. Our aim is to compare the AVL and ECV for each type of interval.

Qin and Lawless (1994) also report the AVL and ECV of the three already defined confidence intervals. They generate 1000 pseudo-random samples of sizes $n = \{30, 60\}$ from $N\left(\theta, \theta^2 + 1\right)$ for $\theta = \{0, 1\}$, and nominal (asymptotic) 90% and 95% confidence intervals. They find that the two empirical methods agree closely and that for smaller samples their coverage probability is less than the nominal 90% and 95%. The parametric approach based on the correct parametric likelihood yields intervals close to the nominal coverage probability. The AVL of the EL confidence intervals are always shorter than that of the parametric intervals.

Our experiments extend Qin and Lawless (1994) simulation evidence using a different framework. We use a setting that incorporates information relating the first and second moments of a $\chi^2_{(\theta)}$ random variable, where θ is the degrees of freedom.

We assess further properties of EL by investigating the behaviour of its overidentification test (refer to Section 1.6.2). We study the finite-sample size properties of this test statistic

[3]Refer to 1.34.

and compare them to those obtained through overidentification tests based on the GMM and the KLIC (refer to Sections 1.8.1 and 1.7.1; respectively). This comparison arises naturally because these tests have a $\chi^2_{(r-q)}$ asymptotic distribution under the null hypothesis, *i.e.* the overidentifying restrictions are valid. We adopt the models proposed by Qin and Lawless (1994) and Hall and Horowitz (1996).

Qin and Lawless (1994) do not assess the finite-sample size properties of the ELR overidentification test. Bravo (2000) complements their work and gives simulation evidence to assess the finite-sample sizes of this statistic. He draws observations from a random variable distributed as $N(1, 2)$ and reports that the ELR overidentification test is slightly oversized for $n = 50$ and that the rejection probabilities of the nominal values improve for $n = 100$. Here, we extend Bravo's (2000) work by considering other distributions than the normal to assess if the asymptotic result is a reasonable approximation to the finite-sample distribution of this test statistic. Studying a broader class of distributions is important because in practice there are well known examples of variables that are not normally distributed. For example, the non-normality of high frequency financial variables has been widely documented in financial literature (Mandelbrot, 1963; Harvey and Zhoe, 1993; Dacorogna *et al*, 1995). Because the density underlying most financial data is more peaked and heavy tailed than the normal distribution, many authors believe that other specifications like the Student's t, gamma, chi-square and mixtures of normal distributions may be more suitable for financial variables (see among others, Blattberg and Gonedes, 1974; Hamilton 1991 and McDonald and Xu, 1995). In response to this literature, our simulation experiments consider the following distributions: normal, chi-square, gamma and Student's t. We also examine the effects of varying the sample size. We use two sample sizes that are typically encountered in empirical work: $n = 50$ and $n = 100$.

The second model that we use to assess the finite-sample size properties of the ELR J-test is that proposed by Hall and Horowitz (1996). Hall and Horowitz examine the bootstrap within a simplified version of an asset pricing model. They generate two moment conditions by incorporating a utility function with constant relative risk aversion into a typical consumption problem. The result is an Euler equation which in combination with restrictions on the distribution of the gross growth rate of consumption yields the orthogonality conditions that characterize our Monte Carlo experiments. Bravo (2000) and Imbens *et al* (1998) use this model to assess the finite-sample behaviour of several overidentification tests, including that based on the ELR (Hall and Horowitz, 1996; do not study the ELR overidentification test). Bravo (2000), Imbens *et al* (1998) and Hall and Horowitz (1996) generate pseudo-random samples from a bivariate normal distribution. It is well known that the asymptotic properties of the ELR overidentifying statistic within this setting and under the DGP investigated by Hall and Horowitz (1996) are a poor guide to finite-sample behaviour (see Table 3 of Bravo, 2000; Table 2 of Imbens *et al*, 1998).

First, we replicate the Monte Carlo experiments of Bravo (2000) and Imbens *et al* (1998). Then, we complement their studies by investigating the extent to which the poor performance of the ELR overidentification test is extended to non-normal distributions. We draw pseudo-random samples from a bivariate chi-square distribution and a bivariate gamma distribution.

Finally, in response to Bravo's (2000) and Imbens' *et al* (1998) simulation experiments (and our findings, see Tables 2.7 and 2.9) –*i.e.* oversized tests– we use efficient bootstrap[4] critical values as an alternative to asymptotic ones. We present results for the overidentification tests based on the EL-bootstrap and the GMM-bootstrap (both tests are defined in Section 1.9.1). As far as we are aware, this is the first study that provides simulation evidence for the EL-bootstrap overidentification test. Moreover, for the first time we apply Brown and Newey's (2001) method of bootstrapping –efficient bootstrap– for the EL.[5]

The rest of the Chapter is structured as follows:

Section 2.2 lays out the moment equations, $E[g(x,\theta)] = 0$, that characterize the three models that we use as laboratories in our simulation experiments.

Computational aspects of EL are studied in Sections 2.3 and 2.4. We discuss and illustrate simultaneous and sequential algorithms in Section 2.3. Section 2.4 examines whether our estimations are sensitive to starting values, as well as the computation speed of our estimates. The adequacy of the asymptotic approximations of EL estimators and test statistics is assessed in Sections 2.5 and 2.6. We examine the ECV and AVL of confidence intervals based on empirical and parametric likelihoods in Section 2.5. Section 2.6 analyses the finite-sample properties of several overidentification tests. Finally, the size properties of the

[4]Efficient bootstrap resamples from the EL distribution, that imposes the moment restrictions, rather than the empirical distribution (refer to Section 1.9).

[5]Brown and Newey (2001) apply this bootstrapping technique for the GMM.

EL-bootstrap and GMM-bootstrap statistics are examined in Section 2.7. Conclusions are given in Section 2.8.

2.2 The Models

The following models have been widely examined in previous research. This is valuable in the sense that we are able to compare the results of our simulations to those well established in literature. Thus, these models represent a natural starting point to assess the validity of our computer programs.

2.2.1 The Chi-Squared Moments Model

The first two moments of $x_1, x_2, ...x_n \backsim \chi^2_{(\theta)}$ can be written as

$$E[x-\theta]=0, \tag{2.1}$$

$$E\left[x^2-2\theta-\theta^2\right]=0. \tag{2.2}$$

Equations (2.1) and (2.2) can be incorporated into our optimization procedures by noting that:

$$g(x,\theta)=\begin{pmatrix} x-\theta \\ x^2-2\theta-\theta^2 \end{pmatrix}. \tag{2.3}$$

2.2.2 The Qin and Lawless Model

Qin and Lawless (1994) consider a setting with first and second moments satisfying

$$E(x-\theta)=0, \tag{2.4}$$

$$E\left(x^2 - 2\theta^2 - 1\right) = 0. \tag{2.5}$$

Here,

$$g(x,\theta) = \begin{pmatrix} x - \theta \\ x^2 - 2\theta^2 - 1 \end{pmatrix}. \tag{2.6}$$

2.2.3 The Hall and Horowitz Model

The model used in Hall and Horowitz (1996) is defined by the following moment conditions:

$$E\left[\exp\{\mu - \theta(x+y) + 3y\} - 1\right] = 0, \tag{2.7}$$

$$E\left\{y\left[\exp\{\mu - \theta(x+y) + 3y\} - 1\right]\right\} = 0; \tag{2.8}$$

where θ is the parameter to be estimated, μ is a known normalization constant, and x and y are random variables.

Here,

$$g(x,y,\theta) = \begin{pmatrix} \exp\{\mu - \theta(x+y) + 3y\} - 1 \\ y\left[\exp\{\mu - \theta(x+y) + 3y\} - 1\right] \end{pmatrix}. \tag{2.9}$$

2.3 Simultaneous and Sequential Methods

The GAUSS' simultaneous procedure EqSolve is used to compute the MEL solution values for the Lagrange multiplier, t, and the parameters, θ; *i.e.* $t\left(\tilde{\theta}_{EL}\right)$ and $\tilde{\theta}_{EL}$. The purpose of EqSolve is to solve a system of nonlinear equations.[6] A starting value must be prespecified to initiate the algorithm. Some care is needed because the solution sought is one of many saddlepoints of the function $h(\theta,t) = \sum_i \log\{1 + t^{\tau} g(x_i,\theta)\}$ and, in particular, must satisfy

[6] A GAUSS alternative to EqSolve is NLSYS.

$1 + t^{\tau} g(x_i, \theta) \geq n^{-1}$ for each i. In some cases, simultaneous estimation of the parameters may not work well due to the scale or complexity of the estimation problem. As noted by Imbens *et al* (1998), the computation of solutions of constrained optimization problems can represent formidable numerical challenges. This can be because the fundamental method by which EL resolves the undetermined nature of the empirical moment conditions is to choose sample weights that ultimately transform the r moment conditions into a functionally dependent, lower rank ($q < r$) system of equations capable of being solved uniquely for the parameters. This could create instability in gradient-based constrained optimization algorithms regarding the representation of the feasible spaces and feasible directions for such problems. Moreover, attempting to solve the optimization problems in primal form is complicated by the dimensionality of the problem. There are as many sample weights as there are observations, and requires that explicit constrained optimization methods be used to enforce the moment conditions and the convexity properties of the sample weights (Mittelhammer *et al*, 2003).

In our GAUSS program, the sequential programming method used is constrained optimization. In this method the parameters are updated in a series of iterations beginning with the provided starting values. We first fix a value of θ, the starting value, and then find the optimal t under the MEL criterion. The optimal value of θ can be determined conditional on the preceding value of t, and the sequential process continues until convergence is achieved. The sequential solution method requires the calculation of a Hessian, various gradients and Jacobians. The Hessian may be very expensive to compute at every iteration. We use a quasi-Newton algorithm for updating the Hessian rather than computing it directly at each iteration. After several iterations the quasi-Newton algorithm should do nearly as well as Newton iteration with much less computation.

2.3.1 The Model

The model that we use in our Monte Carlo simulations is that characterized by the first and second moments of a chi-squared random variable; Equations (2.1) and (2.2).

2.3.2 The Data Generating Process

We generate a pseudorandom sample of size $n = 100$ from $\chi^2_{(1)}$. Hence, moment conditions (2.1) and (2.2) are satisfied; *i.e.* the null hypothesis, $E\left[g\left(x,\theta\right)\right] = 0$, is true. The starting value is set to $\underline{\theta}$=1. Results for our Monte Carlo experiment are reported in Table 2.1. We provide $\tilde{\theta}_{EL}$ and the number of iterations until convergence occurs.

2.3.3 Results

Due to the simplicity of this MEL problem, the simultaneous solution algorithm only requires a few (four) iterations and 11 hundredths of a second until convergence is achieved. The sequential algorithm is more computationally expensive in this case, with 1066 iterations and 115.7 hundredths of a second until convergence occurs. Mittelhammer *et al* (2000) suggest that sequential methods may be better suited to solve MEL problems with highly nonlinear constraints or several unknown parameters.

MEL estimator n=100 $E(x) = \theta$ and $E(x^2) = 2\theta + \theta^2$		
	Simultaneous	Sequential
$\tilde{\theta}_{EL}$	1.0184	1.0184
iterations	4	1066
time	11	115.7

$\tilde{\theta}_{EL}$ is the MEL estimator and n is the sample size
time is given in hundredths of a second

Table 2.1: Algorithms

2.4 Effects of Starting Values and Time

Numerical searches typically require starting values in order to succeed. We analyse the sensitivity of the ELR J-test to different starting values. Since obtaining EL estimators may be computationally intensive, we assess if our calculations are fast by computing the time involved in each estimation for different sample sizes and different starting values.

2.4.1 The Model

We use a framework characterized by Equations (2.1) and (2.2).

2.4.2 The Data Generating Process

We consider two random variables, namely $x \sim N(\theta, 2\ \theta)$ and $x \sim \chi^2_{(\theta)}$, and focus on $\theta = 1$. In both cases the null hypothesis of the ELR overidentification test, $E[g(x, \theta)] = 0$, holds.

2.4.3 Finite-Sample Size Properties of the ELR Overidentification Test

We report the empirical levels of the ELR J-test. Note that $\chi^2_{(1)}$ is the approximating distribution of the statistic. We set 5 different starting values: $\underline{\Theta}=\{0,.5,1,2,5\}$, and analyse 3 sample sizes: n=$\{20,100,500\}$. We are particularly interested in $\underline{\theta}$=5 because it is a poor starting value for both DGPs, whereas $\underline{\theta}$=0, $\underline{\theta}$=.5, $\underline{\theta}$=1 and $\underline{\theta}$=2 are relatively better choices. Results for 5000 replications are given in Tables 2.2 and 2.3. The time involved in each estimation is reported alongside the empirical levels. Time is given in hundreths of a second.

2.4.4 Results

Empirical Levels of ELR J-test
$E(x)=\theta$ and $E(x^2)=2\theta+\theta^2$

	$N(1,2)$ starting value				
Level	$\underline{\theta}$=0	$\underline{\theta}$=.5	$\underline{\theta}$=1	$\underline{\theta}$=2	$\underline{\theta}$=5
	n=20				
.10	.1526	.1566	.1540	.1542	.1490
.05	.0986	.0998	.1044	.0962	.0942
.01	.0424	.0434	.0496	.0428	.0340
(time)	815.6	659.4	639.1	671.8	15928.3
	n=100				
.10	.0906	.1112	.0976	.1024	.0916
.05	.0412	.0578	.0492	.0504	.0492
.01	.007	.0126	.0092	.0112	.0086
(time)	1889	1415.5	1218.7	1521.8	112076.3
	n=500				
.10	.0878	.0972	.0990	.0952	.0908
.05	.0420	.0494	.0490	.0458	.0428
.01	.0054	.0098	.0122	.0084	.0086
(time)	7415.6	5815.6	4336	6045.2	709714.9

Empirical levels refer to rejection frequencies as estimates of Type I error probabilities and n is the sample size.

Table 2.2: Effects of Starting Values - normal

Table 2.2 considers a DGP of the form $x \sim N(1,2)$. There is no evidence that choosing a poor starting value hurts the finite-sample size properties of the ELR overidentification test. Nevertheless, if an unfortunate starting value is set into the algorithm, the computation time increases dramatically.

Empirical Levels of ELR J-test
$E(x)=\theta$ and $E(x^2) = 2\theta + \theta^2$

	$\chi^2_{(1)}$ starting value				
Level	$\underline{\theta}$=0	$\underline{\theta}$=.5	$\underline{\theta}$=1	$\underline{\theta}$=2	$\underline{\theta}$=5
	n=20				
.10	.4574	.4556	.4584	.4752	.5556
.05	.4028	.4004	.4034	.4160	.4986
.01	.3528	.3240	.3204	.3358	.4196
(time)	1920.2	1463.9	1414.1	2173.3	41219.1
	n=100				
.10	.2182	.2320	.2318	.2540	.3282
.05	.1582	.1672	.1726	.1852	.2456
.01	.0932	.0914	.0928	.1086	.1360
(time)	3739.2	2067.3	2015.7	3125.1	570293.9
	n=500				
.10	.1270	.1400	.1106	.1344	.1692
.05	.0750	.0796	.0612	.0824	.1026
.01	.0252	.0246	.0172	.0290	.0348
(time)	15,290.1	7,102.7	5972.3	12377.3	7201413.8

Empirical levels refer to rejection frequencies as estimates of Type I error probabilities and n is the sample size.

Table 2.3: Effects of Starting Values - chi

Table 2.3 summarizes the results for a DGP of the form $x \sim \chi^2_{(1)}$. Our findings show that our estimations can be sensitive to starting values. For this DGP, the J-test over-rejects more often when poor initial values are considered (see $\underline{\theta}$=5). The latter is true regardless of the sample size. Also note that an unfortunate starting value has the additional consequence of increasing the computation time.

2.5 Confidence Intervals

In this section we consider the test of the hypothesis $H_0 : \theta = \theta_0$ *versus* $H_a : \theta \neq \theta_0$, and compare three methods of obtaining confidence intervals for θ.

The first two methods are based on the EL and the third one is based on the parametric likelihood ratio. The first method obtains confidence intervals from the ELR and the $\chi^2_{(q)}$ approximation of Theorem 2. We refer to these confidence intervals as CI_ELR. The second method for obtaining confidence intervals for θ is also based on the EL. This considers the limiting distribution for $\widetilde{\theta}_{EL}$ given in Theorem 1. We refer to these confidence intervals as NCI. The third method of obtaining confidence intervals for θ is based on the parametric likelihood ratio statistic, with a $\chi^2_{(q)}$ as the approximating distribution. We refer to these confidence intervals as CI_PLR.

We compare the ECV and AVL for each of the three types of intervals.

2.5.1 Empirical Coverage and Average Length

Ideally, a confidence interval should have exactly the coverage $(1-\alpha)$ for any sample size and any sampling distribution. This means a probability $(1-\alpha)$ to include in its limit the true value of the parameter. But non exact nonparametric confidence regions are quite common. As a result nonparametric confidence intervals are asymptotic, as indeed are most parametric confidence regions. Under mild regularity conditions, the coverage error for EL confidence intervals decreases to zero at a rate *1/n* as n approaches infinite (see Owen, 2001). This is the same rate that typically holds for confidence intervals based on parametric likelihoods but only if the model is true.

Our aim is to calculate the ECV by counting the percentage of times that θ is contained within each of the confidence intervals. For example, if the experiment is repeated 5000 times and if 4500 times θ is found in the interval, this would imply that the ECV equals 90%.

AVL refers to the average difference between the upper and lower ends of the confidence interval.

2.5.2 The Model

To illustrate the ECV and AVL of CI_ELR, NCI and CI_PLR; we use the model characterized by Equations (2.1) and (2.2).

2.5.3 The Data Generating Process

We consider four DGPs.

a) $x \sim N(1,2)$,

b) $x \sim N(3,6)$,

c) $x \sim \chi^2_{(5)}$,

d) $x \sim \chi^2_{(10)}$.

It is easy to see that the four DGPs satisfy (2.1) and (2.2).

2.5.4 Results

Tables 2.4 and 2.5 report results for $n = 50$ and $n = 100$, respectively. We use 5000 replications for each DGP. Note that the approximating distribution for the ELR and the parametric likelihood ratio is $\chi^2_{(1)}$.

From Table 2.4 we note that for the normal distribution, the parametric likelihood yields intervals with ECV close to its nominal counterpart. The two EL methods agree closely and for this sample size their coverage probability is less than the nominal 90%, 95% and 99%.

For variables distributed as $\chi^2_{(5)}$ and $\chi^2_{(10)}$, the three methods of obtaining confidence intervals yield similar results. For these specifications, the ECV is very close to its nominal counterpart.[7] This fact is somewhat surprising because the ELR relies on weaker assumptions than the parametric approach.

[7] Smaller sample sizes (n=20 and n=30) led to undercoverages similar to those reported by Qin and Lawless (1994) for normal distributed variables.

Average Length and Empirical Coverage $n = 50$ $E(x)=\theta$ and $E(x^2) = 2\theta + \theta^2$		90%		95%		99%	
		AVL	ECV	AVL	ECV	AVL	ECV
$N(1,2)$	CI_ELR	.4533	.8688	.5387	.9232	.6824	.9696
	NCI	.4449	.8690	.5299	.9282	.6951	.9778
	CI_PLR	.4724	.8956	.5649	.9526	.7530	.9884
$N(3,6)$	CI_ELR	.8424	.8750	.9911	.9204	1.2462	.9682
	NCI	.8316	.8750	.9832	.9184	1.2921	.9716
	CI_PLR	.8859	.9062	1.0610	.9454	1.4143	.9880
$\chi^2_{(5)}$	CI_ELR	1.3715	.9001	1.6422	.9502	2.1590	.9901
	NCI	1.3810	.9078	1.6410	.9548	2.1588	.9904
	CI_PLR	1.3840	.8966	1.6441	.9508	2.1612	.9900
$\chi^2_{(10)}$	CI_ELR	2.0040	.9004	2.3621	.9510	3.1117	.9917
	NCI	1.9802	.8992	2.3312	.9514	3.1039	.9884
	CI_PLR	2.1180	.9018	2.3626	.9504	3.1141	.9886

AVL is average length, ECV is empirical coverage. CI_ELR and NCI are confidence intervals based on EL and CI_PLR is that based on the parametric likelihood ratio. 90%, 95% and 99% are nominal coverages. n is the sample size.

Table 2.4: Confidence Intervals - n=50

Results for a larger sample size, $n = 100$, are summarized in Table 2.5. Here, the two EL confidence intervals have coverage probabilities close to the nominal 90%, 95% and 99%. For both distributions, the three methods of obtaining confidence intervals are very accurate. Moreover, the AVL is also similar for both the empirical and parametric likelihood procedures.

Average Length and Empirical Coverage
$n = 100$
$E(x)=\theta$ and $E(x^2) = 2\theta + \theta^2$

		90%		95%		99%	
		AVL	ECV	AVL	ECV	AVL	ECV
$N(1,2)$	CI_ELR	.3257	.8962	.3901	.9396	.5013	.9854
	NCI	.3213	.8914	.3823	.9360	.5028	.9890
	CI_PLR	.3316	.9050	.3959	.9490	.5232	.9934
$N(3,6)$	CI_ELR	.6124	.8860	.7314	.9424	.9250	.9790
	NCI	.6027	.8820	.7210	.9420	.9416	.9808
	CI_PLR	.6247	.8992	.7466	.9526	.9872	.9900
$\chi^2_{(5)}$	CI_ELR	.9690	.9071	1.1659	.9512	1.8328	.9900
	NCI	.9773	.9134	1.1648	.9582	1.8320	.9934
	CI_PLR	.9899	.9026	1.1662	.9486	1.8344	.9914
$\chi^2_{(10)}$	CI_ELR	1.4075	.9016	1.6902	.9518	2.2213	.9920
	NCI	1.3970	.9056	1.6836	.9488	2.2128	.9914
	CI_PLR	1.4123	.9044	1.6911	.9490	2.2300	.9922

AVL is average length, ECV is empirical coverage. CI_ELR and NCI are confidence intervals based on EL and CI_PLR is that based on the parametric likelihood ratio. 90%, 95% and 99% are nominal coverages. n is the sample size

Table 2.5: Confidence Intervals - n=100

2.6 Overidentification Tests

We assess the size properties of overidentifying restrictions statistics using the Qin and Lawless (1994) model, Equations (2.4) and (2.5), and the Hall and Horowitz (1996) model, Equations (2.7) and (2.8).

We consider four overidentification tests in what follows. These are based on: ELR, W_J; KLIC, J_{KLIC}; two-step GMM, J_{2GMM}; and continuously updated GMM, J_{CuGMM}.

These tests have as their null hypothesis that there is a unique value of θ consistent with $E\left[g\left(x,\theta\right)\right]=0$.

2.6.1 The Model

We first use the model characterized by Equations (2.4) and (2.5).

The Data Generating Process

We examine four DGPs:

(a) $x \sim \chi^2_{(1)}$.

(b) $x = \theta + z\left(\sqrt{\theta^2+1}\right)$,

where $z = \frac{y-1}{\sqrt{1}}$ and $y \sim \Gamma\left(1,1\right)$.

(c) $x = \theta + z\left(\sqrt{\theta^2+1}\right)$,

where $z = \frac{y-0}{\sqrt{\frac{5}{3}}}$ and $y \sim t\left(5\right)$.

(d) $x \sim N\left(0,1\right)$.

The four DGPs lead to unbiased estimating equations. This is true for (b) and (c) regardless of the value of θ. Assume the simplest case, $\theta = 0$.

Finite-Sample Size Properties

We report rejection frequencies, with particular interest being in cases where these probabilities are poorly approximated by the nominal size. Under the null hypothesis –validity of the overidentifying restrictions– W_j, J_{KLIC}, J_{2GMM} and J_{CuGMM} have an asymptotic $\chi^2_{(1)}$ distribution. Table 2.6 reports the rejection frequencies of the tests at the .10, .05 and .01 nominal critical values for $n = 50$ and Table 2.7 those for $n = 100$. We use 5000 replications.

Results

Empirical Levels of J-tests
$n = 50$
$E(x) = \theta$ and $E(x^2) = 2\theta^2 + 1$

	Levels	W_j	J_{KLIC}	J_{2GMM}	J_{CuGMM}
	.10	.2720	.2904	.2986	.3088
$\chi^2_{(1)}$	.05	.1996	.2392	.2558	.2724
	.01	.1511	.1771	.1930	.2112
	.10	.1934	.2584	.2552	.2726
$\Gamma(1,1)$	.05	.1318	.2058	.2084	.2332
	.01	.0732	.1401	.1398	.1704
	.10	.1876	.2316	.2362	.2535
$t(5)$	.05	.1218	.1762	.1892	.2088
	.01	.0510	.1060	.1084	.1384
	.10	.1326	.1470	.1548	.1662
$N(0,1)$	.05	.0711	.0956	.1016	.1128
	.01	.0230	.0420	.0474	.0546

Empirical levels refer to rejection frequencies as estimates of Type I error prob.
W_j, J_{2GMM}, J_{CuGMM} and J_{KLIC} are overidentification tests
based on ELR, two-step and continuously updated GMM,
and KLIC; respectively. n is the sample size.

Table 2.6: Finite-Sample Size Properties - QL n=50

First, we analyse the results for $\chi^2_{(1)}$ in Table 2.6. For this DGP, the four tests are oversized. In some cases the size distortion[8] is very pronounced, especially for the tests based on the KLIC and the GMM.

For the gamma and t distributions, the null distribution of the three tests is not well approximated by the asymptotic $\chi^2_{(1)}$. However, the ELR overidentification test is better than those tests based on the GMM and KLIC.

It is for the normal distribution that the tests have the closest Monte Carlo sizes to their nominal counterparts but the J-tests based on the GMM and KLIC are still quite oversized.

In general, for $n = 50$ the ELR overidentification test performs better than the other tests.

The effects of increasing the sample size are reported in Table 2.7. The adequacy of the asymptotic approximation, $\chi^2_{(1)}$, is better for a larger sample size. However, even if the rejection frequencies at the three nominal critical values improve; the tests are still oversized. Note that W_j has moderate size distortions for $\Gamma(1,1)$ and $N(0,1)$.

[8]When we refer to distortions or size distortions we mean that the estimated sizes are different to the nominal sizes.

Empirical Levels of J-tests
$n = 100$
$E(x) = \theta$ and $E(x^2) = 2\theta^2 + 1$

	Levels	W_j	J_{KLIC}	J_{2GMM}	J_{CuGMM}
	.10	.2210	.2264	.2326	.2372
$\chi^2_{(1)}$	.05	.1540	.1690	.1906	.1994
	.01	.0910	.1042	.1356	.1412
	.10	.1312	.2130	.2058	.2037
$\Gamma(1,1)$	.05	.0748	.1616	.1540	.1605
	.01	.0326	.0938	.1004	.1046
	.10	.1518	.2044	.1894	.1899
$t(5)$	.05	.0878	.1406	.1386	.1419
	.01	.0272	.0652	.0784	.0794
	.10	.1126	.1418	.1416	.1204
$N(0,1)$	.05	.0578	.0846	.0858	.0766
	.01	.0130	.0270	.0328	.0324

Empirical levels refer to rejection frequencies as estimates of Type I error prob.
W_j, J_{2GMM}, J_{CuGMM} and J_{KLIC} are overidentification tests
based on ELR, two-step and continuously updated GMM,
and KLIC; respectively. n is the sample size.

Table 2.7: Finite-Sample Size Properties - QL n=100

2.6.2 The Model

This Section assesses the finite-sample size properties of W_j, J_{KLIC}, J_{2GMM} and J_{CuGMM} using the Hall and Horowitz (1996) model, Equations (2.7) and (2.8). Hall and Horowitz (1996) consider, among others, the following DGPs:

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .2^2 & 0 \\ 0 & .2^2 \end{pmatrix}\right), \tag{2.10}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .4^2 & 0 \\ 0 & .4^2 \end{pmatrix}\right). \tag{2.11}$$

(2.10) and (2.11) are nested in (2.7) and (2.8) if a value of μ is obtained from

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\exp\left[\mu - \theta\left(x+y\right) + 3y\right] - 1\right) f_{xy} dxdy = 0, \tag{2.12}$$

where $f_{x,y}$ is the bivariate normal probability distribution function.

The corresponding value is

$$\mu = \frac{1}{2}\left[2\mu_y\left(\theta-3\right) - \sigma_y^2\left(\theta-3\right) + \theta\left(2\mu_x - \sigma_x^2\theta\right)\right]. \tag{2.13}$$

Hall and Horowitz (1996) assume $\theta = 3$. Substituting $\theta = 3$ into (2.13) yields

$$\mu = -.72 \quad \text{for} \quad \begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .4^2 & 0 \\ 0 & .4^2 \end{pmatrix}\right),$$

$$\mu = -.18 \quad \text{for} \quad \begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .2^2 & 0 \\ 0 & .2^2 \end{pmatrix}\right).$$

The Data Generating Process

To extend existing simulation evidence based on the Hall and Horowitz (1996) model, we consider two additional DGPs to those studied by Bravo (2000), Imbens *et al* (1998) and Hall and Horowitz (1996).

- Assume that x and y are independent and that

$$x \sim \Gamma\left(\alpha_x, \beta_x\right) \text{ and } y \sim \Gamma\left(\alpha_y, \beta_y\right). \tag{2.14}$$

The DGPs in (2.14) are nested in (2.7) and (2.8) if a value of μ is obtained from

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\exp\left[\mu-\theta\left(x+y\right)+3y\right]-1\right)f_{xy}dxdy=0, \tag{2.15}$$

where $f_{x,y}$ is the bivariate gamma probability distribution function.

This value is

$$\mu=\ln\left[\beta_x^{\alpha_x}\beta_y^{\alpha_y}\left(\frac{1}{\beta_x}+\theta\right)^{\alpha_x}\left(\frac{1}{\beta_y}+\theta-3\right)^{\alpha_y}\right]. \tag{2.16}$$

Let $\alpha_x=1$, $\beta_x=1$, $\alpha_y=1$, β_y=1 and $\theta=3$. Substituting these specific values into (2.16) yields $\mu=\ln\left(4\right)$.

- Assume that x and y are independent and that

$$x\sim\chi^2_{(u_x)} \text{ and } y\sim\chi^2_{(u_y)}. \tag{2.17}$$

The DGPs in (2.17) are nested in (2.7) and (2.8) if a value of μ is obtained from

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\exp\left[\mu-\theta\left(x+y\right)+3y\right]-1\right)f_{xy}dxdy=0, \tag{2.18}$$

where $f_{x,y}$ is the bivariate chi-square probability distribution function.

The corresponding value is

$$\mu=\ln\left[\left(2\theta-5\right)^{u_x/2}\left(2\theta+1\right)^{u_x/2}\right]. \tag{2.19}$$

Let $u_x=2$, $u_y=2$ and $\theta=3$. Substituting these specific values into (2.19) yields $\mu=\ln\left(7\right)$.

Finite-Sample Size Properties

For 5000 replications we register the proportion of the simulated statistics that exceeds the asymptotic critical value. The results in Table 2.8 are for $n = 50$ and those in Table 2.9 are for $n = 100$.

Results

Empirical Levels of J-tests

$n = 50$

$E\left[\exp\{\mu - \theta(x+y) + 3y\} - 1\right] = 0$

$E\left[y\left(\exp\{\mu - \theta(x+y) + 3y\} - 1\right)\right] = 0$

	Levels	W_j	J_{2GMM}	J_{CuGMM}
$N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}.2^2 & 0\\0 & .2^2\end{pmatrix}\right)$, $\mu = -.18$	.10	.1654	.1136	.1211
	.05	.1025	.0626	.0726
	.01	.0372	.0206	.0259
$N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}.4^2 & 0\\0 & .4^2\end{pmatrix}\right)$, $\mu = -.72$	.10	.2332	.1750	.1271
	.05	.1540	.1240	.0749
	.01	.0650	.0650	.0325
$x \sim \chi^2_{(2)}$, $y \sim \chi^2_{(2)}$, $\mu = \ln(7)$	.10	.1530	.1302	.0922
	.05	.0911	.0776	.0389
	.01	.0306	.0294	.0089
$x \sim \Gamma(1,1)$, $y \sim \Gamma(1,1)$, $\mu = \ln(4)$	.10	.1437	.1066	.1071
	.05	.0841	.0542	.0462
	.01	.0256	.0122	.0048

Empirical levels refer to estimates of Type I error probabilities.
W_j, J_{2GMM}, J_{CuGMM} are overidentification tests based on ELR, two-step and continuously updated GMM. n is the sample size.

Table 2.8: Finite-Sample Size Properties - HH n=50

Examination of Table 2.8 reveals that the ELR overidentification is badly oversized for every DGP and for all the nominal critical values that we investigate. The worst size distortions occur with the normal distribution. The tests based on the GMM have better size properties. This is especially true for J_{CuGMM}, regardless of the DGP. There is an improvement in the size properties of all the overidentification tests as the sample size increases (compare Table 2.8 to Table 2.9). However, the ELR overidentification test remains with pronounced size distortions and the normally distributed variables lead to the worst discrepancies between empirical and nominal sizes.

The results reported inside parenthesis, in Table 2.9, are from Imbens *et al* (1998). Note that our results agree in a large extent with theirs, *e.g.* the size properties of the ELR overidentification test are very poor for $n = 100$. Our simulations show that the size distortions for W_j are still large for other DGPs than those involving normality, although in a lower extent.

The usefulness of EL motivated us to look for accurate inference procedures for EL estimators. Some of our findings indicate that asymptotic approximations can be poor for the ELR J-test. It is on the improvement of inference methods for EL that we focus on next.

Empirical Levels of J-tests
$n = 100$
$E\left[\exp\{\mu - \theta(x+y) + 3y\} - 1\right] = 0$
$E\left[y\left(\exp\{\mu - \theta(x+y) + 3y\} - 1\right)\right] = 0$

	Levels	W_j	J_{2GMM}	J_{CuGMM}
$N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}.2^2 & 0\\0 & .2^2\end{pmatrix}\right)$ $\mu = -.18$	.10	.1660	.1102	.1102
	.05	.1012	.0578	.0582
	.01	.0372	.0140	.0168
$N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}.4^2 & 0\\0 & .4^2\end{pmatrix}\right)$ $\mu = -.72$	.10	.2037 (.190)	.1722 (.178)	.1299 (.136)
	.05	.1221 (.125)	.1170 (.129)	.0744 (.076)
	.01	.0536 (.057)	.0562 (.073)	.0261 (.026)
$x \sim \chi^2_{(2)}$ $y \sim \chi^2_{(2)}$ $\mu = \ln(7)$	.10	.1237	.1236	.0972
	.05	.0711	.0698	.0394
	.01	.0191	.0210	.0053
$x \sim \Gamma(1,1)$ $y \sim \Gamma(1,1)$ $\mu = \ln(4)$	.10	.1337	.1054	.1027
	.05	.0661	.0572	.0527
	.01	.0134	.0124	.0089

Empirical levels refer to rejection frequencies as estimates of Type I error probabilities. W_j, J_{2GMM}, J_{CuGMM} are overidentification tests based on ELR, two-step and continuously updated GMM. Values inside parenthesis are Imbens et al (1998). n is the sample size.

Table 2.9: Finite-Sample Size Properties - HH n=100

2.7 EL-bootstrap

Owen (2001) suggests some methods in which EL can be combined with other approaches; *e.g.* EL and bootstrap. Bootstrapping provides one approach to improved inference. Here, we concentrate on a method of bootstrapping for EL based on resampling from the EL distribution, that incorporates the moment restrictions, rather than the empirical distribution. Below we investigate numerically whether using the EL-bootstrap yields an improvement to the usual asymptotic approximation.

2.7.1 The Model

We employ the Qin and Lawless (1994) model, given in Equations (2.4) and (2.5) ; and the Hall and Horowitz (1996) model, characterized by Equations (2.7) and (2.8).

2.7.2 The Data Generating Process

We concentrate on the specifications in which we found the poorest size properties for the ELR overidentification test for $n = 100$ (refer to Tables 2.7 and 2.9). These are the DGPs given in a) and c) in Section 2.6.1 for the Qin and Lawless (1994) model. The DGPs specified in (2.10) and (2.11) correspond to the Hall and Horowitz (1996) model .

EL-bootstrap critical values are based on 1000 replications of the bootstrap sampling. The Monte Carlo experiment is replicated 5000 times. The results are shown in Tables 2.10 and 2.11. Note that for EL-bootstrap and GMM-bootstrap the rejection probabilities denote the proportion of the simulated data test statistics that exceeds the efficient bootstrap critical values. We refer to the EL-bootstrap overidentification test as W_J^b and to the GMM-bootstrap overidentification test as J_{2GMM}^b.

2.7.3 Results

For the first distribution in Table 2.10, $t(5)$, there is a size improvement when efficient bootstrap critical values are used for W_j (compare columns 3 and 4). This is especially true for $\alpha = .10$. However, J^b_{2GMM} remains quite oversized.

The empirical levels for $\chi^2_{(1)}$ also improve for W_j when efficient bootstrap critical values are used, whereas the GMM-bootstrap is not relatively better than the GMM.

Empirical Levels of J-tests
Bootstrap Critical Values
$n = 100$
$E(x) = \theta$ and $E(x^2) = 2\theta^2 + 1$

	Levels	W_j	W_j^B	J_{2GMM}	J^B_{2GMM}
	.10	.1518	.1260	.1894	.1750
$t(5)$	.05	.0878	.0770	.1386	.1353
	.01	.0270	.0220	.0784	.0662
	.10	.2210	.1630	.2488	.2521
$\chi^2_{(1)}$	.05	.1540	.1210	.2002	.2068
	.01	.0910	.0900	.1426	.1482

Empirical levels refer to rejection frequencies as estimates of Type I error prob. W_j, W_j^B, J_{2GMM}, J^B_{2GMM} are J- tests based on ELR, EL-bootstrap, two-step GMM and GMM-bootstrap. n is the sample size.

Table 2.10: Bootstrap Critical Values - QL

Results for the Hall and Horowitz (1996) model are reported in Table 2.11. We analyse the first DGP and compare columns 3 and 4. The size properties of the W_j^b are better than those of W_j. Moreover, the size distortions have been almost removed. An improvement is also observed for J_{2GMM}, compare columns 5 and 6, but this test is still oversized.

Consider the second DGP. Although W_j^b is slightly undersized, its size properties are better than those of W_j. By comparing both columns of the GMM statistic, columns 5 and 6, we find that the bootstrap is not relatively better for this specification. Here, the empirical levels of the tests have moderate size distortions for both critical values.

Empirical Levels of J-tests
Bootstrap Critical Values
$n = 100$
$E\left[\exp\{\mu - \theta\,(x+y) + 3y\} - 1\right] = 0$
$E\left[y\,(\exp\{\mu - \theta\,(x+y) + 3y\} - 1)\right] = 0$

	Levels	W_j	W_j^B	J_{2GMM}	J_{2GMM}^B
$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .4^2 & 0 \\ 0 & .4^2 \end{pmatrix}\right)$	.10	.2037	.1028	.1722	.1427
	.05	.1221	.0498	.1170	.0861
	.01	.0436	.0178	.0562	.0230
$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .2^2 & 0 \\ 0 & .2^2 \end{pmatrix}\right)$	.10	.1660	.0910	.1102	.1126
	.05	.1012	.0418	.0578	.0540
	.01	.0372	.0086	.0140	.0078

Empirical levels refer to rejection frequencies as estimates of Type I error prob. W_j, W_j^B, J_{2GMM}, J_{2GMM}^B are overidentification tests based on ELR, EL-bootstrap, two-step GMM and GMM-bootstrap. n is the sample size.

Table 2.11: Bootstrap Critical Values - HH

2.8 Conclusions

This Chapter has examined the computational aspects of EL and the adequacy of the asymptotic approximations of its estimators and test statistics.

We first described and illustrated simultaneous and sequential methods and found that both approaches led to identical estimates. However, the simultaneous solution algorithm requires less iterations to converge than those required by the sequential algorithm.

We have examined the sensitivity of our estimations to different starting values. Our findings show that for $N(0,1)$ variables, the finite-sample size properties of W_j are insensitive to starting values. Whereas for $\chi^2_{(1)}$ variables, an unfortunate initial value can lead to large discrepancies between empirical and nominal sizes. Calculations based on poor starting values and large sample sizes led to a dramatic increase in the time that our iterations required to converge.

To assess the adequacy of the asymptotic approximation of EL estimators and statistics, we first examined the ECV and AVL of three methods of obtaining confidence intervals. Our findings show that for a "small" sample size, $n = 100$, methods based on the EL have similar ECV and AVL to those obtained through the parametric likelihood ratio. These findings complement existing simulation evidence by exploring a new setting that examines the ECV and AVL of confidence intervals based on $\widetilde{\theta}_{EL}$.

We also assessed the adequacy of the asymptotic approximation of the ELR overidentification test within the Qin and Lawless (1994) and Hall and Horowitz (1996) models. We extended the existing simulation evidence based on these models by using distributions that have not been studied in the past. For the Qin and Lawless (1994) model we examined four distributions: chi-square, gamma, t and normal. We reported that all the tests were oversized for every DGP and for all the critical values. However, the ELR test had better size properties than those tests based on the GMM and KLIC. Increasing the sample size led to a reduction in rejection frequencies. For the Hall and Horowitz (1996) model we drew observations from random variables which are distributed as normal, chi-square and gamma. We found large size distortions for the ELR overidentification test for every DGP and for all critical values.

For the Qin and Lawless (1994) model the asymptotic approximation for W_j was especially poor for the t and chi-squared distributed variables. Whereas for the Hall and Horowitz (1996) model the normal variables led to the largest size distortions. In response to these findings, we introduced the EL-bootstrap and the GMM-bootstrap overidentification tests. The size properties of W_j improved by considering efficient bootstrap critical values within both models. Moreover, the size distortions were almost removed in the Hall and Horowitz setting (1996).

Chapter 3

CAPM and Overidentifying Restrictions Tests

3.1 Introduction

One of the most interesting areas of finance is that related to asset pricing theory. Asset pricing theory aims to explain why some assets pay higher average returns than others. Common sense suggests that risky investments such as the stock market will generally yield higher returns than investments free of risk. However, it was only with the development of asset pricing models that economists were able to quantify risk and the reward for bearing it.

Asset pricing models have sometimes been proved wrong when tested. Nevertheless these might only be describing how the world should work and not how it does work. If the world does not obey a model's predictions, some skeptics might think that the model needs improvement. However, it can also be argued that the world is "wrong" and that some assets are "mispriced". The latter view accounts for much of the popularity and practical application of asset pricing models.

The CAPM, based on the work of Markowitz (1959) and extended by Sharpe (1964) and Lintner (1965), was the first and probably most widely used model in asset pricing. The Mean-Variance CAPM states that the expected return of an asset is linearly related to the covariance of its return with the return of the market portfolio. A voluminous literature presenting empirical evidence on the Mean-Variance CAPM has evolved since its development. The early evidence was mostly positive (Black *et al*, 1972; Fama and MacBeth, 1973; Blume and Friend, 1973; reported evidence consistent with the model). In the late 1970s less favourable empirical results for the Mean-Variance CAPM came out (see Basu, 1977; Banz, 1981; Fama and French, 1992; De Bondt and Thaler, 1985; Jegadeesh and Titman, 1995). There is still controversy over how these discrepancies must be interpreted. Yet despite growing criticism, the CAPM remains widely used in finance.

Seeking a refinement and/or uniformity in results, some authors proposed extensions of the Mean-Variance framework to incorporate higher moments, while others advised alternative estimation procedures.

The inclusion of higher moments allows the expected return of an asset to be related not only to the covariance but to the co-skewness and co-kurtosis of its return with the return of the market portfolio. Arditti (1967), Jean (1971,1973), Ingersoll (1975), Kraus and

Litzenberger (1976), Friend and Westerfield (1980), Sears and Wei (1985, 1988), Lim (1989) and Homaifar and Graddy (1990) incorporate higher moments into the CAPM. However, empirical results on higher order frameworks –Three-Moment CAPM– show that the evidence is still contradictory, as in the Mean-Variance framework. Two of the most famous papers, Friend and Westerfield (1980) and Kraus and Litzenberger (1976), lead to different conclusions. The first reports significant coefficients on beta[1] and co-skewness, while the latter does not.

In respect of the inference procedures, at the beginning these were mainly carried out using Ordinary Least Squares (OLS) and Maximum Likelihood (ML). However, when deviations from the assumptions that returns are jointly normal and independent through time were accounted for, methods which accommodate non-normality, heteroscedasticity and temporal dependence of returns are to be preferred. Since the development of the GMM by Hansen (1982), this has dominated most of the literature. Within the GMM framework, the distribution of returns is not specified. It can be both serially dependent and conditionally heteroscedastic, the only assumption necessary being that excess asset returns are stationary and ergodic with finite fourth moments (Campbell *et al*, 1997). The GMM gained even more popularity as several published writings advocated a moment equations view of asset pricing theory and the associated empirical procedures. Moreover, when empirical work on extensions of the Mean-Variance framework emerged, it was argued that the GMM was the only appropriate method to test the validity of the Three-Moment CAPM (Lim, 1989). The

[1]Beta is defined as the ratio of the covariance of the return of a risky asset with the return of the market portfolio, and the variance of the market portfolio.

main argument is centered on the assertion that there is no obvious multivariate distribution of returns that also exhibits skewness. Cochrane (2001) discusses several advantages that GMM has over the ML approach within the context of CAPM. He argues that the GMM can handle nonlinearity, especially including conditioning information in an easier way than the ML does.

The debate in respect of which approach –OLS, ML or GMM– is the appropriate estimating method to employ in asset pricing models is practically over. Nowadays, most empirical work uses the GMM.

This chapter focuses on a new debate. It is devoted to an alternative estimation procedure to the GMM: EL. The ELR, as defined in (1.34), is a nonparametric analogue of likelihood estimation. It possesses an asymptotic variance that is the same as for the efficient GMM, thus it is asymptotically efficient. The overidentification test based on the ELR (see Section 1.6.2) is similar to that based on the GMM (see Section 1.8.1). They are asymptotically first-order equivalent and have the same interpretation. Both tests are distribution-free and their general setting is moment-condition models. But despite all the appealing properties of EL statistics, it has had limited diffusion in the area of asset pricing.

This chapter investigates the finite-sample properties, size and power, of moment restrictions tests based on GMM and EL within the Mean-Variance and Three-Moment CAPM through simulation evidence.

The finite-sample properties of GMM overidentification tests using a Two-Moment framework have been widely studied in the past (see, among others: Vorkink, 2003; Dahlquist and Soderlind, 1997; Hansen *et al,* 1996; Neely, 1995; Kocherlakota, 1990; Tauchen, 1986). However, the finite-sample size properties of the ELR overidentifying restrictions statistic remain practically unexplored in this context and the power properties more generally[2]. We are not aware of any study that uses simulations to assess the finite-sample properties of overidentifying restrictions statistics, including those based on the GMM, within a Three-Moment setting. Investigating the asymptotic efficiency of moment restrictions tests based on the GMM and EL should give further insight into the relative advantages of one approach over the other. The comparison arises automatically due to the fact that EL implements the same set of orthogonality conditions as the GMM.

There are several contributions to existing literature arising from this chapter:

1. The finite-sample properties of the ELR overidentification test are compared to those of GMM in a widely used empirical framework. We believe that it is important to assess the ability of these tests within useful empirical settings.

2. We provide simulation evidence to assess the power properties of the ELR J-test. At present, little is known about these properties.

[2]One of the few studies that compares the asymptotic optimality of EL for testing moment restrictions to tests based on two-step, ten-step and continuously updating versions of the GMM is that of Kitamura (2001). Kitamura's (2001) experimental design follows Hall and Horowitz's (1996) simulation study (refer to Chapter 2). Kitamura's (2001) experiments show that every method has moderate size distortions. After computing size-corrected critical values, Kitamura (2001) compares the power properties of the overidentification tests in 32 different experiments. The distribution under the alternative is altered by varying the parameters (mean and variance) in the simulations. His findings show that EL had the greatest power 22 times, two-step updating did this 5 times, 10-step updating 7 times and continuous updating never had the greatest power. EL's power ranking was best at hypotheses farther from the null. When any of the simulated methods achieved power over 80%, EL had the greatest power.

3. We use a Three-Moment framework, for the first time, to test the finite-sample properties of overidentifying restrictions statistics.

4. We assess power under the alternative hypothesis that the Mean-Variance CAPM is valid. We are not aware of any other study which assesses the power properties of tests of overidentifying restrictions using this interpretation.[3]

The rest of the chapter is organized as follows: In Section 3.2 the main theory underlying the CAPM is presented. Technical references are provided for further insight into the theoretical account. The presentation allows the derivation of the Mean-Variance CAPM and the Three-Moment CAPM as particular cases of the general model.

Section 3.3 focuses on the Mean-Variance CAPM. We examine the finite-sample properties of tests of overidentifying restrictions based on the ELR, W_j; two-step GMM, J_{2GMM}; and continuously updated GMM, J_{CuGMM}. To study the size properties of these tests, we use Monte Carlo techniques to simulate their finite-sample distribution. We consider two DGPs that are nested in the Mean-Variance framework: a DGP based on Mexican information and a linear market model. We also assess the power properties of overidentification tests. After an adjustment to make coverage 95%, the power is compared in simulations that vary the means and variances through the null. In total, we carry out 96 experiments.

Section 3.4 studies the size and power properties of overidentification tests using the Three-Moment CAPM. We consider a quadratic market model to simulate data consistent with

[3]Refer to Magdalinos and Symeonides (1996) for a discussion of different interpretations of the tests of overidentifying restrictions.

this setting and report rejection frequencies. We equalize the size of the different tests and examine their power using two different experiments. The first experiment varies the mean of the error term through the null. The second experiment assesses power under the alternative hypothesis that the Mean-Variance CAPM is valid.

Section 3.5 concludes. Proofs are provided in the Appendices.

3.2 Theoretical Background

The aim of this section is to derive the CAPM. This is important because by doing this we will get further insight into the assumptions and implications of the model.

The CAPM is based on theories related to utility, arbitrage, portfolio formation and efficient markets. Mean-Variance analysis offers a basis for the derivation of the model. Cochrane (2001), Copeland and Weston (1998), Campbell *et al* (1997) and Ingersoll (1987) give a complete background and a detailed description of the Mean-Variance CAPM. Homaifar and Graddy (1990), Lim (1989), Sears and Wei (1985), Friend and Westerfield (1980), Kraus and Litzenberger (1976), Ingersoll (1975), Jean (1973, 1971) and Arditti (1967) concentrate on the Three-Moment CAPM.

In general terms, the main problem of the CAPM can be stated as that of an investor with a specified utility function facing an investment environment with a riskless asset and N risky assets. Her aim is to maximize her utility by combining the risky assets and the riskless one in an optimal way. This maximization leads to the expected return of the risky asset being expressed in terms of its relationship with the market.

We follow Hwang and Satchell (1999) in the derivation of the CAPM.

There is a representative investor and all returns are in units of period one consumption. There is a riskless asset whose return is R_f and N risky assets whose i^{th} return is represented as R_i. Investment proportions on the riskless asset and N risky assets are x_0 and x_i $(i = 1, ..., N)$, respectively; where:

$$x_0 + \sum_i x_i = 1.$$

For the investor, the initial investment is one and the end of period wealth is represented as ω. Hence, her end of period wealth is

$$\omega = x_0 (1 + R_f) + \sum_i x_i (1 + R_i).$$

Consider a portfolio composed of combinations of the risky assets and the riskless one. The return of the portfolio is

$$R_P = x_0 R_f + \sum_i x_i R_i.$$

It is sensible to argue that the expected return on a security should be positively related to its risk. That is, individuals will hold risky securities only if its expected return compensates for their risk. According to Sharpe (1964), every investment carries two distinct risks. The systematic risk, which cannot be diversified away, and the unsystematic risk, which is specific to individual securities. Since the latter can be eliminated through appropriate diversification, the expected return hinges not on the asset's variance, skewness and kurtosis

–which are common measures of dispersion– but on the covariances, co-skewnesses and co-kurtosis of the returns. The systematic risk measures are given by beta, systematic skewness and systematic kurtosis[4], *i.e.:*

$$\beta_{iP} = \frac{E\left[\left(R_i - E\left(R_i\right)\right)\left(R_P - E\left(R_P\right)\right)\right]}{E\left[\left(R_P - E\left(R_P\right)\right)^2\right]}, \tag{3.1}$$

$$\gamma_{iP} = \frac{E\left[\left(R_i - E\left(R_i\right)\right)\left(R_P - E\left(R_P\right)\right)^2\right]}{E\left[\left(R_P - E\left(R_P\right)\right)^3\right]}, \tag{3.2}$$

$$\vartheta_{iP} = \frac{E\left[\left(R_i - E\left(R_i\right)\right)\left(R_P - E\left(R_P\right)\right)^3\right]}{E\left[\left(R_P - E\left(R_P\right)\right)^4\right]}. \tag{3.3}$$

To link the systematic risk measures to the investor, information about the investor's preferences must be incorporated. The investor's expected utility is a function of the expected value of end of period wealth and higher moments: variance, skewness and kurtosis. The standard assumption is that preferences induce the favouring of higher means, smaller variances, higher skewness and smaller kurtosis. The investor is concerned as to the proportions to allocate to the riskless and risky assets and be compensated for bearing risk. Loosely put, the investor will maximize her utility, which depends on her wealth and hence on the combination of risky and riskless assets, by obtaining the optimal proportions of assets to allocate into her portfolio.

At this point it is useful to establish the relationship among the moments of the end of period wealth $-E(\omega)$, $\sigma(\omega)^2$, $\gamma(\omega)^3$ and $\vartheta(\omega)^4-$ and the measures of systematic risk

[4] The terms co-skewness and systematic skewness as well as co-kurtosis and systematic kurtosis are interchangeably applied.

$-\beta_{iP}$, γ_{iP} and ϑ_{iP}. After some algebraic manipulations (see Appendix 1) we obtain

$$\sigma(\omega) = \sum_i x_i \beta_{iP} \sigma(R_P), \tag{3.4}$$

$$\gamma(\omega) = \sum_i x_i \gamma_{iP} \gamma(R_P), \tag{3.5}$$

$$\vartheta(\omega) = \sum_i x_i \vartheta_{iP} \vartheta(R_P), \tag{3.6}$$

where

$$\sigma(z) = \left[E(z - E(z))^2\right]^{1/2},$$
$$\gamma(z) = \left[E(z - E(z))^3\right]^{1/3},$$
$$\vartheta(z) = \left[E(z - E(z))^4\right]^{1/4},$$

and z is a random variable.

We now define a constrained optimization problem

$$Max\ E[U(\omega)] = f(E(\omega), \sigma(\omega), \gamma(\omega), \vartheta(\omega)),$$

subject to

$$x_0 + \sum_i x_i = 1.$$

This optimization may be solved though Lagrange Multipliers. Let

$$\mathfrak{L} = f(E(\omega), \sigma(\omega), \gamma(\omega), \vartheta(\omega)) - \lambda\left(x_0 + \sum_i x_i - 1\right). \tag{3.7}$$

Using the relations stated in (3.4), (3.5) and (3.6); the Lagrange Multiplier problem in (3.7) can be rewritten as:

$$\mathfrak{L} = f\left(E(\omega), \sum_i x_i \beta_{iP} \sigma(R_P), \sum_i x_i \gamma_{iP} \gamma(R_P), \sum_i x_i \vartheta_{iP} \vartheta(R_P)\right) - \lambda\left(x_0 + \sum_i x_i - 1\right). \tag{3.8}$$

Then using

$$\frac{\partial E(\omega)}{\partial x_i} = 1 + E(R_i),$$

$$\frac{\partial \sigma(\omega)}{\partial x_i} = \beta_{iP}\sigma(R_P),$$

$$\frac{\partial \gamma(\omega)}{\partial x_i} = \gamma_{iP}\gamma(R_P),$$

$$\frac{\partial \theta(\omega)}{\partial x_i} = \vartheta_{iP}\vartheta(R_P);$$

we can write the FOC as

$$\frac{\partial \mathfrak{L}}{\partial x_0} = \frac{\partial E[U(\omega)]}{\partial E(\omega)}(1 + R_f) - \lambda = 0,$$

$$\frac{\partial \mathfrak{L}}{\partial x_i} = \frac{\partial E[U(\omega)]}{\partial E(\omega)}(1 + E(R_i)) + \frac{\partial E[U(\omega)]}{\partial \sigma(\omega)}\beta_{iP}\sigma(R_P) + \frac{\partial E[U(\omega)]}{\partial \gamma(\omega)}\gamma_{iP}\gamma(R_P) + \frac{\partial E[U(\omega)]}{\partial \vartheta(\omega)}\vartheta_{iP}\vartheta(R_P) - \lambda = 0.$$

Rearranging the FOC we obtain

$$E(R_i) - R_f = -\left[\frac{\frac{\partial E[U(\omega)]}{\partial \sigma(\omega)}}{\frac{\partial E[U(\omega)]}{\partial E(\omega)}}\right]\beta_{iP}\sigma(R_P) - \left[\frac{\frac{\partial E[U(\omega)]}{\partial \gamma(\omega)}}{\frac{\partial E[U(\omega)]}{\partial E(\omega)}}\right]\gamma_{iP}\gamma(R_P) - \left[\frac{\frac{\partial E[U(\omega)]}{\partial \vartheta(\omega)}}{\frac{\partial E[U(\omega)]}{\partial E(\omega)}}\right]\vartheta_{iP}\vartheta(R_P). \tag{3.9}$$

At the maximum, the expected utility is constant and the changes in expected return and variance are zero for a given level of skewness and kurtosis, *i.e.:*

$$dE[U(\omega)] = \tfrac{\partial E[U(\omega)]}{\partial E(\omega)}dE(\omega) + \tfrac{\partial E[U(\omega)]}{\partial \sigma(\omega)}d\sigma(\omega) = 0,$$

$$dE[U(\omega)] = \tfrac{\partial E[U(\omega)]}{\partial E(\omega)}dE(\omega) + \tfrac{\partial E[U(\omega)]}{\partial \gamma(\omega)}d\gamma(\omega) = 0,$$

$$dE[U(\omega)] = \tfrac{\partial E[U(\omega)]}{\partial E(\omega)}dE(\omega) + \tfrac{\partial E[U(\omega)]}{\partial \vartheta(\omega)}d\theta(\omega) = 0.$$

Incorporating these results into Equation (3.9) gives:

$$\begin{aligned} E(R_i) - R_f &= \left[\frac{dE(\omega)}{d\sigma(\omega)}\right] \beta_{iP}\sigma(R_P) + \left[\frac{dE(\omega)}{d\gamma(\omega)}\right] \gamma_{iP}\gamma(R_P) \\ &\quad + \left[\frac{dE(\omega)}{d\vartheta(\omega)}\right] \vartheta_{iP}\vartheta(R_P). \end{aligned} \tag{3.10}$$

It is important to stress that Equation (3.10) is defined in terms of a risky asset and a portfolio denoted by the subindex *P;* thus the pricing results denote an individual equilibrium. To move from an individual equilibrium to a market one –to derive the CAPM– it must be the case that an investor's choice of a risky investment portfolio is separate from her attitude towards risk. This property is often referred to as a portfolio separation principle.[5] Before we formally introduce this principle, its main assumptions are summarized:

(i) Each investor chooses a portfolio with the objective of maximizing a derived utility function, $f(E(\omega), \sigma(\omega), \gamma(\omega), \vartheta(\omega))$, where the utility function is concave and preferences induce the favouring of higher means, smaller variances, higher skewness and smaller kurtosis.

(ii) All investors have a common time horizon and homogeneous beliefs about $E(\omega)$, $\sigma(\omega)$, $\gamma(\omega)$ and $\vartheta(\omega)$.

(iii) Each asset is infinitely divisible.

(iv) The riskless asset can be bought or sold in unlimited amounts.

[5]This principle is also known as portfolio separation theorem or mutual fund theorem.

Theorem 3 *If assumptions (i) to (iv) hold, the optimal combination of risky assets for an investor can be determined without any knowledge of the investor's preferences towards risk and return.*

Theorem 3 is the so-called portfolio separation theorem. Under this theorem the investor makes two separate decisions:

1. After estimating the expected returns, variances, covariances, skewnesses, co-skewnesses, kurtosis and co-kurtosis of securities; the investor calculates an efficient set of risky assets. This is a set formed by the combination of assets that for a given level of variance, covariance, kurtosis, co-kurtosis, skewness and co-skewness yield the highest return. No personal characteristics, such as degree of risk aversion, are needed in this step. Intuitively, no other portfolio could be optimal since all investors working with the same inputs, sketch out the same efficient set of risky assets. If all investors choose the same portfolio of risky assets it is possible to determine what that portfolio is. Common sense points to it being a market valued-weighted portfolio of all existing securities: the market portfolio.

2. The investor must now determine how to combine the portfolio of risky assets with the riskless one. This allocation is determined by her tolerance towards risk.

Theorem 3 is fundamental to understanding the CAPM. It ensures that all individual investors maximize their utility with two funds: a riskless asset and the market portfolio.[6]

[6] For further insight of the Mutual Fund Theorem refer to Bottazzi *et al* (1995), Nielsen (1993) and Ingersoll (1987).

Therefore, after evoking the portfolio separation theorem, Equation (3.10) can be rewritten in terms of the market portfolio

$$E(R_i) - R_f = \left[\frac{dE(\omega)}{d\sigma(\omega)}\right] \beta_{im}\sigma(R_m) + \left[\frac{dE(\omega)}{d\gamma(\omega)}\right] \gamma_{im}\gamma(R_m) + \left[\frac{dE(\omega)}{d\vartheta(\omega)}\right] \vartheta_{im}\vartheta(R_m). \tag{3.11}$$

Note that (3.11) is identical to (3.10) except for R_m, the rate of return of the market portfolio, which is substituted for R_P. The main theoretical difference between both is that (3.11) is a market equilibrium whereas (3.10) is an individual equilibrium. From this point onwards, the subindex m labels the variables and parameters specific to the market portfolio.

Equation (3.11) is an extension of the Kraus and Litzenberger (1976) Three-Moment CAPM (henceforth, K-L CAPM). Following their notation, (3.11) can be rewritten as

$$E(R_i) - R_f = b_1\beta_{im} + b_2\gamma_{im} + b_3\vartheta_{im}, \tag{3.12}$$

where

$$b_1 = \left[\frac{dE(\omega)}{d\sigma(\omega)}\right] \sigma(R_m), \tag{3.13}$$

$$b_2 = \left[\frac{dE(\omega)}{d\gamma(\omega)}\right] \gamma(R_m), \tag{3.14}$$

$$b_3 = \left[\frac{dE(\omega)}{d\vartheta(\omega)}\right] \vartheta(R_m). \tag{3.15}$$

Equation (3.12) is the Four-Moment CAPM.

Note that:

(a) $b_1 > 0$,

since $\left[\frac{dE(\omega)}{d\sigma(\omega)}\right] > 0$ and $\sigma(R_m) > 0$.

(b) $b_2 > 0$ if $\gamma(R_m) < 0$,

$b_2 < 0$ if $\gamma(R_m) > 0$,

since $\frac{dE(\omega)}{d\gamma(\omega)} = -\frac{\partial E[U(\omega)]/\partial\gamma(\omega)}{\partial E[U(\omega)]/\partial E(\omega)} < 0$.

(c) $b_3 > 0$,

since $\frac{dE(\omega)}{d\vartheta(\omega)} = -\frac{\partial E[U(\omega)]/\partial\vartheta(\omega)}{\partial E[U(\omega)]/\partial E(\omega)} > 0$ and $\vartheta(R_m) > 0$.

Multiperiod Framework

Due to the fact that the CAPM is a single period model, all the previous equations do not have a time dimension. For econometric analysis of the CAPM, it is sufficient to assume *i.i.d.* returns to estimate the model over time (Campbell et al, 1997).

Lim (1989) tests the validity of the Three-Moment CAPM through the GMM by defining the CAPM in terms of orthogonality conditions. This specification is convenient since the EL is also a moments-based model. The extension of Lim's (1989) analysis to a Four-Moment framework arises naturally. Following his work, first define the deflated excess returns for the i^{th} asset and the market portfolio as:[7]

$$r_{it} = \left(\frac{R_{it} - R_{ft}}{R_{ft}}\right),$$

$$r_{mt} = \left(\frac{R_{mt} - R_{ft}}{R_{ft}}\right).$$

We now define the moment conditions, $E[g(r_{it}, r_{mt}, \theta)] = 0$, for estimating the Four-

[7]The rates of return on the riskless asset are not constant through time. Thus, the deflated excess returns are used to make moments of the rate of returns intertemporal constants under a changing riskless interest rate (Fama, 1970).

Moment CAPM:[8]

$$E\left[r_{it}-\left(b_1\beta_{im}+b_2\gamma_{im}+b_3\vartheta_{im}\right)\right]=0 \quad i=1,...N, \tag{3.16}$$

$$E\left[r_{it}r_{mt}-\mu\left(r_m\right)r_{it}-\beta_{im}\left\{r_{mt}-\mu\left(r_m\right)\right\}^2\right]=0 \;\; i=1,...,N, \tag{3.17}$$

$$E\left[r_{it}r_{mt}^2-2\mu\left(r_m\right)r_{it}r_{mt}+\mu\left(r_m\right)^2r_{it}-\sigma\left(r_m\right)^2r_{it}\right. \tag{3.18}$$

$$\left.-\gamma_{im}\left\{r_{mt}-\mu\left(r_m\right)\right\}^3\right]=0 \quad i=1,...,N,$$

$$E\left[r_{it}r_{mt}^3-3\mu\left(r_m\right)r_{it}r_{mt}^3+3\mu\left(r_m\right)^2r_{it}r_{mt}-\mu\left(r_m\right)^3r_{it}\right. \tag{3.19}$$

$$\left.-\gamma\left(r_m\right)^3r_{it}-\vartheta_{im}\left\{r_{mt}-\mu\left(r_m\right)\right\}^4\right]=0 \;\; i=1,..,N,$$

$$E\left[r_{mt}-\mu\left(r_m\right)\right]=0, \tag{3.20}$$

$$E\left[\left\{r_{mt}-\mu\left(r_m\right)\right\}^2-\sigma\left(r_m\right)^2\right]=0, \tag{3.21}$$

$$E\left[\left\{r_{mt}-\mu\left(r_m\right)\right\}^3-\gamma\left(r_m\right)^3\right]=0, \tag{3.22}$$

$$E\left[\left\{r_{mt}-\mu\left(r_m\right)\right\}^4-\vartheta\left(r_m\right)^4\right]=0. \tag{3.23}$$

These equations are better analysed by dividing them into two groups.

The first group, Equations (3.16) – (3.19), specify the relationship between the returns of the risky asset and the market. The N moment conditions in (3.16) come from the Four-Moment CAPM as defined in (3.12). The following $3N$ orthogonality conditions, Equations (3.17)–(3.19), are N conditions for beta, N conditions for co-skewness and N conditions for co-kurtosis.

[8] $g\left(.,\theta\right)$ was defined in Chapter 1, Section 1.4.

The second group, Equations (3.20) – (3.23), are particular to the market and they denote common measures for the mean, variance, skewness and kurtosis; respectively.

In total, there are 4N+4 equations and 3N+7 parameters to be estimated, $\theta = (b_1, b_2, b_3, \mu(r_m), \sigma(r_m), \gamma(r_m), \vartheta(r_m), \beta_{im}, \gamma_{im}, \vartheta_{im})^{\tau}$.

For simplicity it is convenient in what follows to make the assumption that $N = 1$. We will denote β_{im}, γ_{im}, ϑ_{im}, and r_{it} as β_m, γ_m, ϑ_m, and r_t; respectively.

3.3 Mean-Variance CAPM

Markowitz (1959) set down the basis for the CAPM. He formulated the investor's portfolio selection problem in terms of expected return and variance of return. He showed that investors would optimally hold a portfolio with the highest expected value for a given level of variance, *i.e.* a Mean-Variance efficient portfolio. Sharpe (1964) and Lintner (1965) extended the work of Markowitz (1959) to develop a general equilibrium model, the CAPM. They showed that if investors have homogeneous expectations and optimally hold Mean-Variance efficient portfolios then, in the absence of market frictions, the market portfolio will itself be a Mean-Variance portfolio. The Mean-Variance CAPM states that the expected return of an asset must be linear in the covariance of its return with the return of the market portfolio.

In this section, a comparison of the EL and GMM, in the context of the Mean-Variance CAPM, is carried out. Essentially what we do is to assess the finite-sample properties, size and power, of their moment restrictions tests.

First, we formally introduce the Mean-Variance CAPM as a particular case of the general model, given in Equation (3.12).

3.3.1 Moment Equations

The moment equations for estimating the Mean-Variance CAPM are Equations (3.16), (3.17) and (3.20); where $b_1 = E(r_{mt})$, $b_2 = 0$ and $b_3 = 0$ in (3.16).

Hence, there are 3 equations and 2 parameters, $\theta = (\mu(r_m), \beta_m)^{\tau}$, to be estimated.

3.3.2 Finite-Sample Properties of Overidentification Tests

The tests of overidentifying restrictions studied in this section have as their null hypothesis that there is a value of θ consistent with $E[g(r_t, r_{mt}, \theta)] = 0$. We analyse three tests of overidentifying restrictions in what follows: W_j, J_{2GMM} and J_{CuGMM}. The three tests have an asymptotic $\chi^2_{(1)}$ distribution under the null.

The Data Generating Process

We consider two DGPs.

Hwang and Satchell (1999) examine the CAPM for the case of emerging markets. The following figures are obtained from their survey, specifically for the case of Mexico:

$$\sigma(r_{MEX}) = 14.41,$$
$$\sigma(r_m) = 4.11, \tag{3.24}$$
$$E(r_{mt}) = .73,$$
$$\rho_{r_{MEX}, r_m} = .65,$$

where

$\sigma(r_{MEX})$ and $\sigma(r_m)$ are the standard deviations of Mexico and the market,

ρ_{r_{MEX}, r_m} denotes the correlation coefficient between the returns of Mexico and the market.

The Mean-Variance CAPM predicts

$$E(r_{MEX}) = \beta_m E(r_{mt}). \tag{3.25}$$

Our aim is to use a DGP consistent with the Mean-Variance CAPM.[9] We consider two processes.

We obtain from substituting (3.24) into (3.25) :

$$\begin{aligned} E(r_{MEX}) &= \frac{(.65)(14.41)(4.11)}{(4.11)^2}(.73) \\ &= 1.68. \end{aligned}$$

Hence, a DGP of the form

$$\begin{bmatrix} r_{MEX} \\ r_{mt} \end{bmatrix} \sim N(\mu, \Sigma), \tag{3.26}$$

[9]Consistent in the sense that there is a value of β_m such that the null hypothesis holds.

where

$$\mu = \begin{pmatrix} 1.68 + \Delta_{11} \\ .73 + \Delta_{21} \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} (14.41 + \Delta_{12})^2 & (.65)\,(14.41 + \Delta_{12})\,(4.11 + \Delta_{22}) \\ (.65)\,(14.41 + \Delta_{12})\,(4.11 + \Delta_{22}) & (4.11 + \Delta_{22})^2 \end{pmatrix},$$

and Δ_{ij} for $i, j = \{1, 2\}$ are constants that allow parameters to vary, is consistent with the Mean-Variance CAPM if $\Delta_{ij} = 0 \; \forall \; i, j$. In other words, if $\Delta_{ij} \neq 0$ then Equations (3.16), (3.17) and (3.20) do not hold.

The second DGP that we consider, a linear market model, has the following general form

$$r_t = a_1 \; r_{mt} + \varepsilon_t, \tag{3.27}$$

where

(i) $E(\varepsilon_t) = 0$,

(ii) r_{mt} and ε_t are uncorrelated.

Note that only if (*i*) and (*ii*) hold then the linear market model in (3.27) satisfies the Mean-Variance CAPM, *e.g.* $a_1 = \beta_m$.

Size of Overidentification Tests

This Section focuses on whether the asymptotic (or nominal) size is a good approximation to that in finite-samples. The following experiments employ either a DGP of the form given in (3.26) with $\Delta_{ij} = 0 \ \forall \ i, j$ or a DGP of the form in (3.27) with $E(\varepsilon_t) = 0$ and no correlation between r_{mt} and ε_t so that the Mean-Variance CAPM holds.

Results We first consider the DGP related to the Mexican figures. We report rejection frequencies, with particular interest being in cases where these probabilities are poorly approximated by the nominal size. Our experiments use 5000 replications. We consider two samples sizes: $n = 50$ and 100. Results are summarized in Table 3.1.

Empirical Levels of J-Tests
Mean-Variance CAPM

	$n = 50$			$n = 100$		
Level	W_j	J_{CuGMM}	J_{2GMM}	W_j	J_{CuGMM}	J_{2GMM}
.10	.1109	.1098	.1124	.1070	.1140	.0940
.05	.0593	.0592	.0561	.0590	.0610	.0520
.01	.0141	.0144	.0110	.0120	.0110	.0100

W_j, J_{CuGMM} and J_{2GMM} are J tests based on the ELR, continuously updated and two-step GMM estimators; respectively. n is the sample size.

Table 3.1: Finite-Sample Size Properties - CAPM Mexican data

Table 3.1 summarizes the rejection frequencies for the tests at the .10, .05 and .01 critical values. Our findings show that the nominal size is a reasonable approximation to finite-sample sizes for the three tests.

Our experiments illustrate that for these sample sizes (and a well behaved DGP), the nominal critical values of the overidentification tests can be a useful guide to finite-sample behaviour.

Consider the second DGP, Equation (3.27). We generate pseudorandom samples with the following characteristics:

$$\begin{bmatrix} r_{mt} \\ \varepsilon_t \end{bmatrix} \sim N \left[\begin{pmatrix} 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right]$$

and we arbitrarily set $a_1 = 1.5$ in (3.27), *i.e.*:

$r_t = 1.5\ r_{mt} + \varepsilon_t$.

The empirical levels for 5000 replications and for two sample sizes: $n = \{50, 100\}$; are summarized in Table 3.2. For n=50, the ELR test is more oversized than its GMM counterparts. However, differences among the three tests are small. We note the familiar decrease in the size distortions as n increases.

Empirical Levels of J-Tests
Mean-Variance CAPM

	$n = 50$			$n = 100$		
Level	W_j	J_{CuGMM}	J_{2GMM}	W_j	J_{CuGMM}	J_{2GMM}
.10	.1250	.1146	.1144	.1180	.1106	.1101
.05	.0706	.0608	.0622	.0598	.0560	.0572
.01	.0182	.0120	.0126	.0146	.0121	.0116

W_j, J_{CuGMM} and J_{2GMM} are J tests constructed through the ELR, continuously updated and two-step GMM estimators; respectively.

Table 3.2: Finite-Sample Size Properties - CAPM linear market model

Power of Overidentification Tests

When drawing inferences using a given test statistic it is important to consider its power. This is the probability that the null hypothesis will be rejected given that the alternative hypothesis is true. Low power suggests that the test is not useful to discriminate between the alternative and the null hypothesis.

To document the power of a test it is necessary to specify the alternative DGP and the size of the test. In what follows we consider a DGP of the form given in (3.26). Under the null hypothesis $\Delta_{ij} = 0 \; \forall \; i, j$. The experiments reported in this section set $\Delta_{ij} \neq 0$ in so that the moment conditions are invalid. We use the rejection frequencies as estimates of one minus the probability of Type II error.

Two main experiments are carried out. The first one considers variations in the means of the returns by setting $\Delta_{i1} \neq 0$ for $i = 1, 2$. The second experiment deals with fluctuations of the variances of the returns by letting $\Delta_{i2} \neq 0$ for $i = 1, 2$.

To separate the effect of size distortions we report the rejection frequencies for the cases where the critical values are given by the (estimated) .10, .05 and .01 critical values of the finite-sample null distribution.

Size Correction To obtain the .10, .05 and .01 finite-sample critical values we perform a Monte Carlo experiment with $\Delta_{ij} = 0 \; \forall \; i, j$. After ordering the simulated values of the overidentification tests from the largest to the smallest we find the 500^{th}, 250^{th} and 50^{th} values (since 5000 replications were performed). These values are the corrected critical values[10]. Results for $n = 100$ are summarized in Table 3.3.

[10] Horowitz and Savin (2000) argue that the size-corrected critical values usually obtained in Monte Carlo studies of power are both misnamed (what is really computed is the exact Type I critical value for essentially arbitrary simple null hypotheses) and irrelevant to empirical research (because the chosen parameter value is arbitrary, the critical value has no empirical analog). Horowitz and Savin (2000) propose an alternative method based on the bootstrap for obtaining critical values in power studies.

Size Correction Finite-Sample Critical Values Mean-Variance CAPM				
	Asymptotic critical value	Corrected Critical Value		
Levels		W_j	J_{CuGMM}	J_{2GMM}
.10	2.7055	2.8091	2.9277	2.6695
.05	3.8414	4.1485	4.1651	3.9135
.01	6.6348	6.8828	6.7997	6.6040

W_j, J_{CuGMM} and J_{2GMM} are J tests constructed through the ELR, continuously updated and two-step GMM estimators; respectively.

Table 3.3: Size Correction Mean-Variance CAPM - Mexican data

Results Experiment 1: Variations in the Means

We set $\Delta_{i1} \neq 0$ in (3.26) for $i = 1, 2$. It is easy to see that deviations from the null hypothesis are given by

$$\Delta_{21} \frac{38.56}{(4.11)^2} - \Delta_{11}.$$

If Δ_{21} and Δ_{11} are both positive (negative), it is ambiguous if these increments (decrements) lead to departures from the null because both effects might cancel each other. Moreover, larger $\Delta_{i1}^{'}s$ are not necessarily interpreted as larger deviations from the null. Hence, we concentrate on the cases in which the means vary in opposite directions: *i.e.* $\Delta_{11} > 0$ ($\Delta_{11} < 0$) and $\Delta_{21} < 0$ ($\Delta_{21} > 0$). Forty eight different cases are studied. The ranges of the variations are between -2 and +2: $\Delta_{i1} = \{-2, ..., +2\}$ for $i = 1, 2$. Results for a significance level of 5% are shown in Table 3.4.

Power of Moment Restrictions Tests Nominal Size=.05 Mean-Variance CAPM										
	Δ_{21} / Δ_{11}	-2	-1	-.5	-.2	0	.2	.5	1	2
W_j	2	1	.973	.818	.628	.458				
J_{CuGMM}		1	.970	.806	.625	.437				
J_{2GMM}		1	.972	.803	.595	.415				
W_j	1	.998	.852	.518	.285	.159				
J_{CuGMM}		.998	.848	.511	.277	.142				
J_{2GMM}		.995	.833	.471	.236	.146				
W_j	.5	.992	.708	.323	.148	.086				
J_{CuGMM}		.989	.690	.310	.121	.061				
J_{2GMM}		.986	.673	.277	.130	.054				
W_j	.2	.982	.634	.242	.099	.070				
J_{CuGMM}		.980	.619	.213	.097	.051				
J_{2GMM}		.978	.610	.221	.075	.047				
W_j	0	.975	.538	.172	.063		.084	.171	.490	.934
J_{CuGMM}		.972	.515	.163	.062		.071	.150	.485	.915
J_{2GMM}		.970	.524	.135	.051		.067	.150	.463	.916
W_j	-.2					.061	.112	.218	.533	.955
J_{CuGMM}						.051	.093	.214	.489	.936
J_{2GMM}						.044	.082	.207	.492	.930
W_j	-.5					.094	.149	.320	.665	.964
J_{CuGMM}						.058	.140	.294	.631	.964
J_{2GMM}						.070	.129	.283	.618	.964
W_j	-1					.158	.267	.472	.787	.985
J_{CuGMM}						.148	.251	.450	.783	.979
J_{2GMM}						.139	.241	.468	.763	.978
W_j	-2					.446	.604	.792	.953	1
J_{CuGMM}						.436	.572	.782	.933	.997
J_{2GMM}						.440	.594	.761	.947	.997

W_j, J_{CuGMM} and J_{2GMM} are J-tests based on the ELR, continuously updated and two-step GMM estimators; respectively.

Table 3.4: Power Properties Mean-Variance CAPM - variations in means

Each coordinate in Table 3.4 represents $(\Delta_{11}, \Delta_{21})$, where Δ_{11} are changes induced to the mean of r_{MEX} and Δ_{21} are changes induced to the mean of r_{mt}. We do not experiment with the coordinate $(0,0)$ since the null holds.

Our results suggest that the ELR overidentification test is more able to detect deviations from the null than tests based on the GMM. Of 48 experiments EL has the greatest power in all 48 of the cases and the two-step and continuously updated GMM are as powerful as the EL in 2 cases.

As expected, power increases as the variations in the means increase and it is also noteworthy that there are no important differences between positive and negative values of $\Delta_{i1}' s$.

We carry out a second experiment to assess the power of overidentification tests.

Experiment 2: Variations in the Variances

We set $\Delta_{i2} \neq 0$ in (3.26) for $i = 1, 2$. The new expected return for the risky asset implied by the Mean-Variance CAPM is

$$E\left(r_{MEX1}\right) = \frac{.65\ \left(14.41 + \Delta_{12}\right)}{\left(4.11 + \Delta_{22}\right)}\left(.73\right).$$

However, we generate random numbers considering the original expected value

$$E\left(r_{MEX}\right) = \frac{.65\ \left(14.41\right)}{\left(4.11\right)}\left(.73\right) = 1.68.$$

Note that if Δ_{12} and Δ_{22} are both positive (negative) it is ambiguous if these increases (decreases) lead to departures from the null hypothesis. Hence, we concentrate on the cases in which the variances vary in opposite directions: $\Delta_{12} > 0$ $(\Delta_{12} < 0)$ and $\Delta_{22} < 0$ $(\Delta_{22} > 0)$.

Power of Moment Restrictions Tests Nominal Size=0.05 Mean-Variance CAPM									
	Δ_{12} \ Δ_{22}	-3	-2	-1	0	1	2	3	4
W_j		.956	.385	.145	.063				
J_{CuGMM}	4	.943	.382	.121	.057				
J_{2GMM}		.951	.382	.102	.042				
W_j		.952	.404	.126	.059				
J_{CuGMM}	3	.945	.371	.090	.055				
J_{2GMM}		.936	.327	.126	.044				
W_j		.935	.350	.103	.056				
J_{CuGMM}	2	.918	.321	.101	.050				
J_{2GMM}		.926	.316	.100	.055				
W_j		.937	.311	.099	.055				
J_{CuGMM}	1	.915	.273	.079	.047				
J_{2GMM}		.918	.279	.094	.052				
W_j		.931	.304	.091		.051	.096	.101	.127
J_{CuGMM}	0	.920	.266	.061		.049	.088	.093	.127
J_{2GMM}		.917	.269	.072		.049	.075	.089	.118
W_j					.062	.072	.099	.125	.164
J_{CuGMM}	-1				.056	.064	.082	.124	.144
J_{2GMM}					.046	.057	.076	.117	.128
W_j					.054	.111	.116	.153	187
J_{CuGMM}	-2				.041	.077	.113	.143	.178
J_{2GMM}					.054	.071	.109	.137	.156
W_j					.071	.123	.163	.196	.219
J_{CuGMM}	-3				.064	.123	.157	.193	.216
J_{2GMM}					.063	.091	.146	.170	.169
W_j					.100	.146	.203	.232	.276
J_{CuGMM}	-4				.091	.133	.200	.214	.269
J_{2GMM}					.089	.143	.193	.204	.263
W_j, J_{CuGMM}, J_{2GMM} are J-tests based on the ELR, continuously updated and two-step GMM estimators; respectively.									

Table 3.5: Power Properties Mean-Variance CAPM - variations in variances

Forty eight different cases are studied. The ranges of the variations are between -4 and +4: $\Delta_{i2} = \{-4, ..., +4\}$ for $i = 1, 2$. We omit $\Delta_{22} = -4$ because we encountered several problems when generating random numbers given that the variance is close to zero: $(4.11 - 4) = .11$. Results for a significance level of 5% and $n = 100$ are shown in Table 3.5. We performed 5000 replications.

Our findings shown in Table 3.5 suggest that W_j is more able to detect false moment conditions than tests based on the GMM. In the 48 experiments W_j has the greatest power in all cases. J_{2GMM} is as powerful as W_j in 2 of the cases and J_{CuGMM} is as powerful as W_j in 2 of the replications. As would be expected, power increases as Δ_{i2} increases. The latter results are consistent with the findings of our first experiment. The GMM tests fail to detect the invalidity of the moment conditions in a greater extent than the ELR test does.

3.4 Three-Moment CAPM

When contradictory empirical results for the traditional form of the Sharpe-Lintner model emerged, authors such as Kraus and Litzenberger (1976) extended the Mean-Variance framework to incorporate the effect of skewness on valuation. They argue that prior empirical findings that were interpreted as inconsistent with the traditional theory can be attributed to misspecification of the CAPM by omission of systematic skewness.

By setting $b_3 = 0$ in (3.12), the K-L CAPM follows.

It is crucial to address the fact that the market price of beta reduction, b_1, and the market price of gamma, b_2, can be expressed in terms of the market's return. To illustrate this assertion consider the special case in which all investors have logarithmic utility functions.[11]

[11]The logarithmic function is representative of utility functions displaying decreasing absolute risk aversion and constant relative risk aversion.

A Taylor approximation of the investor's expected utility of end of period wealth, $E\left[U\left(\omega\right)\right]=f\left(E\left(\omega\right),\sigma\left(\omega\right),\gamma\left(\omega\right)\right)$, yields:

$$E\left[U\left(\omega\right)\right]=\log\left(E\left(\omega\right)\right)-\frac{\sigma\left(\omega\right)^2}{2E\left(\omega\right)^2}+\frac{\gamma\left(\omega\right)^3}{3E\left(\omega\right)^3}. \tag{3.28}$$

When we differentiate (3.28) with respect to $E\left(\omega\right)$, $\sigma\left(\omega\right)$ and $\gamma\left(\omega\right)$ we obtain

$$\frac{\partial E\left[U\left(\omega\right)\right]}{\partial E\left(\omega\right)} = \frac{1}{E\left(\omega\right)}+\frac{\sigma\left(\omega\right)^2}{E\left(\omega\right)^3}-\frac{\gamma\left(\omega\right)^3}{E\left(\omega\right)^4}, \tag{3.29}$$

$$\frac{\partial E\left[U\left(\omega\right)\right]}{\partial \sigma\left(\omega\right)} = -\frac{\sigma\left(\omega\right)}{E\left(\omega\right)^2}, \tag{3.30}$$

$$\frac{\partial E\left[U\left(\omega\right)\right]}{\partial \gamma\left(\omega\right)} = \frac{\gamma\left(\omega\right)^2}{E\left(\omega\right)^3}. \tag{3.31}$$

Substitution of (3.29), (3.30) and (3.31) into (3.13) and (3.14) yields

$$b_1 = \left(\frac{-\frac{\sigma\left(\omega\right)}{E\left(\omega\right)}}{1+\frac{\sigma\left(\omega\right)^2}{E\left(\omega\right)^2}-\frac{\gamma\left(\omega\right)^3}{E\left(\omega\right)^3}}\right)\sigma\left(r_m\right), \tag{3.32}$$

$$b_2 = \left(\frac{\frac{\gamma\left(\omega\right)^2}{E\left(\omega\right)^2}}{1+\frac{\sigma\left(\omega\right)^2}{E\left(\omega\right)^2}-\frac{\gamma\left(\omega\right)^3}{E\left(\omega\right)^3}}\right)\gamma\left(r_m\right). \tag{3.33}$$

Since the initial investment is set to one, the moments of end of period wealth are equivalent to those of the rate of return on the portfolio, in equilibrium the market portfolio. Therefore, we can rewrite b_1 and b_2 as

$$b_1 = \frac{-\frac{\sigma\left(r_m\right)^2}{E\left(r_{mt}\right)}}{1+\frac{\sigma\left(r_m\right)^2}{E\left(r_{mt}\right)^2}-\frac{\gamma\left(r_m\right)^3}{E\left(r_{mt}\right)^3}}, \tag{3.34}$$

$$b_2 = \frac{\frac{\gamma\left(r_m\right)^3}{E\left(r_{mt}\right)^2}}{1+\frac{\sigma\left(r_m\right)^2}{E\left(r_{mt}\right)^2}-\frac{\gamma\left(r_m\right)^3}{E\left(r_{mt}\right)^3}}. \tag{3.35}$$

Note that as soon as information about the investor's preferences is incorporated, b_1 and b_2 can be expressed in terms of the market.

Thus, the special case of K-L CAPM, where all investors have logarithmic utility functions, is

$$E(r_t) = \left[\frac{-\frac{\sigma(r_m)^2}{E(r_{mt})}}{1+\frac{\sigma(r_m)^2}{E(r_{mt})^2}-\frac{\gamma(r_m)^3}{E(r_{mt})^3}}\right]\beta_m + \left[\frac{\frac{\gamma(r_m)^3}{E(r_{mt})^2}}{1+\frac{\sigma(r_m)^2}{E(r_{mt})^2}-\frac{\gamma(r_m)^3}{E(r_{mt})^3}}\right]\gamma_m. \tag{3.36}$$

We can alternatively use a variant of the K-L CAPM that provides information about the structure of the risk premiums, b_1 and b_2, by using the Euler condition for the investor's utility maximization problem as in Seirs and Wei (1985). This is

$$E(r_t) = \left(\left[\frac{\phi\sigma(r_m)}{\phi\sigma(r_m)-\gamma(r_m)}\right]\beta_m - \left[\frac{\gamma(r_m)}{\phi\sigma(r_m)-\gamma(r_m)}\right]\gamma_m\right) r_{mt}, \tag{3.37}$$

where ϕ is the marginal rate of substitution of γ for σ (refer to Seirs and Wei, 1985). Note that as for the logarithmic utility case, b_1 and b_2 are now expressed in terms of the market return.

3.4.1 Moment Equations

The orthogonality conditions that characterize the Three-Moment CAPM are given by Equations (3.16), (3.17), (3.18), (3.20), (3.21) and (3.22); where we set $b_3 = 0$ in (3.16).

3.4.2 Finite-Sample Properties of Overidentification Tests

The tests of overidentifying restrictions studied in this section have as their null hypothesis that there is a value of θ consistent with $E[g(r_t, r_{mt}, \theta)] = 0$; where $\theta = (b_1, b_2, \beta_m, \gamma_m, \mu(r_m), \sigma(r_m), \gamma(r_m))^\tau$. We again consider the three tests of overidentifying restrictions studied in the Mean-Variance setting: W_j, J_{2GMM} and J_{CuGMM}. These tests have a $\chi^2_{(r-q)}$ distribution under the null, where $\dim(g) = r$ and $\dim(\theta) = q$.

The Data Generating Process

Assume the following quadratic market model:

$$r_t = a_1 r_{mt} + a_2(r_{mt} - E(r_{mt}))^2 + \varepsilon_t, \tag{3.38}$$

where

$$\begin{aligned} &(i)\ \ a_i \neq 0 \text{ for } i = \{1, 2\}, \\ &(ii)\ \ \varepsilon_t \text{ is independent of } r_{mt} \text{ and } (r_{mt} - E(r_{mt}))^2, \\ &(iii)\ \ E(\varepsilon_t) = 0 + \Delta \\ &\qquad \text{and } \ \Delta = 0. \end{aligned} \tag{3.39}$$

Then applying the definitions of β_m and γ_m to the quadratic market model we obtain:[12]

$$\beta_m = a_1 + a_2 \frac{\gamma(r_m)^3}{\sigma(r_m)^2}, \tag{3.40}$$

$$\gamma_m = a_1 + a_2 \frac{\left(\vartheta(r_m)^4 - \sigma(r_m)^4\right)}{\gamma(r_m)^3}. \tag{3.41}$$

It is helpful to seek to express a_1 and a_2 in terms of β_m and γ_m. Solving (3.40) and (3.41) for a_1 and a_2 yields

$$a_1 = \frac{-\beta_m \vartheta(r_m)^4 \sigma(r_m)^2 + \beta_m \sigma(r_m)^6 + \gamma(r_m)^6 \gamma_m}{\gamma(r_m)^6 - \vartheta(r_m)^4 \sigma(r_m)^2 + \sigma(r_m)^6}, \tag{3.42}$$

$$a_2 = \frac{\sigma(r_m)^2 \gamma(r_m)^3 (-\gamma_m + \beta_m)}{\gamma(r_m)^6 - \vartheta(r_m)^4 \sigma(r_m)^2 + \sigma(r_m)^6}. \tag{3.43}$$

[12]The proofs of (3.40) and (3.41) are in Appendix 2.

Thus, the DGP in (3.38) can be rewritten as:

$$\begin{aligned} r_t &= \left[\frac{-\beta_m \vartheta (r_m)^4 \sigma (r_m)^2 + \beta_m \sigma (r_m)^6 + \gamma (r_m)^6 \gamma_m}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6}\right] r_{mt} \\ &+ \left[\frac{\sigma (r_m)^2 \gamma (r_m)^3 (-\gamma_m + \beta_m)}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6}\right] (r_{mt} - E(r_{mt}))^2 + \varepsilon_t. \end{aligned} \tag{3.44}$$

Factorizing β_m and γ_m yields

$$r_t = A_1 \beta_m + A_2 \gamma_m + \varepsilon_t, \tag{3.45}$$

where

$$\begin{aligned} A_1 &= \frac{-\vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} r_{mt} \\ &+ \frac{\sigma (r_m)^2 \gamma (r_m)^3}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} (r_{mt} - E(r_{mt}))^2, \end{aligned} \tag{3.46}$$

$$\begin{aligned} A_2 &= \frac{\gamma (r_m)^6}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} r_{mt} \\ &- \frac{\sigma (r_m)^2 \gamma (r_m)^3}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} (r_{mt} - E(r_{mt}))^2. \end{aligned} \tag{3.47}$$

Note that the proposed quadratic market model, Equation (3.38), has been rewritten as (3.45). Here, we have two terms: one which factorizes beta and one which factorizes systematic kurtosis.

For simplicity, we consider a specification of the K-L CAPM that provides information about the structure of the risk premiums[13]. From this point onwards we focus on the model given in (3.37). We introduce a normalization variable to generate data consistent with this framework. Let ϕ in (3.37) be a normalization variable such that

$$\int\int\left\{\left(r_t-\left[\left\{\frac{\phi\sigma(r_m)}{\phi\sigma(r_m)-\gamma(r_m)}\right\}\beta_m - \left\{\frac{\gamma(r_m)}{\phi\sigma(r_m)-\gamma(r_m)}\right\}\gamma_m\right]r_{mt}\right)f_{r_t,r_{mt}}\right\}dr_{mt}dr_t=0, \tag{3.48}$$

where $f_{r_t,r_{mt}}$ is the joint density function of the risky and market returns.

Consider a DGP of the form given in (3.38) where the conditions in (3.39) hold. Let $r_{m_t} \sim \chi^2_{(k)}$ and $\varepsilon_t \sim N(0,1)$. Then:

$$\begin{aligned} E(r_{mt}) &= k, \\ \sigma^2(r_m) &= 2k, \\ \gamma^3(r_m) &= 8k, \\ \vartheta^4(r_m) &= 12k(k+4). \end{aligned}$$

Substitution of these values into (3.40) and (3.41) yields:

$$\beta_m = a_1 + 4a_2,$$

$$\gamma_m = a_1 + a_2(6+k).$$

[13] The advantage of expressing b_1 and b_2 in terms of the market is that these are no longer parameters to be estimated.

Hence, we can rewrite (3.48) as:

$$\int\int\left\{\left(a_1 r_{mt} + a_2(r_{mt}-k)^2 + \varepsilon_t - \left[\left\{\frac{\phi\,(2k)^{1/2}}{\phi\,(2k)^{1/2}-(8k)^{1/3}}\right\}(a_1+4a_2) \right.\right.\right. \tag{3.49}$$
$$\left.\left.\left. - \left\{\frac{(8k)^{1/3}}{\phi\,(2k)^{1/2}-(8k)^{1/3}}\right\}(a_1+a_2\,(6+k))\right] r_{mt}\right) f_{r_{mt}} f_{\varepsilon_t}\right\} dr_{mt}d\varepsilon_t = 0,$$

where $f_{r_{mt}}$ is the chi-square marginal density of r_{mt} and f_{ε_t} is the standard normal marginal density function of ε_t.

After some simplification we obtain:

$$\phi = \frac{(.177\ a_1 + 1.06\ a_2 - .177\ k\ \ (a_1+a_2))}{k^{1/6}\left(.516\times 10^{-11} a_2 k^2 + .125 a_1 + .501 a_2 - .125 k a_1 - .25 a_2 k\right)}. \tag{3.50}$$

To generate r_t as in (3.38) we must specify the degrees of freedom for $r_{mt} \sim \chi^2_{(k)}$ and set values for a_1 and a_2. Before doing this we review a key point. In Section 3.2 we define the market price of beta, b_1, and the systematic skewness, b_2, as:

(i) $b_1 = \left[\frac{dE(\omega)}{d\sigma(\omega)}\right]\sigma\left(r_m\right)$ where $b_1 > 0$,

(ii) $b_2 = \left[\frac{dE(\omega)}{d\gamma(\omega)}\right]\gamma\left(r_m\right)$ where $b_2 < 0$ if $\gamma\left(r_m\right) > 0$ and $b_2 > 0$ if $\gamma\left(r_m\right) < 0$.

It is easy to see from (3.12) and (3.37) that

$b_1 = \left[\frac{\phi\sigma(r_m)}{\phi\sigma(r_m)-\gamma(r_m)}\right] r_{mt}$ and $b_2 = -\left[\frac{\gamma(r_m)}{\phi\sigma(r_m)-\gamma(r_m)}\right] r_{mt}$.

Substituting $a_1 = 1.5$, $a_2 = .5$ and $k = 1$ into (3.50) yields $\phi = 3.53$. Since $\gamma\left(r_m\right) > 0$ we only need that $\phi\sigma\left(r_m\right) - \gamma\left(r_m\right) > 0$ so that (i) and (ii) hold, which is actually true for these values.

Size of Overidentification Tests

In this section we examine whether the asymptotic (or nominal) size is a good approximation to the size in finite-samples. The following experiment uses the already defined DGP; *i.e.*:

$r_t = 1.5\ r_{mt} + .5\ (r_{mt} - E(r_{mt}))^2 + \varepsilon_t,$

where $r_{mt} \sim \chi^2_{(1)}$ and $\varepsilon_t \sim N(0,1)$.

Note that for the K-L CAPM specification that we study, (3.37) replaces (3.16) so that the moment equations are (3.37), (3.17), (3.18), (3.20), (3.21) and (3.22) with $\phi = 3.53$. Thus W_j, J_{2GMM} and J_{CuGMM} have an asymptotic $\chi^2_{(1)}$ distribution under the null. We compute 5000 replications for two sample sizes: $n = \{50, 100\}$. The empirical levels of the J-tests are reported in Table 3.6.

Empirical Levels of J-Tests
Three-Moment CAPM

	$n = 50$			$n = 100$		
Level	W_j	J_{CuGMM}	J_{2GMM}	W_j	J_{CuGMM}	J_{2GMM}
.10	.1160	.1114	.1225	.1020	.0932	.1110
.05	.0640	.0658	.0705	.0510	.0436	.0620
.01	.0210	.0238	.0222	.0180	.0096	.0114

W_j, J_{CuGMM} and J_{2GMM} are J-tests based on the ELR, continuously updated and two-step GMM estimators, respectively.

Table 3.6: Finite-Sample Size Properties - Three-Moment CAPM

Results Table 3.6 shows that for both sample sizes the rejection probabilities of the three tests are close to their nominal levels. As n increases size distortions tend to decrease. Given these results, the size properties cannot be used as a criterion for choosing among the overidentification tests that we study. At this point, it is natural to prefer the test whose power is closer to unity. We investigate power properties of J-tests in the following section.

Power of Overidentification Tests

To make the power (percentage of rejections under the alternative hypothesis) of different test procedures comparable we calculate exact 10%, 5% and 1% critical values from the experiment conducted in the previous Section. These size corrected critical values are used, thus making the power of different test procedures comparable.

To examine power, we concentrate on the following cases:

1. Let $\Delta \neq 0$ in (3.39), so that $E(\varepsilon_t) \neq 0$.

2. Let $a_2 = 0$ in (3.38), so that the model is overidentified[14].

Size Correction To obtain the .10, .05 and .01 finite-sample critical values we use Monte Carlo simulations. After ordering the simulated values of the overidentification tests from the largest to the smallest we find the 500^{th}, 250^{th} and 50^{th} values (since 5000 replications were performed). These values are the corrected critical values. Results for $n = 50$ and $n = 100$ are summarized in Table 3.7.

From our results in Table 3.7 we note that for $n = 100$ the $\chi^2_{(1)}$ is a good approximation to the finite-sample distribution of the three test statistics. Hence, using asymptotic critical values for this sample size to assess power seems a safe undertaking.

[14]Note that a linear market model is consistent with the Mean-Variance framework whereas a quadratic market model is consistent with the Three-Moment CAPM.

Size Correction
Finite-Sample Critical Values
Three-Moment CAPM

Level	Asymptotic critical value	Corrected Critical Value W_j	J_{CuGMM}	J_{2GMM}
			n=50	
.10	2.7055	3.0104	2.9204	3.2271
.05	3.8414	4.6525	4.4869	4.9780
.01	6.6348	11.2156	13.8448	13.2016
			n=100	
.10	2.7055	2.7810	2.6025	2.8990
.05	3.8414	3.9650	3.6361	4.1502
.01	6.6348	6.7202	6.5429	6.6801

W_j, J_{CuGMM} and J_{2GMM} are J-tests based on the ELR, continuously updated and two-step GMM; respectively. n is the sample size.

Table 3.7: Size Correction - Three-Moment CAPM

Results Experiment 1: Variations in the error term

We set $\Delta \neq 0$ in (3.39) so that the moment conditions of the Three-Moment CAPM are invalid. Eight departures from the null are considered and the ranges of the variations are between -1 and +1: $\Delta = \{-1, ..., +1\}$.

We calculate the rejection frequencies as estimates of one minus the probability of Type II error at the nominal .10, .05 and .01 critical values. 5000 replications were used and two sample sizes considered: $n = \{50, 100\}$. Results are reported in Table 3.8.

Power of Moment Restriction Tests
Three-Moment CAPM
Variation in the error term

	W_j Levels			J_{CuGMM} Levels			J_{2GMM} Levels		
Δ	.10	.05	.01	.10	.05	.01	.10	.05	.01
					$n = 50$				
-1	1	.878	.374	.926	.779	.039	1	.841	.222
-.5	.902	.701	.231	.917	.769	.041	.863	.608	.184
-.2	.444	.201	.080	.170	.084	.004	.396	.233	.047
-.1	.200	.073	.041	.113	.052	.005	.152	.054	.021
.1	.158	.088	.021	.123	.067	.019	.110	.061	.019
.2	.182	.111	.020	.173	.108	.026	.164	.140	.058
.5	.301	.300	.201	.410	.317	.083	.301	.241	.117
1	1	.706	.489	.701	.636	.334	.694	.613	.497
					$n = 100$				
-1	1	1	1	.965	.941	.803	.990	.950	.921
-.5	.926	.768	.412	.779	.649	.318	.879	.701	.361
-.2	.575	.351	.118	.288	.171	.037	.442	.272	.069
-.1	.183	.107	.039	.151	.081	.016	.166	.093	.021
.1	.127	.073	.027	.103	.058	.020	.100	.052	.018
.2	.149	.121	.058	.156	.111	.049	.162	.131	.041
.5	.311	.309	.367	.433	.375	.250	.301	.300	.190
1	1	1	1	.780	.744	.643	.711	.691	.611

W_j, J_{CuGMM} and J_{2GMM} are J-tests based on the ELR, continuously updated and two-step GMM estimators. n is the sample size.

Table 3.8: Power Properties Three-Moment CAPM - variation in the error term

In most of the cases power increases as the departures from the null increase. We observe that for all the cases that we examine, the ELR test performs better than the GMM tests. Intriguingly, power is higher for negative departures from the null hypothesis than for positive deviations.

These results are new in this kind of literature. Kitamura (2001) found that the power of

ELR was greater than that of GMM tests when power was already high. Here, we find that the power of ELR is uniformly better.

Experiment 2: Three-Moment CAPM versus Mean-Variance CAPM

The null and alternative hypothesis that we consider are:

$H_0 : The\ Three-Moment\ CAPM\ is\ valid$

$H_a : The\ Mean-Variance\ CAPM\ is\ valid.$

We have already shown that while linear market characteristic lines are consistent with the Mean-Variance CAPM, quadratic market lines characterize the Three-Moment CAPM. Hence, if we set $a_2 = 0$ in (3.38) then the model characterized by (3.37), (3.17), (3.18), (3.20), (3.21) and (3.22) is overidentified.

We perform 5000 replications and calculate the rejection frequencies as estimates of one minus the probability of Type II error at the nominal .10, .05 and .01 critical values for two sample sizes: $n = 50$ and $n = 100$. Results are reported in Table 3.9.

Power of Moment Restriction Tests
Three-Moment CAPM

	n=50			n=100		
	Level			Level		
	.10	.05	.01	.10	.05	.01
W_j	.4912	.2618	.1117	.4900	.2883	.1481
J_{CuGMM}	.3416	.1906	.0558	.3522	.2724	.1204
J_{2GMM}	.3754	.2086	.0650	.4065	.2700	.1303

W_j, J_{CuGMM} and J_{2GMM} are J tests based on the ELR, continuously updated and two-step GMM; respectively.

Table 3.9: Power Properties Three-Moment CAPM vs. Mean-Variance CAPM

The ranking among tests show that the EL performs better than GMM tests. Throughout this experiment design, W_j has the highest power for both sample sizes. For most of the cases that we study, J_{CuGMM} has the lowest power.

3.5 Conclusions

We compared the finite-sample size properties of overidentification tests based on EL and GMM within two variants of the CAPM. While there is a large amount of literature on the GMM that uses a Two-Moment framework to examine size and power of its overidentifying restrictions tests, there are no studies which use a higher moment setting. The finite-sample properties of the J-test based on the EL has not been previously assessed in the asset pricing literature. In addition, little is known about its power properties in general.

Our experiments show that there are no clear advantages in terms of size when the GMM overidentification tests are compared to those based on EL within a Two-Moment and Three-Moment setting. The three tests have moderate size distortions. However, our findings illustrate that the ELR overidentification statistic is more powerful in detecting deviations from the null under the alternatives that we analysed. We also found some evidence that this statistic has uniformly greater power than tests based on GMM whereas Kitamura (2001) shows that the ELR J-test has better power when this is already high.

When we compared the power of overidentification tests within the Three-Moment framework, we tested against the alternative that the Mean-Variance CAPM is valid. We are not aware of any other study which assesses the power properties of tests of overidentifying restrictions using this interpretation.

Chapter 4

Dynamic Panel Data Models

4.1 Introduction

Dynamic panel econometric models are of interest in a wide range of economic applications, including Euler equations for household consumption, adjustment cost models for firms' factor demands and models of economic growth. These models enable researchers to analyse dynamic relationships from cross section units. Moreover, dynamic panel data models offer the possibility of investigating heterogeneity in adjustment dynamics between different types of individuals, households or firms.

In this chapter we analyse autoregressive panel data models with individual effects. The basic idea of these models is to introduce a generic individual effect in the random term. This term is divided into two parts or components: an individual component and an overall remainder[1]. This specification is widely used in empirical work since individual effects

[1]While assuming that all the reaction coefficients are fixed and the same for all individuals.

are typically believed to be related to a large number of non-observable random causes. The main interest usually lies in the coefficients of the slope parameters and less in the individual differences. If the investigator wants to make inference with respect to population characteristics and in addition has a large number of cross-section units, each observed for a small number of time periods, these models are adequate. The formal description of autoregressive models with individual effects is presented in Section 4.2 (for a detailed description of these models refer to Chapter 6 in Arellano, 2003; Chapter 7 in Mátyás and Sevestre, 1996).

One of the disadvantages of autoregressive panel data models with individual effects is that the usual OLS method does not lead to consistent estimates for the parameters. This inconsistency is due to the asymptotic correlation between the lagged endogeneous variables and the disturbances because of the presence of the individual effects. To solve this problem, one can either rewrite the model in first differences or in orthogonal deviations. Differencing removes the primary cause of the inconsistency but induces a moving average type serial correlation in the disturbances of the transformed model. Therefore, even then OLS does not lead to consistent estimates. Using orthogonal deviations, when panels are small, induces a correlation between the lagged dependent variable and the transformed error. Hence, again the OLS estimators are not consistent with this latter transforming technique.

The GMM provides a convenient framework for obtaining asymptotically efficient estimators in dynamic panel data settings. First-differenced GMM estimators for the AR(1) panel data model were developed by Arellano and Bond (1991), Holtz-Eakin *et al* (1988), Anderson and Hsiao (1981). We mainly concentrate on Arellano and Bond's (1991) work. They difference the AR(1) model and use all the orthogonality conditions that exist between lagged values of the endogenous variables and the disturbances, the so-called DIF conditions.[2] By doing this they obtain optimal linear GMM estimators under relatively weak auxiliary assumptions about the heterogeneity and error term processes. These estimators are the most efficient in the class of instrumental variables' estimators (Arellano and Bond, 1991). However, it has been extensively documented that if the series are highly autoregressive, the GMM estimators based on DIF conditions have large finite-sample bias and poor precision in simulation studies (see Alonso-Borrego and Arellano, 1999; Blundell and Bond, 1998). One response to these limitations has been to consider further moment restrictions. Ahn and Schmidt (1995) consider non-linear moment conditions implied by the standard error components formulation. Blundell and Bond (1998) propose further restrictions on the initial conditions process, the so-called LEV conditions[3]. Our analysis considers DIF and LEV moment conditions.[4]

In this Chapter we compare the finite-sample size properties of two overidentifying restrictions tests: the GMM test, usually referred in this context as the Sargan test, and that based on EL and efficient bootstrap critical values, *i.e.* the EL-bootstrap test.[5]

The finite-sample behaviour of the Sargan test in an AR(1) dynamic panel data setting has been the subject of prior study. Among others, the work of Brown and Newey (2001) and

[2]We will define DIF moment conditions in Section 4.3.

[3]We will define LEV moment conditions in Section 4.4.

[4]The system formed by both DIF and LEV conditions is known as SYS moment conditions.

[5]These tests were introduced in Chapter 1, Sections 1.9.1 and 1.8.1.

Bowsher (2000, 2000a) have contributed to this literature. Brown and Newey (2001) also examine the GMM-bootstrap overidentification statistic as defined in Chapter 1, Section 1.9.1. They do not find a significant improvement in accuracy for the Sargan test when using efficient bootstrap critical values as an alternative to asymptotic critical values.[6] Bowsher (2000a) devotes a whole chapter of his PhD thesis to examining tilting parameter alternatives to the Sargan statistic.[7] His findings show that tilting parameter tests of overidentifying restrictions have worse finite-sample properties than the Sargan test in the context of the AR(1) dynamic panel data model. Although both tests are sensitive to the number of T –the time periods– becoming large, tilting parameter tests can be very oversized in panels where the Sargan test is well behaved.

We are not aware of any other study which has assessed the size properties of the EL-bootstrap overidentification test within dynamic panel data models. Hence, we concentrate on analysing in depth this statistic and compare it to the conventional two-step GMM overidentification test. The relevance of extending EL to this setting is evident because empirical applications which deal with dynamic panel data models are numerous. Moreover, given the already defined limitations of GMM estimators (large sample biases if series are highly autoregressive and worse size properties of its Sargan test as T increases) it is worthwhile to look for estimation alternatives to GMM.

The rest of the Chapter is organized as follows. Section 4.2 reviews autoregressive models with individual effects and lays out the underlying assumptions. We concentrate on an

[6] Although the improvement is substantial for the coverage probability of the confidence interval.
[7] Tilting parameter tests were introduced by Imbens *et al* (1998).

AR(1) process since the main insights generalize in a straightforward way to higher order multivariate cases. Section 4.3 and Section 4.4 present the moment equations, the so-called DIF and LEV conditions, implied by the model's assumptions. In Section 4.5, we analyze the finite-sample size properties of overidentification statistics through Monte Carlo experiments. An empirical application on an AR(1) univariate panel data model with individual effects using the cash-flow series of 174 firms in the United States from 1981 to 1985 is carried out in Section 4.6. Conclusions are then presented.

4.2 The Model

We consider a first-order univariate autoregressive panel data model of the form

$$y_{it} = \rho y_{i,t-1} + u_{it}, \tag{4.1}$$

$$u_{it} = \eta_i + v_{it}; \tag{4.2}$$

$$for \;\; i = 1, 2, ..., n \;\; and \;\; t = 2, , ..., T,$$

where y_{it} is an observation on some series for individual i in period t, η_i is an unobserved individual-specific time-invariant effect which allows for heterogeneity and v_{it} is a disturbance term.

We assume that n is large, T is fixed, $|\rho| < 1$, and η_i and v_{it} are independently distributed across i.

The following standard assumptions are usually made in connection with (4.1) (Ahn and Schmidt, 1995):

(A1) $E(\eta_i) = 0$, $E(v_{it}) = 0$, $for\ t = 2, ..., T$ and $\forall\ i$.

(A2) $E(v_{it}v_{is}) = 0$, $\forall\ t \neq s$ and $\forall\ i$.

(A3) $E(v_{it}\eta_i) = 0$, $for\ t = 2, ..., T$ and $\forall\ i$.

(A4) $E(y_{i1}v_{it}) = 0$, $for\ t = 2, ..., T$ and $\forall\ i$.

$(A2)$ states that the $v_{it}^{'}s$ are not serially correlated and $(A4)$ specifies that the initial conditions y_{i1} are predetermined. We will also impose $(A5)$, which is discussed in Section 4.4.

Given these assumptions, the OLS estimator of ρ in the level equation (4.1) is inconsistent. The reason is that the explanatory variable, $y_{i,t-1}$, is positively correlated with the error term, u_{it}, due to the presence of the individual effects. Mátyás and Sevestre (1996) show that this correlation does not disappear as the number of individuals in the sample gets larger. Standard results for omitted variables biases indicate that the OLS levels estimator is biased upwards.

The so-called Within Groups estimator eliminates this source of inconsistency by transforming the equation to eliminate η_i. Specifically, the original observations are expressed as deviations from the mean values of y_{it}, $y_{i,t-1}$, η_i and v_{it} across the $T-1$ observations for each individual i. OLS is then used to estimate ρ from

$$y_{it} - \overline{y}_i = \rho \left(y_{i,t-1} - \overline{y}_{i\_1} \right) + v_{it} - \overline{v}_i,$$

where $\overline{y}_i$, $\overline{y}_{i\_1}$, $\overline{\eta}_i$ and $\overline{v}_i$ are the mean values.

In panels where the number of time periods available is small, this transformation induces a correlation between the transformed lagged dependent variable and the transformed error term. Nickell (1981) shows that this correlation is negative. Standard results for omitted variables biases indicate that the Within Groups estimator is biased downwards.[8]

There are two approaches discussed in literature in which one can proceed to tackle the inconsistency of OLS and Within Groups estimators. The first uses a kind of Two-Stage Least Squares estimator as proposed by Balestra and Nerlove (1966). The second uses instrumental variables as proposed by Arellano and Bond (1991) and Anderson and Hsiao (1982, 1981). In what follows we focus on Arellano and Bond's (1991) work and on extensions provided by Blundell and Bond (1998).

4.3 DIF Moment Conditions

Assumptions $(A1)$ – $(A4)$ imply moment conditions that are sufficient to identify and estimate ρ for $T \geq 3$.

Applying first differences to (4.1) yields

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta v_{it}, \tag{4.3}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ and $\Delta v_{it} = v_{it} - v_{i,t-1}$,

for $i = 1, ..., n$ and $t = 3, ..., T$.

[8] Note that the OLS and Within Groups estimators are biased in opposite directions.

Equation (4.1) together with assumptions $(A1)-(A4)$ imply the following $m_d = .5\,(T-1)\,(T-2)$ linear moment restrictions

$$E\left(y_{i,t-s}\Delta v_{it}\right) = 0 \quad for \quad t = 3, ..., T \;\text{ and }\; s \geq 2. \tag{4.4}$$

These equations are known as DIF moment conditions because they involve the use of lagged levels of y_{it} as instruments for the first differenced equations. They can be expressed more compactly as

$$E\left(Z_{di}^{\tau}\Delta v_i\right) = 0, \tag{4.5}$$

where Z_{di} is the $(T-2) \times m_d$ matrix of instruments given by

$$Z_{di} = \begin{bmatrix} y_{i1} & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & y_{i1} & \ldots & y_{iT-2} \end{bmatrix}, \tag{4.6}$$

and Δv_i is the $(T-2)$ vector

$$\Delta v_i = \left(\Delta v_{i3}, \Delta v_{i4}, ..., \Delta v_{iT}\right)^{\tau}. \tag{4.7}$$

The subindex d emphasizes the fact that these are instruments for the differenced equations.

4.4 SYS Moment Conditions

By $(A1)-(A4)$ and the additional assumption on initial conditions

(A5) $E\left(\eta_i \Delta y_{i2}\right) = 0 \;\forall\, i,$

lagged differences are valid instruments for equations in levels (Blundell and Bond, 1998).

Thus, the further $m_l = (T-2)$ moment conditions

$$E\left[u_{it}\Delta y_{i,t-1}\right] = 0 \text{ for } t = 3, ..., T \text{ and } \forall\ i, \tag{4.8}$$

are available. These can be expressed as

$$E\left(Z_{li}^{T}\ u_i\right) = 0, \tag{4.9}$$

where Z_{li} is the $(T-2) \times m_l$ matrix of instruments given by

$$Z_{li} = \begin{bmatrix} \Delta y_{i2} & 0 & \cdots & \cdots & 0 \\ 0 & \Delta y_{i3} & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \\ \vdots & \vdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \Delta y_{iT-1} \end{bmatrix}, \tag{4.10}$$

and u_i is the $(T-2)$ vector $(u_{i3},\ u_{i4},\ ..., u_{it})^{T}$.

We refer to (4.9) as the LEV moment conditions since they involve the use of lagged differences as instruments for equations in levels. Note that the subindex l is used to emphasize the fact that these instruments are valid for equations in levels.

Since y_{i2} is specified by our model given y_{i1}, $(A5)$ is a restriction on the initial conditions process generating y_{i1}. Let $y_{i1} = \dfrac{\eta_i}{1-\rho} + \varepsilon_{i1}$, so that ε_{i1} is the initial deviation from the long run mean of the y_{it} process. Then, necessary and sufficient conditions for $(A5)$ are given by

$$E\left(\varepsilon_{i1}\eta_i\right) = 0. \tag{4.11}$$

The initial deviation from the long run mean must be uncorrelated across individuals with the level of that long run mean. (4.11) is satisfied if an infinite past is assumed for the dynamic process in (4.1) or by any initial deviation from $\frac{\eta_i}{1-\rho}$ which is randomly distributed across individuals.

The system of moment equations formed by the DIF moment conditions, (4.5), and the LEV moment conditions, (4.9), are the so-called SYS conditions. These are the $.5\,(T+1)\,(T-2)$ + $(T-2)$ equations

$$E\left(Z_i^{*\tau}\ \Delta v_i^*\right) = 0, \tag{4.12}$$

where Z_i^* is the instrument matrix given by

$$Z_i^* = \begin{pmatrix} Z_{di} & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & 0 & \dots & 0 \\ 0 & 0 & \Delta y_{i3} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \Delta y_{i,T-1} \end{pmatrix} = \begin{pmatrix} Z_{di} & 0 \\ 0 & Z_{li} \end{pmatrix}, \tag{4.13}$$

and

$$\Delta v_i^* = \left(\Delta v_i^\tau,\ u_{i3},\ u_{i4}, \dots u_{iT}\right)^\tau. \tag{4.14}$$

4.5 Finite-Sample Size Properties of Overidentification Tests

In what follows we employ Monte Carlo experiments to study the finite-sample size properties of two overidentification tests within a dynamic panel data framework. We analyse the EL-bootstrap overidentification test, W_j^b, and two-step GMM test, J_{2GMM}, based on DIF and SYS moment conditions. These tests are used to evaluate the null hypothesis that there is a value of ρ consistent with $E\left(Z_{di}^{\tau}\ \Delta v_i\right) = 0$ for DIF estimation and $E\left(Z_i^{*\tau}\ \Delta v_i^{*}\right) = 0$ for SYS estimation. They are therefore model specification tests since our maintained AR(1) model implies these moment conditions. Under the null hypothesis the two statistics have a $\chi^2_{(m-1)}$ distribution asymptotically. Where $m = .5\,(T+1)\,(T-2)$ for DIF estimation and $m = .5\,(T+1)\,(T-2) + (T-2)$ for SYS estimation.

To revise how the tests are calculated, we introduce the following notation. Let the $N\,(T-2) \times\ 1$ vector $\Delta v = (\Delta v_1^{\tau} ..., \ \Delta v_N^{\tau})^{\tau}$ be formed by vertically stacking the Δv_i for $i = 1, ...N$. Similarly, let $\Delta y_i = (\Delta y_{i3}, ..., \Delta y_{iT})^{\tau}$ and $\Delta y = (\Delta y_i^{\tau}, ..., \Delta y_N^{\tau})^{\tau}$. Also define $\Delta y_{i,-1} = (\Delta y_{i2}, ..., \Delta y_{i,T-1})^{\tau}$ and its stacked version Δy_{-1}. Let $Z_d = (Z_{d1}^{\tau} .....Z_{dN}^{\tau})^{\tau}$ be the $N\,(T-2) \times m_d$ matrix formed by vertically stacking the instrument matrices Z_{di} used for GMM-DIF estimation.

The one-step GMM-DIF estimator, $\overline{\rho}$, is obtained from

$$\overline{\rho} = \min_{\rho}\left(\Delta v^{\tau} Z_d W Z_d^{\tau} \Delta v\right) = \left(\Delta y_{-1}^{\tau} Z_d W Z_d^{\tau} \Delta y_{-1}\right)^{-1} \left(\Delta y_{-1}^{\tau} Z_d W Z_d^{\tau} \Delta y\right)$$

where W is the identity matrix. The two-step GMM-DIF estimator, $\widehat{\rho}_{2GMM}$, may be calculated by setting

$$\widehat{W} = \left(n^{-1}\sum Z_{di}^{\tau}\Delta\widehat{v}_i^{*}\Delta\widehat{v}_i^{*\tau}Z_{di}\right), \tag{4.15}$$

where $\Delta\widehat{v}_i^{*} = \Delta y_i - \overline{\rho}\ \Delta y_{i,-1}$.

The GMM overidentifying statistic based on the DIF moment conditions is denoted by

$$J_{2GMM} = \Delta\widehat{v}^{\tau} Z_{di}\left(\sum Z_{di}^{\tau}\Delta\widehat{v}_i\Delta\widehat{v}_i^{\tau}Z_{di}\right)^{-1} Z_{di}\Delta\widehat{v},$$

where $\Delta\widehat{v} = \Delta y - \ \widehat{\rho}_{2GMM}\Delta y_{-1}$.

The procedure for obtaining the Sargan statistic based on SYS conditions is essentially the same. We proceed in an analogous way to GMM-DIF estimation but using Z_i^{*} in (4.13) and Δv_i^{*} in (4.14) instead of Z_{di} in (4.6) and Δv_i in (4.7).

W_j^b, is obtained as described in Chapter 1, Section 1.9.1. We define

$$g\left(y_i, \rho\right) = Z_{di}^{\tau}\Delta v_i,$$

for the statistic based on DIF moment conditions and

$$g\left(y_i, \rho\right) = Z_i^{\tau *}\Delta v_i^{*},$$

for that based on SYS moment conditions. Here, $y_i = \left(y_{i1}, y_{i2}, ...y_{iT}\right)^{\tau}$.

4.5.1 The Data Generating Process

We generate y_{it} as

$$y_{it} = \rho\ y_{i,t-1} + \eta_i + v_{it},$$

$$y_{i0} = \frac{\eta_i}{1-\rho} + e_i,$$

where $0 \leq \rho < 1$, $t = 1, 2, ..., T$ and $i = 1, ...n$.

η_i and v_{it} are $N(0,1)$ random variables, $e_i \sim N(0, 1/1-\rho^2)$ and the ν_{it} 's, $\eta_i' s$ and e_i are mutually independent. Note that y_{it} is stationary over time because of the specification of the equation describing y_{i0}.

Assumptions $(A1) - (A5)$ are thus satisfied and both the DIF moment conditions, (4.5), and the SYS moment conditions, (4.12), are valid for a DGP of this form. In other words, the null hypothesis that the true DGP is nested in the model given in equation (4.1) and assumptions $(A_1) - (A_5)$ is true.

Our aim is to examine the effects of varying the sample size and the dimensions of the panels within our estimations. We are also interested in assessing the implications of using weak and strong instruments[9] and the benefits, if any, of exploiting additional moment conditions in our calculations. The importance of each one is discussed below.

(i) Sample Size

Within dynamic panel data models, asymptotic theory is based on $n \to \infty$ (rather than on T). Given this, it is interesting to assess if the asymptotic approximation of the overidentifying restrictions tests improves as n increases.

(ii) Dimensionality

[9] When the lagged levels of the series are only weakly correlated with subsequent first differences, then the instruments available for the differenced equations are weak. This may arise when marginal processes for y_{it} are highly persistent or close to random walk processes.

The number of available moment conditions increases rapidly as T increases (keeping n fixed) due to the dependence on T^2 (see Table 4.1). Bowsher (2000*a*) examines the implications of dimensionality on the performance of Sargan tests and a tilting parameter overidentification statistic. His findings show that when the number of moment conditions increases, the size properties of both tests deteriorate and that the tilting parameter test of overidentifying restrictions has worse size properties than the Sargan test in the context of the AR(1) dynamic panel data model. We analyse whether this dimensionality problem is also found for the EL-bootstrap overidentifying statistic.

(iii) Strong and weak instruments

As we have previously discussed, Blundell and Bond (1998) illustrate that GMM-DIF estimators have pronounced bias in the presence of weak instruments. We are interested in examining if the GMM overidentifying statistic is also sensitive to high values of the autoregressive coefficient. In addition, we want to see if the size properties of the EL-bootstrap overidentification test are worse for weak instruments than for strong instruments.

(iv) SYS *versus* DIF conditions

Blundell and Bond (1998) find evidence of large benefits from introducing additional restrictions on the initial conditions of the AR(1) process (in terms of bias and/or precision) in GMM estimators, in the presence of highly persistent series. For example, they show that estimators based on SYS conditions are better than those based on DIF conditions. Specifically, for values of $\rho \geq .8$ they find that the DIF estimator has a pronounced downward bias. By adding LEV conditions into GMM estimation, biases are significantly reduced. In light of these results we examine whether these gains are extended to its overidentification test. In other words, we assess if the finite-sample size properties of overidentification tests based on SYS moment conditions are better than those based on DIF equations. We distinguish two opposite effects: a positive effect of incorporating correct moment equations into our calculations and a negative dimensionality effect. Here, we are interested in the overall effect of exploiting extra moment equations.

Number of estimating equations
Dynamic Panel Data

	DIF	SYS
	$\frac{1}{2}(T-1)(T-2)$	$\frac{1}{2}(T-1)(T-2)+(T-2)$
T=3	1	2
T=4	3	5
T=5	6	9
T=6	10	14
T=10	36	44

T is the time periods.

Table 4.1: Dynamic Panel Data - Number of Estimating Equations

All of our results are based on 5000 Monte Carlo replications with 1000 bootstrap trials in each experiment. We consider two sample sizes: $n = 100$ and $n = 175$;[10] four different values

[10]n=175 was chosen to match the sample size of our empirical example, given in Section 4.6.

for the autoregressive component: $\rho = \{.2, .5, .7, .9\}$; and three time periods: $T = \{4, 5, 6\}$. By considering these values of ρ we intend to cover representative stationary cases. It would have been more illustrative to have considered longer time periods to study the dimensionality effect. However, computational restrictions played a decisive part in this respect.

First, we investigate the implications of varying the sample size within our estimations using both DIF and SYS moment conditions. We concentrate on T=6.[11]

Empirical Levels of J-tests
Dynamic Panel Data
Sample Size effects

$T = 6$
DIF Moment Conditions

		W_j^b		J_{2GMM}	
Levels	ρ	$n = 100$	$n = 175$	$n = 100$	$n = 175$
.10		.0772	.0948	.0984	.1098
.05	.2	.0394	.0482	.0448	.0528
.01		.0080	.0094	.0086	.0100
.10		.1160	.0948	.1066	.1126
.05	.5	.0524	.0474	.0472	.0552
.01		.0062	.0074	.0076	.0088
.10		.1158	.0972	.1186	.1166
.05	.7	.0532	.0398	.0594	.0610
.01		.0050	.0226	.0084	.0136
.10		.1282	.1162	.1092	.1188
.05	.9	.0598	.0668	.0512	.0604
.01		.0108	.0134	.0070	.0128

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM; respectively. T is the time periods, ρ the autoregressive coefficient and n is the sample size.

Table 4.2: Finite-Sample Size Properties - Dynamic Panel Data (Sample Size Effects: DIF)

[11] Alternative time periods lead to similar conclusions to those corresponding to T=6. The results for the complete set of time periods are provided in Appendix 3.

Table 4.2 summarizes the rejection frequencies[12] of W_j^b and J_{2GMM} based on DIF moment conditions. Those corresponding to SYS estimation are reported in Table 4.3. Note that there is a single parameter to be estimated, ρ. The reference distributions for this specification are $\chi^2_{(9)}$ and $\chi^2_{(13)}$ for the DIF and SYS context; respectively (refer to Table 4.1).

The Tables show that increasing n does not necessarily lead to better size properties of both tests. Moreover, there are several cases in which the size distortions of both statistics increase as n increases.

Empirical Levels of J-tests
Dynamic Panel Data
Sample Size effects

$T = 6$
SYS Moment Conditions

Levels	ρ	W_j^b $n = 100$	W_j^b $n = 175$	J_{2GMM} $n = 100$	J_{2GMM} $n = 175$
.10		.1047	.1235	.1028	.1058
.05	.2	.0541	.0704	.0490	.0564
.01		.0095	.0190	.0090	.0112
.10		.1148	.1159	.1144	.1118
.05	.5	.0534	.0569	.0522	.0552
.01		.0104	.0100	.0086	.0120
.10		.0926	.0728	.1218	.1312
.05	.7	.0362	.0334	.0612	.0694
.01		.0062	.0034	.0134	.0166
.10		.1042	.1002	.1256	.1434
.05	.9	.0452	.0478	.0684	.0746
.01		.0062	.0078	.0100	.0176

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM; respectively. T is the time periods, ρ the autoregressive coefficient and n is the sample size.

Table 4.3: Finite-Sample Size Properties - Dynamic Panel Data (Sample Size Effects: SYS)

[12]Note that for W_j^b the rejection frequencies are the proportion of the simulated data test statistics that exceeds the bootstrap critical values.

Given these findings, we run an additional experiment for this specification ($n = 500, T = 6$, $\rho = \{.2, .5, .7, .9\}$, $m = 5000$ and 1000 bootstrap trials).[13] Tables 4.4 and 4.5 summarize the rejection frequencies for the analysis based on DIF and SYS moment conditions, respectively.

Empirical Levels of J-tests
Dynamic Panel Data
Sample Size effects

$T = 6$
DIF Moment Conditions

Levels	ρ	W_j^b $n = 100$	W_j^b $n = 500$	J_{2GMM} $n = 100$	J_{2GMM} $n = 500$
.10		.0772	.0952	.0984	.1010
.05	.2	.0394	.0495	.0448	.0520
.01		.0080	.0012	.0086	.0116
.10		.1160	.1015	.1066	.1033
.05	.5	.0524	.0510	.0472	.0510
.01		.0062	.0089	.0076	.0095
.10		.1158	.1142	.1186	.1090
.05	.7	.0532	.0570	.0594	.0555
.01		.0050	.0018	.0084	.0113
.10		.1282	.1140	.1092	.1290
.05	.9	.0598	.0530	.0512	.0642
.01		.0108	.0115	.0070	.0155

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM; respectively. T is the time periods, ρ the autoregressive coefficient and n is the sample size.

Table 4.4: Finite-Sample Size Properties - Dynamic Panel Data (Sample Size Effects: DIF n=500

[13]This is worthwhile because n=175 is not very large relative to n=100.

Empirical Levels of J-tests
Dynamic Panel Data
Sample Size effects

$T = 6$
SYS Moment Conditions

		W_j^b		J_{2GMM}	
Levels	ρ	$n = 100$	$n = 500$	$n = 100$	$n = 500$
.10		.1047	.1087	.1028	.1002
.05	.2	.0541	.0540	.0490	.0520
.01		.0095	.0110	.0090	.0114
.10		.1148	.1120	.1144	.1130
.05	.5	.0534	.0513	.0522	.0515
.01		.0104	.0121	.0086	.0100
.10		.0926	.0734	.1218	.1210
.05	.7	.0362	.0335	.0612	.0598
.01		.0062	.0044	.0134	.0180
.10		.1042	.1017	.1256	.1510
.05	.9	.0452	.053	.0684	.0920
.01		.0062	.0091	.0100	.0123

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM; respectively. T is the time periods, ρ the autoregressive coefficient and n is the sample size.

Table 4.5: Finite-Sample Size Properties - Dynamic Panel Data (Sample Size Effects: SYS n=500

Table 4.4 shows that some of our results improve –or are very similar to those corresponding to $n = 100$– as the sample size increases from $n = 100$ to $n = 500$, except for J_{2GMM} and $\rho = .9$. Interestingly, this is also the case for J_{2GMM} when the analysis is based on SYS moment conditions (see Table 4.5). In other words, when the instruments are weak $-\rho = .9-$ the size distortions for the Sargan statistic can increase as n increases. Also note that W_j^b is still undersized for $\rho = .7$ and $n = 500$ (compare Tables 4.3 and 4.5).

Now, we examine the effect of increasing the number of time periods, the so-called dimensionality effect, on the finite-sample size properties of both overidentifying statistics. We consider the weak instrument case, $\rho = \{.7, .9\}$.[14] Results based on DIF moment conditions are summarized in Table 4.6. The reference distributions for these tests are respectively $\chi^2_{(2)}$, $\chi^2_{(5)}$, $\chi^2_{(9)}$ for T=$\{4, 5, 6\}$. Results corresponding to SYS conditions are given in Table 4.7. Here, the reference distributions are respectively $\chi^2_{(4)}$, $\chi^2_{(8)}$, $\chi^2_{(13)}$ for T=$\{4, 5, 6\}$.[15]

For the experiments based on DIF moment conditions there is no evidence of a worsening in the size properties of both overidentification tests as T increases. However, when SYS conditions are used (see Table 4.7) there are several cases in which the asymptotic χ^2 approximation of the finite-sample distribution deteriorates for J_{2GMM} as T increases. These results are consistent with prior simulation evidence (see Table 3.2 of Bowsher, 2000a). The results for W_j^b show that there are some specifications for which this test becomes undersized as T increases. We observe that W_j^b is not particularly sensitive to variations in T.

[14] There is no evidence of a dimensionality effect for $\rho = \{.2, .5\}$. Refer to Appendix 3.

[15] To calculate the degrees of freedom refer to Table 4.1.

		Empirical Levels of J-tests Dynamic Panel Data Dimensionality Effect DIF Moment Conditions					
				n=100			
		T=4		T=5		T=6	
Levels	ρ	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}
.10		.1006	.1270	.0884	.1292	.1158	.1186
.05	.7	.0492	.0710	.0392	.0656	.0532	.0594
.01		.0110	.0204	.0048	.0136	.0050	.0084
.10		.1282	.1018	.1142	.1056	.1282	.1092
.05	.9	.0682	.0492	.0615	.0464	.0598	.0512
.01		.0154	.0068	.0071	.0072	.0108	.0070
				n=175			
.10		.0946	.1080	.1143	.1326	.0972	.1166
.05	.7	.0506	.0588	.0603	.0688	.0398	.0610
.01		.0084	.0152	.0126	.0128	.0226	.0136
.10		.1010	.0990	.1006	.1148	.1162	.1188
.05	.9	.0544	.0524	.0520	.0604	.0668	.0604
.01		.0118	.0096	.0110	.0114	.0134	.0128

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM. T is the time periods, n is the sample size and ρ is the autoregressive coefficient.

Table 4.6: Finite-Sample Size Properties - Dynamic Panel Data (DIF: Dimensionality Effect)

Empirical Levels of J-Tests Dynamic Panel Data Dimensionality Effect SYS Conditions							
		T=4		T=5		T=6	
		W_j^b	J_{2GMM}	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}
Levels	ρ						
				$n = 100$			
.10		.0992	.1164	.1024	.1196	.0926	.1218
.05	.7	.0480	.0578	.0500	.0592	.0362	.0612
.01		.0074	.0102	.0094	.0106	.0062	.0134
.10		.0994	.1060	.0954	.1208	.1042	.1256
.05	.9	.0474	.0540	.0410	.0578	.0452	.0684
.01		.0086	.0096	.0036	.0100	.0062	.0100
				$n = 175$			
.10		.1106	.1094	.0910	.1072	.0728	.1312
.05	.7	.0608	.0614	.0396	.0556	.0334	.0694
.01		.0114	.0146	.0038	.0114	.0034	.0166
.10		.1050	.1220	.0924	.1198	.1002	.1434
.05	.9	.0636	.0636	.0420	.0638	.0478	.0746
.01		.0136	.0136	.0064	.0146	.0078	.0176

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM; respectively. T is the time periods, ρ is the autoregressive coefficient and n is the sample size.

Table 4.7: Finite-Sample Size Properties - Dynamic Panel Data (SYS: Dimensionality Effect)

The effects of weak and strong instruments are analysed in Table 4.8. Consider the DIF moment conditions, n=100 and T=$\{4,5,6\}$[16]. The reference distributions are $\chi^2_{(2)}$, $\chi^2_{(5)}$ and $\chi^2_{(9)}$.[17]

[16] Our main interest is to assess the effects of weak instruments using DIF moment conditions. Refer to Appendix 3 for the results corresponding to SYS estimation.

[17] To calculate the degrees of freedom refer to Table 4.1.

Empirical Levels of J-tests
Dynamic Panel Data
Weak and Strong Instruments

		n=100					
		T=4		T=5		T=6	
Levels	ρ	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}
.10		.1006	.1084	.1044	.1042	.0772	.0984
.05	.2	.0526	.0532	.0576	.0526	.0394	.0448
.01		.0124	.0086	.0108	.0084	.0080	.0086
.10		.1196	.1074	.1040	.1114	.1160	.1066
.05	.5	.0604	.0564	.0530	.0564	.0524	.0472
.01		.0134	.0148	.0124	.0126	.0062	.0076
.10		.1006	.1270	.0884	.1292	.1158	.1186
.05	.7	.0492	.0710	.0392	.0656	.0532	.0594
.01		.0110	.0204	.0048	.0136	.0050	.0084
.10		.1282	.1018	.1142	.1056	.1282	.1092
.05	.9	.0682	.0492	.0615	.0464	.0598	.0512
.01		.0154	.0068	.0071	.0072	.0108	.0070

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM. T is the time periods, n is the sample size and ρ is the autoregressive coefficient.

Table 4.8: Finite-Sample Size Properties - Dynamic Panel Data (Weak and Strong Instruments: DIF n=100)

The main result is that for $n = 100$ and $\rho = .7$ the asymptotic approximation for J_{2GMM} is the worst. In Monte Carlo results reported by Blundell and Bond (1998) GMM estimators are biased for highly autoregressive series and these biases are dramatic for $\rho = .9$. However, note that J_{2GMM} has better sizes for $\rho = .9$ than for $\rho = .7$. Moreover, the results for $\rho = .2$ do not differ to those corresponding to $\rho = .9$ in a large extent. We find that W_j^b is not very sensitive to the problem of weak instruments (although we observe that W_j^b is more oversized for T=4 and $\rho = .9$).

The objective of the next set of simulations is to test whether there is an improvement in accuracy for the overidentification tests from using additional moment conditions. We are particularly interested on assessing whether this is the case for a weak instruments specification: $\rho = .9$. For DIF estimation and T=$\{4,5,6\}$; the reference distributions are $\chi^2_{(2)}$, $\chi^2_{(5)}$, $\chi^2_{(9)}$. For the analysis based on SYS moment equations and T=$\{4,5,6\}$; these are $\chi^2_{(4)}$, $\chi^2_{(8)}$, $\chi^2_{(13)}$.[18] Results for W^b_j are provided in Table 4.9 and those for J_{2GMM} are reported in Table 4.10.

Empirical Levels of W^b_j
Dynamic Panel Data
DIF versus SYS

	T=4		T=5		T=6	
Levels	DIF	SYS	DIF	SYS	DIF	SYS
			$n = 100$ and $\rho = .9$			
.10	.1282	.0994	.1142	.0954	.1282	.1042
.05	.0582	.0474	.0615	.0410	.0598	.0452
.01	.0154	.0086	.0071	.0036	.0108	.0062
			$n = 175$ and $\rho = .9$			
.10	.1010	.1050	.1006	.0924	.1162	.1002
.05	.0544	.0470	.0520	.0420	.0668	.0478
.01	.0118	.0108	.0110	.0064	.0134	.0078

W^b_j is an overidentification tests based on EL-bootstrap. T is the time periods, n is the sample size and ρ is the autoregressive coefficient

Table 4.9: Finite-Sample Size Properties - EL-bootstrap: Dynamic Panel Data (DIF versus SYS)

Although the size properties of W^b_j at the .10 and .05 levels are better for SYS than for DIF

[18] To calculate the degrees of freedom refer to Table 4.1.

moment conditions for n=100, this is not the case for n=175.

Our results for J_{2GMM} in Table 4.10 are unexpected. The finite-sample size properties of the Sargan tests based on SYS moment conditions are worse than those based on DIF conditions. These findings have the implication that as we add moment conditions to reduce the sample bias of GMM estimators due to weak instrumentation, as recommended by Blundell and Bond (1998), we could negatively be affecting the finite-size properties of its Sargan test.

Empirical Levels of J_{2GMM}
Dynamic Panel Data
DIF versus SYS

	T=4		T=5		T=6	
Levels	DIF	SYS	DIF	SYS	DIF	SYS
			$n = 100$ and $\rho = .9$			
.10	.1018	.1060	.1056	.1208	.1092	.1256
.05	.0492	.0540	.0464	.0578	.0512	.0684
.01	.0068	.0096	.0072	.0100	.0070	.0100
			$n = 175$ and $\rho = .9$			
.10	.0990	.1220	.1148	.1198	.1188	.1434
.05	.0524	.0636	.0604	.0638	.0604	.0746
.01	.0096	.0136	.0114	.0146	.0128	.0176

J_{2GMM} is an overidentification tests based on the two-step GMM estimator.
T is the time periods, n is the sample size and ρ is the autoregressive coefficient

Table 4.10: Finite-Sample Size Properties - Two-step GMM: Dynamic Panel Data (DIF versus SYS)

4.6 Empirical Application

4.6.1 Data

The data set we use was kindly provided by Bronwyn Hall. It is a balanced panel of 174 firms for the United States for 1978-1989. Hall *et al* (1998) use this data set to test for causal relationship among sales and cash-flow, and research and development and investment. These 174 firms belong to the science-based industries and include Chemicals, Pharmaceuticals, Electrical Machinery, Computing Equipment, Electronics, and Scientific Instruments. The original data set consists of 863 firms and the variables analysed are sales, research and development, investment, cash-flow and employment. Hall *et al* (1998) apply the following "cleaning" rules:

(i) Only firms with growth rates between -90% and 900% were considered.

(ii) In order to remove erroneous data values, firms with at least one of the following characteristics were eliminated:

- Sequential employment and/or sales growth rates that were large, *e.g.* below -50% or above 100%, and alternate in sign.
- Sequential investment and/or cash-flow growth rates that were large, *e.g.* below -80% or above 400%, and alternate in sign.
- Sequential research and development growth rates that were large, *e.g.* between -67% and 200%, and alternate in sign.

(iii) Firms with negative cash-flows and with jumps in observations were removed from the data set.

We choose the years 1981-1985. This reflects the desire to have a short panel, *e.g.* $T = 5$. This leaves us with a total of 870 observations for each series. The series were deflated as in Hall *et al* (1998). Among the 5 different variables we looked for those compatible with models where heterogeneity across firms is summarized by an individual effect. Another important feature that we explored was the stationarity of the process and the order of the autoregressive component. We now present the analysis for the series corresponding to cash-flow.

4.6.2 Cash-flow: Descriptive Statistics

Empirical First and Second Moments real log (cash-flow)							
	Mean	St Dev	Correlation Matrix				
Year			**1981**	**1982**	**1983**	**1984**	**1985**
1981	4.2033	1.9862	1	.9818	.9745	.9645	.9598
1982	4.1548	1.9806	.9818	1	.9835	.9754	.9671
1983	4.3062	1.9158	.9745	.9835	1	.9877	.9750
1984	4.4531	1.8744	.9645	.9754	.9877	1	.9829
1985	4.3888	1.8588	.9598	.9671	.9750	.9829	1

Table 4.11: Cash-flow Descriptive Statistics

Table 4.11 shows the means, standard deviations, and autocorrelations for the cash-flow series. In general, the means and standard deviations of cash-flow do not change much over time. Note that the correlation matrix illustrates the fact that cash-flow is highly

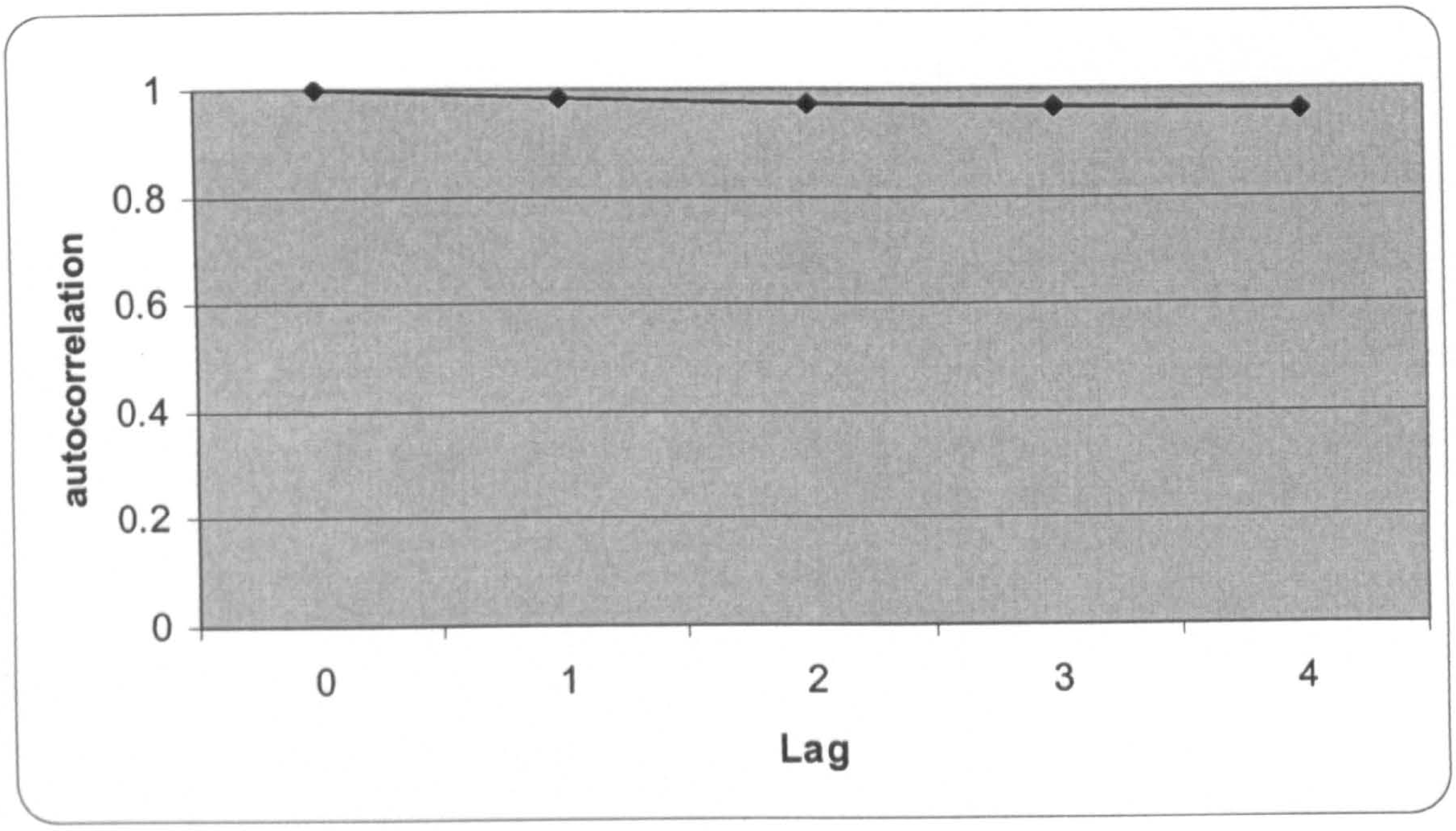

Figure 4.1: Autocorrelation Cash-flow

autocorrelated.

Figure 1 shows the autocorrelation plots for the cash-flow series. It is clear from this plot that the autocorrelation decays very slowly, which suggests either that the time series process used to describe these data will have a root close to one, or that the series are dominated by permanent differences in the level of the variables across firms.

4.6.3 Cash-flow: The Model

The model is

$$cf_{it} = \sum_{k=1}^{K} \rho_k cf_{it-k} + u_{it}, \tag{4.16}$$

$$u_{it} = \eta_i + v_{it} \text{ for } i = 1, ...174 \text{ and } t = 1, ...5;$$

where cf_{it} is the logarithm of real cash-flow of the i^{th} individual at time t.

Before analyzing the results it is important to review some key points.

1. First, in micro panels properties such as orders of integration and cointegration are crucial for identification of econometric models. Where differencing transformations are employed to eliminate unobserved individual effects, identification requires the existence of instrumental variables that are correlated with first-differences of the series. In the case of a pure random walk, lagged values of the series are uncorrelated with first-differences, thus the widely used first-differenced instrumental variables estimators will provide no information on the parameter of interest. In other words, the presence of a unit root will invalidate the commonly used GMM specification.[19]

   It is therefore important to assess the time series properties of the series under consideration. In this regard, our analysis is greatly influenced by the studies of Bond *et al* (2002) and Hall and Mairesse (2002). They investigate the properties of several unit roots tests in short panel data. Their findings illustrate that a test based on the model estimated under the null of a unit root (that is, where the OLS can be used be-

[19]In the presence of a unit root, the identifiability of GMM is preserved if this method is based on quadratic moments (see Alvarez and Arellano (2004)).

cause there are no "individual effects") provides a simple robust test with high power. We rely on this test, denoted by BNW, and allow for heteroscedasticity by using a Seemingly Unrelated Regression (SUR) framework with each year being an equation (as in Hall and Mairesse, 2002).

2. Second, assumption $(A2)$ states that there is no serial correlation in the $v_{it}'s$. This is the crucial assumption allowing the identification of ρ in our model. If the assumption of no serial correlation is not valid then the null hypothesis, $E\left(Z_{di}^{T}\Delta u_i\right) = 0$, will be false since the moment conditions would not hold. Thus, it is important to report tests of serial correlation in the first differenced residuals. If the errors in levels are serially independent, those in first-differences will exhibit first-order –but not second– serial correlation. Moreover, the first-order serial correlation coefficient must be equal to -0.5. An informal but often useful test diagnostic is provided by inspecting the autocorrelation matrix of the errors in first differences (see Chapter 6 of Arellano, 2003). Arellano and Bond (1991) propose formal tests of serial correlation: m_2 and m_1. The former tests for lack of second order serial correlation in the first differenced residuals. This will be certainly the case if the errors in the model in levels are not serially correlated, but also if the errors in levels follow a random walk process. To discriminate between the two situations we calculate an m_1 statistic to test for lack of first-order serial correlation in the differenced residuals (see Arellano and Bond, 1991).

Summing up, if the disturbances v_{it} are not serially correlated there should be evidence of significant negative first order serial correlation in differenced residuals and no evidence of second order serial correlation in the differenced residuals. The statistics m_1 and m_2 are based on the standardized average residual autocovariance. These tests are asymptotically distributed as $N(0,1)$ under the null of no autocorrelation.

3. Finally, the fact that the OLS and Within Groups estimators are likely to be biased in opposite directions is very useful (recall that OLS is biased upwards and Within Groups is biased downwards). Thus if the cash-flow series is well represented by an autoregressive model with individual effects, the GMM estimator will lie between the OLS and Within Groups estimator (or at least not be significantly higher than the former or significantly lower than the latter (Bond, 2002)).

We now analyse in depth an AR(1) model.[20] A constant term and time dummies are included. Our estimations are solely based on DIF conditions. Most of our calculations are performed using DPD98 for GAUSS and our own GAUSS programs.

For an AR(1) model the DIF moment equations are $E\left(Z_{di}^T \Delta v_i\right) = 0$. Here Z_{di} is the matrix of instruments given by

$$Z_{di} = \begin{bmatrix} cf_{i1} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & cf_{i1} & cf_{i2} & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & cf_{i1} & cf_{i2} & cf_{i3} & 1 & 0 & 1 \end{bmatrix}; \tag{4.17}$$

[20]We first examined an AR(2) model (a higher order was not studied since we have a short series: $T = 5$, and three cross sections are already lost in constructing lags and taking first differences for this specification). We discriminated between an AR(1) model and an AR(2) model using conventional procedures. Main results for an AR(2) model are given in Appendix 4.

and Δv_i is denoted by $(\Delta v_{i3}, \Delta v_{i4}, \Delta v_{i5})^T$,

where

$$\begin{aligned}\Delta v_{i3} &= cf_{i3} - cf_{i2} - \zeta_1 1 - \rho_1 (cf_{i2} - cf_{i1}), \\ \Delta v_{i4} &= cf_{i4} - cf_{i3} - \zeta_1 1 - \rho_1 (cf_{i3} - cf_{i2}) - \zeta_2 1, \\ \Delta v_{i5} &= cf_{i5} - cf_{i4} - \zeta_1 1 - \rho_1 (cf_{i4} - cf_{i3}) - \zeta_3 1.\end{aligned} \tag{4.18}$$

ζ_1 is the constant term and ζ_2 and ζ_2 are the coefficients of the time dummies for 1984 or $T = 4$ and 1985 or $T = 5$; respectively. Note that these coefficients are multiplied by a $(n \times 1)$ row vector of ones.

Therefore, we have 9 moment equations and 4 parameters: ζ_1, ζ_2, ζ_3 and ρ_1; to be estimated. Results for the levels of the OLS, Within Groups and GMM estimators are reported in Table 4.12.

4.6.4 Cash-flow: Results

Since the GMM estimate lies between the OLS estimate and the Within Groups estimate, we have some evidence that the logarithm of cash-flow is well represented by a dynamic AR(1) model with individual effects. For example, the OLS is considerably higher than the Within Groups estimate and the GMM lies between both.

Dynamic Panel Data AR(1) Cash-flow	OLS LEVELS	WITHIN GROUPS	GMM DIF
cf_{t-1}	.9677 (.000)	.1888 (.006)	.6729 (.000)
m_1	−2.299 (.021)	−3.090 (.002)	−3.263 (.001)
m_2	−.4859 (.627)	−.6345 (.526)	.2628 (.793)
BNW	−5.22 [*std.error* .00532]		
J_{2GMM} [5 df]			10.01 (.075)

p-values are reported inside parenthesis. df refers to the degrees of freedom. m_1 and m_2 test for serial correlation. J_{2GMM} is the Sargan test. BNW is a unit root test suggested by Bond et al (2002).

Table 4.12: AR(1) Cash-flow

To test for serial correlation we examine informal and formal tests. A serial matrix for cash-flow based on GMM residuals is

$$\begin{pmatrix} 1 & -.5501 & .0296 \\ -.5501 & 1 & -.4122 \\ .0296 & -.4122 & 1 \end{pmatrix},$$

which broadly conforms to the expected pattern.

Formal tests of serial correlation are provided by the m_1 and m_2 statistics in Table 4.12. m_1

and m_2 are not reliable for the OLS and the Within Groups methods because the estimators of ρ and hence the estimates of the first-differenced residuals are likely to be biased. The serial correlation tests based on GMM are consistent with our assumptions: m_1 is negative and significant whilst m_2 is insignificant.

To test for unit roots, $H_0 : \rho = 1$, we follow Bond *et al* (2002). Our OLS test is based on the following model

$$y_{it} = \rho_k y_{it-1} + v_{it} \text{ for } i = 1, ..., 174 \text{ and } t = 1, ..., 5,$$

$$E\left[v_i v_i'\right] = \Omega,$$

where $v_i = (v_{i1}, v_{i2}, ...v_{i5})$.

Under the null BNW has an asymptotic standard normal distribution. The method of estimation is SUR with a weighting matrix based on the first stage estimate of Ω. According to BNW, in Table 4.12, there is no evidence of unit roots.

The Sargan statistic, J_{2GMM}, is 10.01 with a p-value equal to .075. It would certainly be appealing to have a stronger result (a higher p-value) to assess whether the AR(1) model is well specified for cash-flow. Because our simulations showed that the GMM statistic rejects too frequently for this particular specification, *e.g.* $n = 175$ and $\rho = .7$ (refer to Table 4.6), we now consider the EL-bootstrap overidentification test. Although W_J^b is also oversized, this is to a lower extent. Hence, if the validity of the moment equations is not rejected by

W_j^b at any conventional significance level we would have stronger evidence to support the hypothesis that the AR(1) model is well defined for the cash-flow series.

We also note that it is likely that $\widehat{\rho}_{2GMM} = .6729$ is biased downwards (see the simulation evidence given by Blundell and Bond, 1998). Blundell and Bond (1998) show that for persistent series GMM-SYS estimators were better than those obtained through DIF conditions. Using GMM-SYS estimation for our cash-flow series yields a ρ estimate close to .90. From Table 4.6 it is the case that although J_{2GMM} and W_j^b are both oversized for $n = 175$ and $\rho = .9$, the latter statistic over-rejects to a lesser extent. Hence, it is still worthwhile to report the EL-bootstrap overidentification statistic in this case.

To calculate the EL-bootstrap overidentification test we use the same set of estimating equations that were used for the GMM estimations.

1000 bootstrap trials are considered and the coefficients given in Table 4.12 are taken as the initial values in our algorithms. From our experiments we obtain the following efficient bootstrap critical values for 10%, 5% and 1%: $\{11.12, 12.23, 16.59\}$. These values are larger than the asymptotic $\chi^2_{(5)}$ values: $\{9.24, 11.07, 15.09\}$.

The EL-bootstrap overidentification test yields a statistic

$$W_j^b = 10.90,$$

which is smaller than the efficient bootstrap critical values. Therefore, the validity of the moment equations is not rejected at any conventional significance level.

We can now can conclude that there is evidence that the logarithm of cash-flow is well represented by an AR(1) model with individual effects.

4.7 Conclusions

The objectives of this chapter were twofold:

- To extend EL estimation to a widely used framework: dynamic panel data models.
- To examine EL as an alternative to GMM estimation in the context of autoregressive models with individual effects.

We studied the finite-sample size properties of the overidentification test based on EL and bootstrap, which we referred to as EL-bootstrap, and compared them to those of the Sargan statistic.

Asymptotic theory, in the context that we examine, is based on the sample size rather than on the number of time periods. We analysed the effect of increasing the sample size within the finite-sample size properties of both overidentification tests. We found no indication of better size properties for both tests based on DIF and SYS conditions as n increases from $n = 100$ to $n = 175$. We carried out an extra experiment for $T = 6$ and assessed the effect of increasing the sample size from $n = 100$ to $n = 500$. Our simulations do not uniformly support the conclusion that increasing the sample size leads to better size properties for the overidentification tests. Moreover we found some evidence of worse size properties for the J_{2GMM} for highly autoregressive series, $\rho = .9$, within both sets of moment conditions. This finding presents an area for future research.

According to Bowsher (2000a), tilting parameter tests of overidentifying restrictions have worse size properties than the conventional Sargan test in the context of the AR(1) dynamic panel data model. The former tests appear to be more sensitive to the problem of T becoming large and can be very oversized in panels where the Sargan test is well behaved. Therefore, we analysed the extent to which the dimensionality effect was also a problem for the EL-bootstrap statistic. For the three periods that we analysed and for the specifications of our experiments, there was no evidence of a size distortion effect in the size properties of this statistic.

Several simulation studies have found that for high values of the autoregressive coefficient, GMM estimators based on DIF conditions have large finite-sample bias and poor precision. It turned out that this might also be true for its Sargan test, for $\rho = .7$, as we found evidence of its finite-sample size properties being worse for this specification. However, the Sargan statistic has better sizes for $\rho = .9$ than for $\rho = .7$. Moreover, the results corresponding to $\rho = .2$ do not differ to those corresponding to $\rho = .9$ in a large extent. Our findings suggest that the finite-sample size properties of the EL-bootstrap statistic based on DIF conditions are not sensitive to weak instruments (except for T=4 and $\rho = .9$)

It has been widely documented that incorporating information relating to initial conditions is an effective way of reducing the sample bias and imprecision of GMM estimators in the weak instruments case. However, contrary to our initial expectations, our experiments show that the size-properties of the Sargan statistic can be worse for estimations based on

SYS conditions than for those based on DIF conditions. This means that while trying to reduce some of the bias in GMM estimators –due to the presence of persistent series– by incorporating additional conditions we could be negatively affecting the size properties of its overidentification test (this conclusions holds for $\rho = .9$). Thus, it was interesting to examine the extent to which this behaviour was also applicable to the EL-bootstrap statistic. We found some evidence of better size properties derived from exploiting additional moment conditions for n=100 and $\rho = .9$ (not for n=175).

Finally, we carried out an empirical application. We considered the cash-flow series of 174 firms from the United States from 1981-1985. Except for the Sargan statistic, the different tests that we studied –both formal and informal– provided strong evidence that pointed to cash-flow being well-represented as an AR(1) model with individual effects. The p-value of the Sargan statistic was only 7.5%. Our simulations showed that the Sargan test over-rejected the null hypothesis for the same sample size and the same number of time periods for our empirical example. Whereas even if the EL-bootstrap over-rejected for this specification, this was to a lesser extent. Hence, we calculated the EL-bootstrap statistic for the cash-flow series. The null hypothesis was not rejected at any conventional statistical level. Given these results, we have stronger evidence that supports cash-flow as being well represented by an AR(1) panel data model with individual effects.

Conclusions

This thesis outlined the theory of EL, provided numerical illustrations of its performance in a number of applications and examined EL as an alternative to GMM estimation.[21] The aim of our research was to provide some evidence of EL's practical value in econometrics.

An important element of EL estimation is the fact that the solutions to problems often cannot be written in closed form and must be computed numerically. We explored some of the standard computational aspects –such as the sensitivity of our estimations to poor starting values, how long our iterations take to achieve convergence and the accuracy of our results given different sample sizes– involved in the estimating procedures.

We presented EL as an alternative to GMM estimation. We compared the finite-sample properties, size and power, of their overidentification tests through Monte Carlo simulations. As a starting point, we used models in which these properties have already been explored. These are: the Qin and Lawless Model (1994), that proposed by Hall and Horowitz (1996) and a chi-squared moments model. Then, we looked for alternative frameworks to assess

[21] Although in Chapter 2 we also analysed other approaches such as KLIC and parametric likelihood estimation.

the finite-sample properties of overidentification tests. We concentrated on those models with potential empirical applications. In particular, we focused on the Mean-Variance and Three-Moment CAPM and a dynamic panel data model with individual effects.[22]

Finally, to complement and complete our research, we carried out an empirical application on an autoregressive cash-flow series with individual effects for 174 firms in the United States from 1981 to 1985.

This thesis extends the work of Bravo (2000), Imbens et al (1998), Hall and Horowitz (1996) and Qin and Lawless (1994) in a number of ways:

1. We exploited an alternative framework to that studied in Qin and Lawless (1994) to document the empirical coverage and average length of confidence intervals. Qin and Lawless (1994) explored parametric and non-parametric methods to construct confidence regions within a model characterized by the first and second moments of a random variable. They found that the parametric likelihood confidence intervals have empirical coverages closer to their nominal counterparts (assuming that the likelihood is correctly specified) than those corresponding to EL. Our experiments differ from Qin and Lawless' (1994) by: (*i*) using a different framework, a chi-squared moments model, to assess the empirical coverage and average length of confidence intervals and, (*ii*) examining other sample sizes (they report results for $n = 30$ and $n = 60$, whereas our samples sizes are $n = 50$ and $n = 100$). Our Monte Carlo experiments support Qin

[22]Despite the fact that both models have already been used to assess the finite-sample properties of overidenitifcation tests before, existing evidence is for the GMM and not for the EL. Our literature search suggests that the Three-Moment CAPM has never been used to assess these properties.

and Lawless' (1994) simulation evidence and illustrate the fact that for a relatively small sample size, n=50, the empirical coverages of tests based on EL are fairly close to those based on the parametric likelihood.

2. We examined the finite-sample size properties of overidentification tests within the Qin and Lawless (1994) and the Hall and Horowitz (1996) models using unexplored distributions and parameter values. Prior research by Bravo (2000) based on the Qin and Lawless (1994) model illustrates that the ELR J-test has small size distortions for normally distributed variables. This thesis extended Bravo's (2000) research by investigating the extent in which the latter result holds for variables distributed as chi-square, t and gamma. Our simulations show that it is only in the case of the normal distribution that the empirical sizes of the tests are close to their nominal counterparts. Other distributions led to poor sizes. This is especially true for tests based on $t(5)$ and $\chi^2_{(1)}$ processes. However, even if the large-sample approximations are not reliable, the ELR overidentification test performed better than statistics based on the GMM (two-step and continuously updated) and the KLIC. Prior research using the Hall and Horowitz (1996) model by Bravo (2000) and Imbens *et al* (1998) illustrates the poor size properties of the ELR overidentification test based on normally distributed variables. Our findings show that the latter results also extend to other DGPs. We found important size distortions for $\chi^2_{(2)}$ and $\Gamma(1,1)$ processes, although to a lesser extent than for the normal specification.

Some of the contributions and findings of this thesis are:

1. Although we investigated standard computational aspects of EL, prior to this research little had been documented about them. We illustrated that despite the fact that sequential algorithms are more computational intensive than simultaneous algorithms, both led to identical estimates. Our experiments suggest that the ELR test is sensitive to starting values –in terms of accuracy and computation time– for variables distributed as $\chi^2_{(1)}$. However, there is no evidence of accuracy being related to poor starting values for $N(0,1)$ variables.

2. We assessed the size and power properties of overidentification tests using the Mean-Variance and the Three-Moment CAPM. Prior research (Vorkink, 2003; Dahlquist and Soderlind, 1997; Hansen et al, 1996) entirely focused on the Two-Moment version and mainly on GMM tests. As far as we are aware, this is the first study that has assessed the finite-sample properties of EL statistics within asset pricing models (both for the Mean-Variance and the Three-Moment CAPM). We found that sizes are very similar for GMM and ELR overidentification statistics. The asymptotic approximation for these tests was good in most of our experiments within both versions of the CAPM. However, our findings show that EL estimation has better power.

3. Prior research (Kitamura, 2001) on the power properties of the ELR overidentification test considered the Hall and Horowitz (1996) model. Within this setting, the ELR J-test is more powerful than its GMM counterparts. The latter is especially true for

alternatives farther from the null. Our experiments to assess power, based on the CAPM, support Kitamura's (2001) findings in that the ELR moment restrictions test has better power than the GMM tests. However, we found that differences in power were larger not only when power was already high, like in Kitamura (2001), but more uniformly.

4. Earlier research assessed the power properties of overidentification tests under the alternative of misspecification (Kitamura, 2001; Bowsher, 2000a). DGPs incorporate deviations from the null hypothesis either by violating one of the main assumptions of the model, within dynamic panel data frameworks, or by varying the parameters through the null. The contribution of our Three-Moment CAPM experiments is that we assessed power under the alternative that the Mean-Variance CAPM is valid. We simulated data consistent with the Mean-Variance CAPM, while considering a set of estimating equations consistent with the Three-Moment CAPM. Our simulations show that the ELR J-test has better power than GMM tests.

5. Brown and Newey (2001) developed a novel method of bootstrapping based on the MEL weights, which incorporate the information contained in the moment equations, rather than on uniform weights (efficient bootstrap). Brown and Newey (2001) applied this technique to the GMM (GMM-bootstrap) and found important improvements in the coverage properties of GMM estimators and a slight improvement in the finite-sample properties of its overidentification test. Our research introduced the use of efficient bootstrap critical values for EL (EL-bootstrap) in three different frameworks:

Qin and Lawless (1994) model, that proposed by Hall and Horowitz (1996) and a dynamic panel data model with individual effects. Our experiments show that within the three settings and for our distributional assumptions and parameter values, the EL-bootstrap had good size properties. Moreover, under specifications in which we had previously found very poor sizes, the EL-bootstrap substantially reduced the size distortions and in some cases these were virtually eliminated. The EL-bootstrap led to better size-properties than Brown and Newey's GMM-bootstrap (within the models and under the assumptions that we examined).

6. We assessed the finite-sample properties of the EL-bootstrap overidentification test within dynamic panel data models. Up to this point, research on dynamic panel data models has been dominated by the GMM approach. It has been extensively documented that GMM estimators can be biased and imprecise due to weak instruments. However, the behaviour of its overidentification test in the presence of weak instruments has not been thoroughly treated in the literature. Our findings show that the Sargan test is less sensitive to weak instruments than GMM estimators. It is common-practice to introduce further moment conditions into the analysis to reduce sample bias and the imprecision of GMM estimators. However, we found a pattern that suggests that incorporating moment conditions could lead to a deterioration in the size properties of the Sargan test. Our simulation experiments show that the EL-bootstrap overidentification test is not sensitive to weak instruments (except for T=4 and $\rho = .9$). Moreover, there is not enough evidence to support the conclusion that

EL-bootstrap tests based on SYS conditions dominate those based on DIF conditions.

7. The cash-flow empirical example illustrates how EL can be used in applied work. Our Monte Carlo experiments, for the same sample size and number of time periods as our empirical application, provided evidence to suspect that the Sargan test was over-rejecting the null. According to our simulations, the test based on EL-bootstrap was more accurate for this specification. Thus we relied on the latter statistic. Our conclusion points to cash-flow being well represented by an AR(1) panel data model with individual effects.

In summary, although our experiments do not uniformly support the conclusion that one estimator dominates the other, we found some evidence that EL and EL-bootstrap could be good alternatives to GMM in some econometric applications.

Appendices

Appendix 1

Since the initial investment is set to one, the moments of end of period wealth are equivalent to those of the rate of return on the portfolio, *i.e.*:

$$\begin{aligned}\sigma(\omega) &= \sigma(R_p), \\ \gamma(\omega) &= \gamma(R_p), \\ \theta(\omega) &= \theta(R_p).\end{aligned}$$

Using $\sum_i x_i R_i = R_p - x_0 R_f$ and $\sum_i x_i E(R_i) = E(R_p) - x_0 R_f$ gives

$$\begin{aligned}\sum_i x_i \beta_{ip} &= \sum_i x_i \frac{E\left[\{R_i - E(R_i)\}\{R_p - E(R_p)\}\right]}{E\left[\{R_p - E(R_p)\}^2\right]} \\ &= \frac{E\left[\left\{\sum_i x_i R_i - \sum_i x_i E(R_i)\right\}\{R_p - E(R_p)\}\right]}{E\left[\{R_p - E(R_p)\}^2\right]} \\ &= 1.\end{aligned}$$

Therefore,

$$\sigma(\omega) = \sum_i x_i \beta_{ip} \sigma(R_p).$$

Following the same procedure we obtain:

$$\sum_i x_i \gamma_{ip} = 1,$$
$$\sum_i x_i \theta_{ip} = 1,$$

which leads to

$$\gamma(\omega) = \sum_i x_i \gamma_{ip} \gamma(R_p) \ and \ \theta(\omega) = \sum_i x_i \theta_{ip} \theta(R_p).$$

Appendix 2

We define β_m as

$$\beta_m = \frac{Cov(r_t, r_{mt})}{Var(r_{mt})}.$$

Substituting the proposed DGP into β_m yields

$$\begin{aligned}
\beta_m &= \frac{Cov\left(a_1 r_{mt} + a_2 (r_{mt} - E(r_{mt}))^2 + \varepsilon_t \ , \ r_{mt}\right)}{Var(r_{mt})} \\
&= \frac{a_1 Cov(r_{mt}, r_{mt})}{Var(r_{mt})} + \frac{a_2\left[Cov\left(r_{mt}^2, r_{mt}\right) - 2E(r_{mt}) Cov(r_{mt}, r_{mt})\right]}{Var(r_{mt})} \\
&= a_1 + a_2 \frac{\left[E\left(r_{mt}^3\right) - 3E(r_{mt}) E\left(r_{mt}^2\right) + 2\left(E(r_{mt})\right)^3\right]}{Var(r_{mt})} \\
&= a_1 + a_2 \frac{\gamma(r_m)^3}{\sigma(r_m)^2}.
\end{aligned}$$

We define γ_m as

$$\gamma_m = \frac{Cov(r_t, r_{mt}) - 2E(r_{mt}) Cov(r_t, r_{mt})}{\gamma(r_{mt})^3}.$$

Substituting the proposed DGP into γ_m yields

$$\gamma_m = \frac{Cov\left(a_1 r_{mt} + a_2 \left(r_{mt} - E(r_{mt})^2 + \varepsilon_t \ , \ r_{mt}\right)\right.}{\gamma\left(r_{mt}\right)^3}$$
$$- \frac{2E\left(r_{mt}\right)\left[a_1\left(E\left(r_{mt}\right) - E\left(r_{mt}\right)\right) + a_2\right]\gamma\left(r_{mt}\right)^3}{\gamma\left(r_{mt}\right)^3}$$
$$= a_1 + a_2 \frac{E\left(r_{mt}^4\right) - \left(E\left(r_{mt}\right)\right)^2 - 2E\left(r_{mt}\right)\left[E\left(r_{mt}^3\right) - \left(E\left(r_{mt}\right)\right)E\left(r_{mt}^2\right)\right]}{\gamma\left(r_m\right)^3}$$
$$- a_2 \frac{2E\left(r_{mt}\right)E\left(r_{mt}^3\right) + 6\left(E\left(r_{mt}\right)\right)^2 E\left(r_{mt}^2\right) - 4\left(E\left(r_{mt}\right)\right)^4}{\gamma\left(r_{mt}\right)^3}.$$

After rearranging:

$$\gamma_m = a_1 + a_2 \frac{E\left(r_{mt}^4\right) - \left(E\left(r_{mt}\right)\right)^2 - 4E\left(r_{mt}\right)E\left(r_{mt}^3\right)}{\gamma\left(r_{mt}\right)^3}$$
$$+ a_2 \frac{8\left(E\left(r_{mt}\right)\right)^2 E\left(r_{mt}^2\right) - 4\left(E\left(r_{mt}\right)\right)^4}{\gamma\left(r_{mt}\right)^3}.$$

To simplify this expression we note that

$$E\left[\left(r_{mt} - E\left(r_{mt}\right)\right)^4\right] - E\left[\left(r_{mt} - E\left(r_{mt}\right)\right)^2\right]^2 = E\left(r_{mt}^4\right) - \left(E\left(r_{mt}\right)\right)^2$$
$$- 4E\left(r_{mt}\right)E\left(r_{mt}^3\right) + 8\left(E\left(r_{mt}\right)\right)^2 E\left(r_{mt}^2\right) - 4\left(E\left(r_{mt}\right)\right)^4.$$

Hence,

$$\gamma_m = a_1 + a_2 \frac{\left[\varphi\left(r_{mt}\right)^4 - \sigma\left(r_{mt}\right)^2\right]}{\gamma\left(r_{mt}\right)^3}.$$

Appendix 3

Empirical Levels of J-tests
Dynamic Panel Data
DIF Moment Conditions

		n=100					
		T=4		T=5		T=6	
Levels	ρ	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}
.10		.1006	.1084	.1044	.1042	.0772	.0984
.05	.2	.0526	.0532	.0576	.0526	.0394	.0448
.01		.0124	.0086	.0108	.0084	.0080	.0086
.10		.1196	.1074	.1040	.1114	.1160	.1066
.05	.5	.0604	.0564	.0530	.0564	.0524	.0472
.01		.0134	.0148	.0124	.0126	.0062	.0076
.10		.1006	.1270	.0884	.1292	.1158	.1186
.05	.7	.0492	.0710	.0392	.0656	.0532	.0594
.01		.0110	.0204	.0048	.0136	.0050	.0084
.10		.1282	.1018	.1142	.1056	.1282	.1092
.05	.9	.0682	.0492	.0615	.0464	.0598	.0512
.01		.0154	.0068	.0071	.0072	.0108	.0070

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM. T is the time periods, n is the sample size and ρ is the autoregressive coefficient.

Table 4.13: Finite-Sample Size Properties - Dynamic Panel Data (DIF conditions n=100)

Empirical Levels of J-tests Dynamic Panel Data DIF Moment Conditions							
		n=175					
		T=4		T=5		T=6	
Levels	ρ	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}
.10		.1162	.0968	.1214	.0936	.0948	.1098
.05	.2	.0596	.0466	.0632	.0506	.0482	.0528
.01		.0148	.0078	.0118	.0094	.0094	.0100
.10		.0978	.1004	.1004	.1140	.0948	.1126
.05	.5	.0450	.0518	.0462	.0562	.0474	.0552
.01		.0068	.0166	.0100	.0130	.0074	.0088
.10		.0946	.1080	.1143	.1326	.0972	.1166
.05	.7	.0506	.0588	.0603	.0688	.0398	.061
.01		.0084	.0152	.0126	.0128	.0226	.0136
.10		.1010	.0990	.1006	.1148	.1162	.1188
.05	.9	.0544	.0524	.0520	.0604	.0668	.0604
.01		.0118	.0096	.0110	.0114	.0134	.0128

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM estimators. T is the time periods, n is the sample size and ρ is the autoregressive coefficient

Table 4.14: Finite-Sample Size Properties - Dynamic Panel Data (DIF conditions n=175)

Empirical Levels of J-tests
Dynamic Panel Data
SYS Moment Conditions

n=100

Levels	ρ	T=4 W_j^b	T=4 J_{2GMM}	T=5 W_j^b	T=5 J_{2GMM}	T=6 W_j^b	T=6 J_{2GMM}
.10		.0894	.1094	.1035	.1118	.1047	.1028
.05	.2	.0552	.0538	.0544	.0590	.0541	.0490
.01		.0188	.0096	.0123	.0092	.0095	.0090
.10		.0994	.1036	.1164	.1148	.1148	.1144
.05	.5	.0450	.0496	.0558	.0572	.0534	.0522
.01		.0052	.0098	.0108	.0106	.0104	.0086
.10		.0992	.1164	.1024	.1196	.0926	.1218
.05	.7	.0480	.0578	.0500	.0592	.0362	.0612
.01		.0074	.0102	.0094	.0106	.0062	.0134
.10		.0994	.1060	.0954	.1208	.1042	.1256
.05	.9	.0474	.0540	.0410	.0578	.0452	.0684
.01		.0086	.0096	.0036	.0100	.0062	.0100

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM estimators. T is the time periods and ρ is the autoregressive coefficient.

Table 4.15: Finite-Sample Size Properties - Dynamic Panel Data (SYS conditions n=100)

Empirical Levels of J-Tests
Dynamic Panel Data
SYS Moment Conditions

		n=175					
		T=4		T=5		T=6	
Level	ρ	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}	W_j^b	J_{2GMM}
.10		.1031	.0938	.1250	.1076	.1235	.1058
.05	.2	.0510	.0458	.0658	.0500	.0704	.0564
.01		.0119	.0100	.0138	.0088	.0190	.0112
.10		.1204	.1034	.0993	.1044	.1159	.1118
.05	.5	.0652	.0540	.0487	.0548	.0569	.0552
.01		.0124	.0092	.0096	.0120	.0100	.012
.10		.1106	.1094	.0910	.1072	.0728	.1312
.05	.7	.0608	.0614	.0396	.0556	.0334	.0694
.01		.0114	.0146	.0038	.0114	.0034	.0166
.10		.1050	.1220	.0924	.1198	.1002	.1434
.05	.9	.0470	.0636	.0420	.0638	.0478	.0746
.01		.0108	.0136	.0064	.0146	.0078	.0176

W_j^b and J_{2GMM} are overidentification tests based on EL-bootstrap and two-step GMM estimators. T is the time periods, n is the sample size and ρ is the autoregressive coefficient.

Table 4.16: Finite-Sample Size Properties - Dynamic Panel Data (SYS conditions n=175)

Appendix 4

Dynamic Panel Data AR(2) Cash-flow GMM estimation (DIF conditions)				
	Coeff.	s.e.	t-value	p-value
cf_{t-1}	.4892	.1798	2.72	.007
cf_{t-2}	.0426	.0806	.529	.597
Constant	.0749	.0306	2.44	.015
T1985	-.2176	.0454	-4.78	.000
Wald (joint) (2 df)	7.910			.019
Wald (dummy) (2df)	23.18			.000
Wald (time) (2 df)	23.18			.000
J_{2GMM} (3 df)	6.473			.091
m_1	-2.075			.038

p-values are reported inside parenthesis. df refers to the degrees of freedom. m_1 is a test for serial correlation. J_{2GMM} is the Sargan test. Note: m_2 could not be calculated because there are not enough observations (for an AR(2) process we need $T \geq 6$).

Table 4.17: AR(2) Cash-flow

Bibliography

[1] Ahn, S.C. and P. Schmidt (1995), "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics,* Vol. 68, pp. 5-27.

[2] Aitchison, J. and S.D. Silvey (1958), "Maximum Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, Vol. 29, pp. 813-828.

[3] Alderfer, C. and H. Bierman (1970), "Choice with Risk: Beyond the Mean and Variance," *Journal of Business*, Vol. 43, pp. 341-353.

[4] Alonso-Borrego, C. and M. Arellano (1999), "Symmetrically Normalized Intsrumental-Variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, Vol. 17, pp. 36-49.

[5] Alvarez, J. and M. Arellano (2004), "Robust Likelihood Estimation of Dynamic Panel Data Models," CEMFI Working Papers 0421/04, Centro de Estudios Monetarios y Financieros.

[6] Anderson, T.W. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics,* Vol. 19, pp. 578-606.

[7] Anderson, T.W. and C. Hsiao (1981), "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association,* Vol. 76, pp. 598-606.

[8] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics,* Vol. 30, pp. 46-63.

[9] Arditti, Fred (1967), "Risk and the Required Return on Equity," *The Journal of Finance*, Vol. 22, pp. 19-36.

[10] Arellano, Manuel (2003), *Panel Data Econometrics*, Oxford: Oxford University Press.

[11] Arellano, M. and S. Bond (1998), "Dynamic Panel Data Estimation Using DPD98 for GAUSS: A Guide for Users," Institute of Fiscal Studies, London: Photocopied.

[12] Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, Vol. 58, No. 2, pp. 277-297.

[13] Baggerly, K.A. (1998), "Empirical Likelihood as a Goodness of Fit Measure," *Biometrika*, Vol. 85, pp. 535-547.

[14] Balestra, P. and M. Nerlove (1966), "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model," *Econometrica*, Vol. 51, pp. 1635-1659.

[15] Banz, R. (1981), "The Relation between Return and Market Value of Common Stocks," *Journal of Financial Economics*, Vol. 9, pp. 3-18.

[16] Basu, S. (1977), "The Investment Performance of Common Stock in Relation to their Price to Earning Ratios: A Test of the Efficient Market Hypothesis", *Journal of Finance*, Vol. 32, pp. 663-682.

[17] Beran, R. and Ducharme, G.R. (1991), "Asymptotic Theory for Bootstrap Methods in Statistics," Les Publications Centre de Recherches Mathematiques: Universitë dë Montrëal, Montrëal.

[18] Black, F., M. Jensen and M. Scholes (1972), "The Capital Asset Pricing Model: Some Empirical Tests," in Jensen, M. (ed.), *Studies in Theory of Capital Markets*, New York: Praeger.

[19] Blattberg, R. and N. Gonedes (1974), "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices," *Journal of Business*, Vol. 47, pp. 244-280.

[20] Blume, M. and I. Friend (1973), "A New Look at the Capital Asset Pricing Model," *Journal of Finance*, Vol. 28, pp. 19-33.

[21] Blundell, R. and S. Bond. (1999), "GMM Estimation With Persistent Panel Data: An Application to Production Functions," IFS Working Papers W99/4, The Institute for Fiscal Studies.

[22] Blundell, R. and S. Bond. (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 87, pp. 115-143.

[23] Blundell, R., S. Bond and, F. Windmeijer (2000), "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimators," IFS Working Papers W00/12, The Institute for Fiscal Studies.

[24] Bond, S. (2002), "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice," CEMMAP working papers CWP09/02, Centre for Microdata Methods and Practice, The Institute for Fiscal Studies.

[25] Bond, S., C. Nauges and, F. Windmeijer (2002), "Unit Roots and Identification in Autoregressive Panel Data Models: A Comparison of Alternative Tests," Institute of Fiscal Studies, London: Photocopied.

[26] Bond, S., C. Bowsher and F. Windmeijer (2001), "Criterion-Based Inference for GMM in Autoregressive Panel Data Models," *Economics Letters*, Elsevier, Vol. 73, pp. 379-388.

[27] Bond, S. and F. Windmeijer (2002), "Finite Sample Inference for GMM Estimators in Linear Panel Data Models," CEMMAP working papers CWP04/02, Centre for Microdata Methods and Practice, The Institute for Fiscal Studies.

[28] Bottazzi, J.M., H. Thorsten and, L. András (1995), "Market Demand Function in the CAPM," Discussion Paper in Economics 347/95, The University of Bielefeld.

[29] Bowsher, C. (2000), "On Testing Overidentifying Restrictions in Dynamic Panel Data Models," Nuffield College Discussion Paper in Economics 2000-W28, University of Oxford.

[30] Bowsher, C. (2000a), "Finite Sample Properties of Alternative Methods of Inference in Dynamic Panel Data Models", Unpublished PhD dissertation, University College, University of Oxford.

[31] Bravo, F. (2001), "Empirical Cressie-Read Power Divergence Statistic for Stationary Time Series, Estimation and Testing", Unpublished manuscript, University of York.

[32] Bravo, F. (2000), "Higher Order Asymptotics and the Bootstrap for Empirical Likelihood J Tests", Discussion Papers in Economics 30/00, University of York.

[33] Brown, B. and W. Newey (2001), "GMM, Efficient Bootstrapping, and Improved Inference," *Journal of Business and Economic Statistics*, Vol. 20, pp. 507-517.

[34] Campbell, J., A. Lo and, A. MacKinlay (1997), "The Econometrics of Financial Markets", Princeton, New Jersey: Princeton University Press.

[35] Cass, D. and J. Stiglitz (1970), "The Structure of Investor Preferences, Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds," *Journal of Economic Theory*, Vol. 2, pp. 122-60.

[36] Chen, S. X. (1994), "Comparing Empirical Likelihood and Bootstrap Hypothesis Tests," *Journal of Multivariate Analysis*, Vol. 51, pp. 277-293.

[37] Chung, K.L. (1974), "A Course in Probability Theory," U.S.A.: Academic Press, Inc.

[38] Cochrane, J. (2001), "Asset Pricing," Princeton: Princeton University Press.

[39] Copeland, T. and J.F. Weston (1998), "Financial Theory and Corporate Policy," U.S.A.: Addison-Wesley Publishing Company.

[40] Cressie, N. and T. Read (1984), "Multinomial Goodness of Fit Tests," *Journal of the Royal Statistical Society*, Series B, Vol. 46, pp. 440-464.

[41] Dacorogna, M., Muler, U., Pictet, O and G. de Vries (1995), "The Distribution of Extremal Foreign Exchange Returns in Extremely Large Data Sets", *Mimeo*, Olsen & Associates, Research Group.

[42] Dahlquist, M. and P. Soderlind (1997), "Evaluating Portfolio Performance with Stochastic Discount Factors," C.E.P.R. Discussion Papers 1663, The Centre for Economic Policy Research.

[43] Davidson, J. (1994), "Stochastic Limit Theory," Oxford: Oxford University Press.

[44] Davidson, A.C., D. Hinkley and, B.J. Worton (1992), "Bootstrap likelihoods," *Biometrika*, Vol. 79, pp. 113-130.

[45] Davidson, R. and J.G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypothesis," *Econometrica*, Vol 49, pp. 781-792.

[46] DeBondt, W. and R. Thaler (1985), "Does the Stock Market Overreact?," *Journal of Finance*, Vol. 40, pp. 793-805.

[47] DiCiccio, T., P. Hall and, J. Romano (1988), "Bartlett Adjustment for Empirical Likelihood," Technical Report No. 298, Dept. Statistics, Stanford University.

[48] Doornik, J. and D. Hendry (2001), "Econometric Modelling Using PcGive," Vol. III, Timberlake Consultants Ltd., United Kingdom.

[49] Efron, B. (1979), "Bootstrap Methods: Another Look at the Jacknife," *Annals of Statistics*, Vol. 7, pp. 1-26.

[50] Efron, B. and R.J. Tibshirani (1993), "An Introduction to the Bootstrap," Chapman and Hall, New York.

[51] Fama, E. (1970), "Multiperiod Consumption-Investment Decisions," *The American Economic Review*, Vol. 60, pp. 163-174.

[52] Fama, E. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, Vol. 38, pp. 34-105.

[53] Fama, E. and K. French (1992), "The Cross-Section of Expected Stock Returns," *Journal of Finance*, Vol. 47, pp. 427-465.

[54] Fama, E. and J. MacBeth (1973), "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, Vol. 71, pp. 607-636.

[55] Ferson, W. and S. Foerster (1994), "Finite sample properties of the generalized method of moments in tests of conditional asset pricing models," *Journal of Financial Economics*, Vol. 36, pp. 29-55.

[56] Friend, I. and R. Westerfield (1980), "Co-skewness and Capital Asset Pricing," *Journal of Finance*, Vol. 35, pp. 897-913.

[57] Greene, William (2000), "Econometric Analysis," U.S.A.: Prentice Hall, Inc.

[58] Hall, B. and J. Mairesse (2002), "Testing for Unit Roots in Panel Data: An Exploration Using Real and Simulated Data," In: Andrews and Stock (eds.) *Identification and Inference for Econometric Models: a Festschrift in Honor of Thomas Rothenberg*, Cambridge University Press.

[59] Hall, B., J. Mairesse, L. Branstetter and, B. Crepon (1998), "Does Cash Flow Cause Investment and R&D: An Exploration Using Panel Data from French, Japanese, and United States Scientific Firms," In: Audretsh, D. and A.R. Thurik (eds.) *Innovation, Industry Evolution and Employment*, Cambridge, England: Cambridge University Press.

[60] Hall, P. (1992), "The Bootstrap and Edgeworth Expansion," Springer-Verlag, New York.

[61] Hall, P. and B. Presnell. (1999) "Intentionally biased bootstrap methods," *Journal of the Royal Statistical Society*, Series B, Methodological, Vol. 61, pp. 143-158.

[62] Hall, P. and J. Horowitz (1996), "Bootstrap Critical Values for Tests Based on Generalized Method of Moment Estimators," *Econometrica*, Vol. 64, pp. 891-916.

[63] Hamilton, J.D. (1991), "A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions", *Journal of Business and Economic Statistics*, Vol. 9, pp. 27-39.

[64] Hansen, Bruce (2003), "Econometrics," Unpublished Manuscript, University of Wisconsin.

[65] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, pp. 1029-1054.

[66] Hansen, L.P., J. Heaton and, A. Yaron (1996), "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, Vol. 14, pp. 262-280.

[67] Harvey, C. and G. Zhou (1993), "International Asset Pricing with Alternative Distributional Specifications", *Journal of Empirical Finance*, Vol. 1, pp. 107-131.

[68] Hoeffding, W. (1965), "Asymptotically Optimal Tests for Multinomial Distributions," *The Annals of Mathematical Statistics*, Vol. 36, pp. 369-400.

[69] Holtz-Eakin, D., W. Newey, and H. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica*, Vol. 56, pp. 1371-1395.

[70] Homaifar, G. and D. Graddy (1990), "Variance and Lower Partial Moment Betas as Alternative Risk Measures in Cost of Capital Estimation: A Defence of the CAPM Beta," *Journal of Finance and Accounting*, Vol. 17, pp. 677-688.

[71] Horowitz, J.L. (1997), "Bootstrap Methods in Econometrics: Theory and Numerical Performance," In: D.M. Kreps and K.F. Wallis (eds.) *Advances in Economics and Econometrics: Theory and Applications*, Cambridge: Cambridge University Press.

[72] Horowitz, J. (2000), "The Bootstrap," Unpublished Manuscript, Department of Economics, University of Iowa.

[73] Horowitz, J. and N.E. Savin (2000), "Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach," *Journal of Econometrics*, Vol. 95, pp. 375-389.

[74] Hsiao, C. (1986), "Analysis of Panel Data,"Cambridge: Cambridge University Press.

[75] Hwang, S. and S. Satchell (1999), "Modelling Emerging Market Risk Premia using Higher moments", *International Journal of Finance and Economics*, Vol. 4, pp. 271-296.

[76] Imbens, G. (1997), "One Step Estimators for Over-Identified Generalized Method of Moment Models," *Review of Economic Studies*, Vol. 64, pp. 359-383.

[77] Imbens, G., P. Johnson and, R. Spady (1998), "Information Theoretic Approaches to inference in moment condition models," *Econometrica*, Vol. 66, pp. 333-357.

[78] Ingersoll, J. (1987), "Theory of Financial Decision Making," Totowa, N.J.: Rowman and Littlefield.

[79] Ingersoll, J. (1975), "Multidimensional Security Pricing," *Journal of Financial and Quantitative Analysis*, Vol. 10, pp. 785-798.

[80] Jean, W.H. (1971), "The Extension of Portfolio Analysis to Three or More Parameters," *The Journal of Financial and Quantitative Analysis*, Vol 6, pp. 505-515.

[81] Jean, W.H. (1973), "More on Multidimensional Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, Vol. 8, pp. 475-490.

[82] Jegadeesh, N. and S. Titman (1995), "Overreaction, Delayed Reaction and Contrarian Profits," *Review of Financial Studies*, Vol. 8, pp. 973-993.

[83] Kocherlakota, N. (1990), "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, Vol. 26, pp. 285-304.

[84] Kraus, A. and R.H. Litzenberger (1976), "Skewness Preference and the Valuation of Risk Assets," *Journal of Finance*, Vol. 31, pp. 1085-1100.

[85] Kitamura, Y. (2001), "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, Vol. 69, pp. 1661-1672.

[86] Kitamura, Y. (1997), "Empirical Likelihood Methods with Weakly Dependent Processes," *The Annals of Statistics*, Vol. 25, pp. 2084-2102.

[87] Kitamura, Y. and M. Stutzer (1997), "An Information-Theoretic Alternative to Generalize Method of Moments Estimation," *Econometrica*, Vol. 65, pp 861-874.

[88] Lai, T.L. and J.Q.Z. Wang (1993), "Edgeworth expansions for symmetric statistics with applications to bootstrap methods," *Statistica Sinica*, Vol. 3, pp. 517-542.

[89] Lazar, N.A. and P.A. Mykland (1998), "An Evaluation of the Power and Conditionality Properties of Empirical Likelihood," *Biometrika*, Vol. 85, pp. 523-534.

[90] Levy, H. (1969), "A Utility Function Depending on the First Three Moments," *The Journal of Finance*, Vol. 24, pp. 715-719.

[91] Lim, Kian-Guan (1989), "A New Test of the Three-Moment Capital Asset Pricing Model," *Journal of Financial and Quantitative Analysis*, Vol. 24, 205-216.

[92] Lintner, J. (1965), "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, Vol. 47, pp. 13-37.

[93] Magdalinos, M. and S. Symeonides (1996), "A Reinterpretation of the Tests of Overidentifying Restrictions," *Journal of Econometrics*, Vol. 73, pp. 325-353.

[94] Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, Vol. 36, pp. 394-419.

[95] Mammen, E. (1992), "When does Bootstrap work? Asymptotic Results and Simulations," Heidelberg: Springer-Verlag.

[96] Markowitz, N. (1959), "Portfolio Selection: Efficient Diversification of Investments," New York: John Wiley.

[97] Mátyás, L. and P. Sevestre (1996), "The Econometrics of Panel Data: A Handbook of the Theory with Applications," Netherlands: Kluwer Academic Publishers.

[98] McCullagh, P. and J.A. Nelder (1989), "Generalized Linear Models," London: Chapman and Hall.

[99] McDonald, J.B. and Y. Xu (1995), "A Generalization of the Beta Distribution with Applications," *Journal of Econometrics*, Vol. 66, pp. 133-152.

[100] Mittelhammer, R. C. (1996), "Mathematical Statistics for Economics and Business", New York: Springer-Verlag.

[101] Mittelhammer, R., G. Judge and, R. Schoenberg (2003), "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods", CUDARE Working Paper 945, Department of Agricultural & Resource Economics, UCB.

[102] Mittelhammer, R., G. Judge and, D. Miller (2000), "Econometric Foundations," Cambridge: Cambridge University Press.

[103] Mood, A., F. Graybill and, D. Boes (1974), "Introduction to the Theory of Statistics," McGraw-Hill, International Editions.

[104] Namba, A. (2004), "Simulation Studies on Bootstrap Empirical Likelihood Tests," *Communications in Statistics: Simulation and Computation*, Vol. 33, pp. 99-108.

[105] Neely, C. (1995), "Testing Asset Pricing Models with Euler Equations: it's worse than you think," Working Paper 1995-018A, The Federal Reserve Bank of St. Louis.

[106] Newey, W. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing". In: R. Engle and D. McFadden (eds.) *Handbook of Econometrics*, Vol. IV, North Holland: Amsterdam, pp. 2111-2245.

[107] Nickell, S.J. (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica*, Vol. 49, pp. 1417-1426.

[108] Nielsen, L. (1993), "Robustness of the Market Model," *Economic Theory*, Vol (3), pp. 365-369.

[109] Owen, Art (2001), "Empirical Likelihood," Monographs on Statistics and Applied Probability 92, Chapman and Hall/CRC.

[110] Owen, Art (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, Vol. 75, pp. 237-49.

[111] Owen, Art (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, Vol. 18, pp. 90-120.

[112] Owen, Art (1991), "Empirical Likelihood for Linear Models," *The Annals of Statistics*, Vol. 19, pp. 1725-1747.

[113] Qin, J. and J.F. Lawless (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, Vol. 22, pp. 300-325.

[114] Qin, J. and J.F. Lawless (1992), "Empirical Likelihood Estimation and Tests for Parameters Subject to Constraints," Technical Report 92-13, Dept. Statistics, University of Waterloo.

[115] Rubinstein, M. (1973), "A Comparative Statics Analysis of Risk Premiums," *Journal of Business*, Vol. 46, pp. 605-615.

[116] Sargan, J.D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, Vol. 26, pp. 393-415.

[117] Sears, R.S. and K.C.J. Wei (1988), "The structure of Skewness Preferences in Asset Pricing Models with Higher Moments: An Empirical Test," *Financial Review*, Vol. 23, pp. 25-38.

[118] Sears, R.S. and K.C.J. Wei (1985), "Asset Pricing and Higher Moments, and the Market Risk Premium: A Note," *The Journal of Finance*, Vol. 40, pp. 1251-1253.

[119] Serfling, Robert (1980), "Approximation Theorems of Mathematical Statistics," Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

[120] Shao, J. and D. Tung (1995), "The Jacknife and Bootstrap," Springer Series in Statistics, Springer.

[121] Sharpe, W. (1964), "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *The Journal of Finance*, Vol. 19, pp. 425-442.

[122] Stein, C. (1956), "Efficient nonparametric testing and estimation," in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 187-195.

[123] Tauchen, G. (1986), "Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business & Economic Statistics*, American Statistical Association, Vol. 4, pp. 397-416.

[124] Thomas, D.R. and Grunkemeier, G. L. (1975), "Confidence Interval Estimation of Survival Probabilities for Censored Data," *Journal of the American Statistical Association*, Vol. 70, pp. 865-871.

[125] Tukey, John (1958), "Bias and confidence in not quite large samples," *Annals of Mathematical Statistics,* Vol. 29, p. 614-625.

[126] Vorkink, K. (2003), "Return Distribution and Improved Tests of Asset Pricing Models," *The Review of Financial Studies*, Vol. 16, pp. 845-874.

[127] Wilks, S. (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypothesis," *Annals of Mathematical Statistics*, Vol. 9, pp. 60-62.

[128] Windmeijer, F. (1998), "Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models," IFS Working Papers W98/1, The Institute for Fiscal Studies.