Improved Inference for Efficient Method of Moments and Indirect Inference Estimators

Veronika Czellar* — HEC Paris
Eric Zivot† — University of Washington

March 29, 2008
Preliminary and Incomplete. Please do not quote without authors' permission.

Abstract

The efficient method of moments (EMM) and indirect inference (II) are two widely used simulation-based techniques for estimating structural models that have intractable likelihood functions. The poor performance in finite samples of traditional coefficient and overidentification tests based on the EMM or II objective function indicates a failure of first order asymptotic theory for the distribution of these tests, especially for EMM.We propose practically feasible saddlepoint coefficient tests for hypotheses on structural coefficients estimated by II and EMM that are asymptotically chi-square distributed and have much better finite sample performance than traditional tests. To construct the tests, we make use of the fact that II and EMM estimators have asymptotically equivalent M-estimators and then use the coefficient saddlepoint tests for M-estimators developed by Robinson, Ronchetti and Young (2003). We evaluate the finite sample behavior of our coefficient saddlepoint tests by Monte Carlo methods using a first order MA model. Whereas traditional likelihood-ratio type tests can exhibit substantial size distortions, we show that our saddlepoint tests do not. We also find that the size-adjusted power of our saddlepoint tests is similar to and sometimes greater than the power of traditional tests.

1 Introduction

The efficient method of moments (EMM) and indirect inference (II) are two widely used simulation-based techniques for estimating structural models that have intractable likelihood functions. Typical examples include discrete-time stochastic volatility models, continuous-time diffusion models, multinomial choice models, and dynamic stochastic general equilibrium models.

*Department of Decision Science and Information Systems, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France. email: czellarv@hec.fr. Much of this work was completed while V. Czellar was visiting the Departments of Economics and Statistics at the University of Washington, whose hospitality is gratefully acknowledged. She would also like to thank the Swiss National Science Foundation for financial support during this visit.

†Department of Economics, Box 353330, Seattle, WA 98195-3330. email: ezivot@u.washington.edu. Support from the Gary Waterman Distinguished Scholar Fund is gratefully acknowledged.

Gouriéroux and Monfort (1996) showed that EMM and II are asymptotically equivalent under certain general conditions. Gallant and Tauchen (2002) argued that EMM is computationally more attractive than II, and is better suited for estimating models with multiple latent variables. However, several studies have shown that II tends to perform better than EMM in finite samples which suggests that the extra computational burden of II may be worthwhile. For example, Chumacero (2001), Michaleades and Ng (2001), and Ghysels *et. al.* (2003) compared EMM and II for the estimation of a simple first order moving average (MA) model. They found that II estimators had less bias and coefficient and overidentification tests had less size distortion, especially when the MA parameter was near the boundary of the invertibility region of the parameter space. For the simple AR(1) model, Duffee and Stanton (2007) found that inference based on EMM was substantially worse than that based on II especially for highly persistent processes that are calibrated to match typical short-term interest rate behavior.

The poor performance in finite samples of traditional coefficient and overidentification tests based on the EMM or II objective function indicates a failure of first order asymptotic theory for the distribution of these tests, especially for EMM. Possible remedies to improve finite sample performance include bootstrapping or the use of higher order asymptotic expansions. However, given the computational complexity and expense of EMM and II bootstrapping has limited practical appeal and we therefore concentrate on the use of higher order asymptotic expansions in the form of saddlepoint approximations.

We propose practically feasible saddlepoint coefficient tests for hypotheses on structural coefficients estimated by EMM and II that are asymptotically chi-square distributed and have much better finite sample performance than traditional tests. To construct the tests, we make use of the fact that EMM and II estimators have asymptotically equivalent M-estimators and then use the coefficient saddlepoint tests for M-estimators developed by Robinson, Ronchetti and Young (2003). We derive the conditions under which the saddlepoint tests for EMM and II are asymptotically equivalent, and show that the tests for EMM are substantially easier to compute and are more numerically stable than the corresponding tests for II.

We evaluate the finite sample behavior of our coefficient saddlepoint tests by Monte Carlo methods using a first order MA model. Whereas traditional likelihood-ratio type tests can exhibit substantial size distortions, we show that our saddlepoint tests do not. We also find that the size-adjusted power of our saddlepoint tests is similar to and sometimes greater than the power of traditional tests.

Our paper is organized as follows. In section 2, we give an overview of estimation and inference with EMM and II, define two types of EMM and three types of II estimators, and discuss some practical issues associated with implementing these estimators. In section 3, we illustrate the finite sample properties of EMM and II using a simple MA(1) model. Our analysis provides the first comprehensive comparison of the different types of EMM and II estimators with respect to estimation and inference. Our analysis shows that traditional likelihood-ratio type coefficient tests can have substantial size distortions in small samples, and that the size distortions for EMM are much worse than the distortions for II. We propose coefficient saddlepoint tests in section 4, where we review influence functions for EMM and II estimators, discuss the relationship between influence functions and M-estimators, present the saddlepoint coefficient tests for M-estimators developed by Robinson, Ronchetti and Young (2003), and show how to implement the saddlepoint tests for EMM and II estimators. In section 5, we evaluate the finite sample performance of our saddlepoint tests in terms of size and size-adjusted power. Our concluding remarks and suggestions for future research are presented in Section 6.

2 Simulation-based Estimation and Inference

We consider two types of simulation-based estimation and inference for a structural model: the efficient method of moments (EMM) introduced by Bansal, Gallant, Hussey, and Tauchen (1993, 1995) and Gallant and Tauchen (1996) and indirect inference (II) introduced by Smith (1993) and Gouriéroux, Monfort and Renault (1993). Both techniques consist of choosing an auxiliary model, easier to estimate than the original one, and the corresponding estimators are obtained by simulation-based procedures.

Assume that a sample of n observations $\{y_t\}_{t=1\ldots,n}$ are generated from a strictly stationary and ergodic probability model F_θ, $\theta \in \mathbb{R}^p$, with density $p(y_{-m}, \ldots, y_{-1}, y_0; \theta)$ that is difficult or impossible to evaluate analytically. Define an auxiliary model $\widetilde{F}_\mu$ in which the parameter $\mu \in \mathbb{R}^r$, with $r \geq p$, is easier to estimate then θ. For example, the auxiliary model can be defined by an approximation of the original likelihood function, by the exact likelihood function of an approximated model or by a set of moment conditions derived from an approximated model. A general purpose seminonparametric auxiliary model that is capable of accurately approximating a large class of stationary structural models was proposed by Gallant and Tauchen (1992), and is described in detail in Zivot and Wang (2005), and Gallant and Tauchen (2006). In this paper, we consider an auxiliary model that is a conditional likelihood of an approximated model.

Denote by $\tilde{\mu}$ the *auxiliary estimator*, or, the estimator of the auxiliary parameter μ calculated with the original sample $\{y_t\}$:

$$\tilde{\mu} = \arg\max_{\mu} \tilde{Q}_n\left(\{y_t\}_{t=1,\ldots,n}, \mu\right). \tag{1}$$

where $\tilde{Q}_n$ denotes a sample objective function associated with the model $\widetilde{F}_\mu$. We consider the case in which the auxiliary estimator is the quasi-maximum likelihood (QML) estimator of the model $\widetilde{F}_\mu$, so that $\tilde{Q}_n$ can be written as

$$\tilde{Q}_n\left(\{y_t\}_{t=1,\ldots,n}, \mu\right) = \frac{1}{n-m} \sum_{t=m+1}^{n} \tilde{f}(y_t; x_{t-1}, \mu), \tag{2}$$

where $\tilde{f}(y_t; x_{t-1}, \mu)$ is the log density of y_t for the model $\widetilde{F}_\mu$ conditioned on $x_{t-1} = \{y_i\}_{i=t-m,\ldots,t-1}$, $m \in \mathbb{N}$.

EMM and II are estimation methodologies that use the auxiliary model information to obtain estimates of the structural parameters θ. The link between the auxiliary model parameters and the structural parameters is given by the so-called binding function $\mu(\theta)$, which is the functional solution of the asymptotic optimization problem

$$\begin{aligned}\mu(\theta) &= \arg\max_{\mu} \lim_{n\to\infty} \tilde{Q}_n\left(\{y_t\}_{t=1,\ldots,n}, \mu\right) \\ &= \arg\max_{\mu} E_{F_\theta}[\tilde{f}(y_0; x_{-1}, \mu)],\end{aligned} \tag{3}$$

where $\tilde{f}(y_0; x_{-1}, \mu)$ denotes the log density of y_0 given $x_{-1} = (y_{-m}, \ldots, y_{-1})$ for the model $\tilde{F}_\mu$, and $E_{F_\theta}[\cdot]$ means that the expectation is taken with respect to F_θ. In order for $\mu(\theta)$ to define a unique mapping it is assumed that $\mu(\theta)$ is one-to-one and that $\frac{\partial\mu(\theta)}{\partial\theta'}$ has full column rank.

If $\mu(\theta)$ is known then non-simulation based versions of EMM and II may be defined. The non-simulation based EMM estimator is then a generalized method of moments (GMM) estimator that makes use of the population moment condition

$$E_{F_\theta}\left[\frac{\partial\tilde{f}(y_0; x_{-1}, \mu(\theta))}{\partial\mu}\right] = 0$$

and is defined as

$$\hat{\theta}^{\text{EMM}} = \arg\min_{\theta}\ J^{\text{EMM}}(\theta) = \arg\min_{\theta}\ \tilde{g}_n(\theta)'\tilde{\Sigma}\tilde{g}_n(\theta) \tag{4}$$

where $\tilde{g}_n(\theta) = \frac{1}{n-m}\sum_{t=m+1}^{n} \frac{\partial\tilde{f}(y_t; x_{t-1}, \mu(\theta))}{\partial\mu}$ is the sample score evaluated at $\mu(\theta)$, and $\tilde{\Sigma}$ is a positive definite (pd) and symmetric weight matrix which may depend on the data through the auxiliary model. The non-simulation based II estimator is then a minimum distance estimator of the form

$$\hat{\theta}^{\text{II}} = \arg\min_{\theta}\ J^{\text{II}}(\theta) = \arg\min_{\theta}\ (\tilde{\mu} - \mu(\theta))'\tilde{\Omega}(\tilde{\mu} - \mu(\theta)) \tag{5}$$

where $\tilde{\Omega}$ is a positive definite (pd) and symmetric weight matrix which may depend on the data through the auxiliary model.

In general, the analytic form of $\mu(\theta)$ is not known. If it is possible to simulate from F_θ,then simulation-based versions of (4) and (5) can be solved to obtain the EMM and II estimators of θ. With simulation-based EMM, $\tilde{\mu}$ is used to estimate $\mu(\theta)$ and simulations are used to approximate the sample score as a function of θ in (4). With simulation-based II, simulations are used to approximate $\mu(\theta)$ in (5). Gouriéroux and Monfort (1996) discuss five types of simulation-based estimators: two types of EMM estimators and three types of II estimators. These estimators vary in how the data is simulated and how the EMM and II objective functions are formed. These estimators are described in the following sub-sections.

2.1 EMM estimators

For the first type of EMM estimator, we draw pseudo-observations from the model F_θ and obtain a long pseudo-data series of size $S \cdot n$:

$$\{y_t(\theta)\}_{t=1,\ldots,Sn}, \quad S \geq 1. \tag{6}$$

Consider the score vector associated with the simulated sample $\{y_t(\theta)\}_{t=1,\ldots,Sn}$ evaluated at $\tilde{\mu}$:

$$\tilde{g}_{Sn}(\theta, \tilde{\mu}) = \frac{\partial \tilde{Q}_{Sn}}{\partial \mu} \left(\{y_t(\theta)\}_{t=1,\ldots,Sn}, \tilde{\mu} \right).$$

The EMM estimator with a long pseudo-data series is defined by

$$\hat{\theta}_S^{\mathrm{EL}}(\tilde{\Sigma}) = \arg\min_\theta \ \mathrm{J}_S^{\mathrm{EL}}(\theta) = \arg\min_\theta \tilde{g}_{Sn}(\theta, \tilde{\mu})' \tilde{\Sigma} \tilde{g}_{Sn}(\theta, \tilde{\mu}), \tag{7}$$

where $\tilde{\Sigma}$ is a positive definite (pd) and symmetric weight matrix, possibly depending on the data, such that $\tilde{\Sigma} \xrightarrow{p} \Sigma$ pd. The EL superscript indicates that the EMM estimator is exploiting a *long series* simulation principle. This type of EMM estimator is utilized by Gallant and Tauchen (2001) and in most empirical applications of EMM in macroeconomics and finance.

For the second type of EMM estimator, we draw S pseudo-data series of size n from the model F_θ:

$$\{y_t^s(\theta)\}_{t=1,\ldots,n}, \quad s = 1, \ldots, S, \quad S \geq 1. \tag{8}$$

Denote by $\tilde{g}_n^s(\theta, \tilde{\mu})$ the score vector associated with the simulated sample $\{y_t^s(\theta)\}_{t=1,\ldots,n}$ evaluated at $\tilde{\mu}$:

$$\tilde{g}_n^s(\theta, \tilde{\mu}) = \frac{\partial \tilde{Q}_n}{\partial \mu} \left(\{y_t^s(\theta)\}_{t=1,\ldots,n}, \tilde{\mu} \right). \tag{9}$$

The EMM estimator using S pseudo-data series of the same length as the observed data is defined by:

$$\hat{\theta}_S^{\mathrm{EA}}(\tilde{\Sigma}) = \arg\min_\theta \ \mathrm{J}_S^{\mathrm{EA}}(\theta) = \arg\min_\theta \ \left(S^{-1} \sum_{s=1}^S \tilde{g}_n^s(\theta, \tilde{\mu}) \right)' \tilde{\Sigma} \left(S^{-1} \sum_{s=1}^S \tilde{g}_n^s(\theta, \tilde{\mu}) \right). \tag{10}$$

The EA superscript indicates that the EMM estimator is exploiting an *aggregate score* simulation principle.

2.2 II estimators

For the first type of II estimator we simulate a long pseudo-data series as in (6) and then compute the auxiliary estimator using the simulated path:

$$\tilde{\mu}_S^{\mathrm{L}}(\theta) = \underset{\mu}{\mathrm{argmax}} \ \tilde{Q}_{Sn} \left(\{y_t(\theta)\}_{t=1,\ldots,Sn}, \mu \right). \tag{11}$$

The II estimator computed with a long simulated series is defined by:

$$\hat{\theta}_S^{\mathrm{IL}}(\tilde{\Omega}) = \arg\min_{\theta}\ \mathrm{J}_S^{\mathrm{IL}}(\theta) = \arg\min_{\theta}\ \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{L}}(\theta)\right)' \tilde{\Omega} \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{L}}(\theta)\right)\ , \tag{12}$$

where $\tilde{\Omega}$ is a pd and symmetric weight matrix, possibly depending on the data, such that $\tilde{\Omega} \xrightarrow{p} \Omega$. The IL superscript indicates that the II estimator is exploiting a *long series* simulation principle. This type of II estimator was originally considered by Smith (1990).

For the second type of II estimator we simulate S pseudo-data series as in (10) and compute the auxiliary estimators for each simulated path:

$$\tilde{\mu}_S^{\mathrm{M}}(\theta) = S^{-1} \sum_{s=1}^{S} \arg\max_{\mu}\ \tilde{Q}_n\left(\{y_t^s(\theta)\}_{t=1,\ldots,n}, \mu\right)\ . \tag{13}$$

An IM estimator is then defined by:

$$\hat{\theta}_S^{\mathrm{IM}}(\tilde{\Omega}) = \arg\min_{\theta}\ \mathrm{J}_S^{\mathrm{IM}}(\theta) = \arg\min_{\theta}\ \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{M}}(\theta)\right)' \tilde{\Omega} \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{M}}(\theta)\right). \tag{14}$$

The IM superscript indicates that the II estimator is exploiting the *mean of auxiliary estimators* principle.

The IM estimator requires S optimizations for the evaluation of the function $\tilde{\mu}_S^{\mathrm{M}}(\theta)$. An alternative and computationally less intensive way to compute the IM estimator involves replacing $\tilde{\mu}_S^{\mathrm{M}}(\theta)$ by:

$$\tilde{\mu}_S^{\mathrm{A}}(\theta) = \arg\max_{\mu}\ S^{-1} \sum_{s=1}^{S} \tilde{Q}_n\left(\{y_t^s(\theta)\}_{t=1,\ldots,n}, \mu\right)\ . \tag{15}$$

We then define the IA estimator by:

$$\hat{\theta}_S^{\mathrm{IA}}(\tilde{\Omega}) = \arg\min_{\theta}\ \mathrm{J}_S^{\mathrm{IA}}(\theta) = \arg\min_{\theta}\ \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{A}}(\theta)\right)' \tilde{\Omega} \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{A}}(\theta)\right). \tag{16}$$

The IA superscript indicates that the II estimator is exploiting an *aggregated auxiliary estimator* principle.

2.3 Computational considerations

The EL and EA estimators are equivalent in terms of computation time and less expensive than the three types of II estimators since the evaluation of the objective function for the EMM estimator does not require any optimizations. This is the main practical advantage of EMM over II. The IM estimator is computationally the most expensive, followed by the IA and IL estimators, which are equivalent in terms of computation time.

2.4 Asymptotic Properties

The asymptotic properties of EMM and II estimators are derived in Gouriéroux *et. al.* (1993), Gouriéroux and Monfort (1996), and Gallant and Tauchen (1996). Under regularity conditions described in Gouriéroux and Monfort (1996), the EMM estimators, with weight matrix $\tilde{\Sigma}$, and the II estimators, with weight matrix $\tilde{\Omega}$, are consistent and asymptotically normally distributed with asymptotic variance matrices given by

$$W_{\text{EMM}} = \left(1+\frac{1}{S}\right)\left(M_\theta'\Sigma M_\theta\right)^{-1} M_\theta'\Sigma\Sigma^{*-1}\Sigma M_\theta\left(M_\theta'\Sigma M_\theta\right)^{-1} \tag{17}$$

$$W_{\text{II}} = \left(1+\frac{1}{S}\right)\left(\frac{\partial\mu(\theta)'}{\partial\theta}\Omega\frac{\partial\mu(\theta)}{\partial\theta'}\right)^{-1}\frac{\partial\mu(\theta)'}{\partial\theta}\Omega\Omega^{*-1}\Omega\frac{\partial\mu(\theta)}{\partial\theta'}\left(\frac{\partial\mu(\theta)'}{\partial\theta}\Omega\frac{\partial\mu(\theta)}{\partial\theta'}\right)^{-1} \tag{18}$$

where[1]

$$\Sigma^* = \mathcal{I}^{-1},\ \Omega^* = M_\mu \mathcal{I}^{-1} M_\mu \tag{19}$$

$$\mathcal{I} = \lim_{n\to\infty} \text{var}\left(\sqrt{n}\tilde{g}_n(\{y_t\}_{t=1,\dots,n}, \mu(\theta)\right), \tag{20}$$

$$M_\mu = E_{F_\theta}\left[\frac{\partial^2 \tilde{f}(y_0; x_{-1}, \mu(\theta))}{\partial\mu\partial\mu'}\right],\ M_\theta = \left\{\frac{\partial}{\partial\theta'}E_{F_\theta}\left[\frac{\partial\tilde{f}(y_0; x_{-1}, \mu)}{\partial\mu}\right]\right\}\Big|_{\mu=\mu(\theta)} \tag{21}$$

The asymptotic equivalence and efficiency of these estimators depend on the choice of $\tilde{\Sigma}$ and $\tilde{\Omega}$. In particular, the EMM and II estimators are asymptotically equivalent and efficient provided $\tilde{\Sigma} \xrightarrow{p} \Sigma^*$ and $\tilde{\Omega} \xrightarrow{p} \Omega^*$. Commonly used estimators are $\tilde{\Sigma}^* = \widetilde{\mathcal{I}}^{-1}$ and $\tilde{\Omega}^* = \tilde{M}_\mu\,\tilde{\Sigma}\,\tilde{M}_\mu$ where

$$\widetilde{\mathcal{I}} = \frac{1}{n-m}\sum_{t=m+1}^{n}\frac{\partial\tilde{f}(y_t; x_{t-1}, \tilde{\mu})}{\partial\mu}\frac{\partial\tilde{f}(y_t; x_{t-1}, \tilde{\mu})}{\partial\mu}'$$

$$\tilde{M}_\mu = \frac{\partial^2\tilde{Q}_n}{\partial\mu\partial\mu'}\left(\{y_t\}_{t=1,\dots,n}, \tilde{\mu}\right).$$

provided $\tilde{F}_\mu$ is a good approximation to F_θ. If $\tilde{F}_\mu$ is not a good approximation to F_θ, then a heteroskedasticity and autocorrelation consistent long-run variance estimate (e.g., Newey and West, 1987) should be used for $\widetilde{\mathcal{I}}$. For notational convenience, we use $\hat{\theta}_S^i$ (i = EL, EA, IL, IM, IA) to denote the efficient EMM and II estimators based on the estimated optimal weight matrices $\tilde{\Sigma}^*$ and $\tilde{\Omega}^*$.

From (17) and (18), the asymptotic variance matrices of the optimal EMM and II estimators become:

$$W^*_{\text{EMM}} = \left(1+\frac{1}{S}\right)\left(M_\theta'\Sigma^* M_\theta\right)^{-1} \tag{22}$$

$$W^*_{\text{II}} = \left(1+\frac{1}{S}\right)\left(\frac{\partial\mu(\theta)'}{\partial\theta}\Omega^*\frac{\partial\mu(\theta)}{\partial\theta'}\right)^{-1} \tag{23}$$

[1] In M_μ and M_θ, the derivatives are taken first, then the expectations and the results are evaluated at $\mu = \mu(\theta)$.

Gouriéroux and Monfort (1996) derived the result

$$\frac{\partial \mu(\theta)}{\partial \theta'} = -M_{\mu}^{-1} M_{\theta} \tag{24}$$

from which it follows that (22) and (23) are equal and the optimal EMM and II estimators are asymptotically equivalent. Estimates of (22) are generally easier to compute and are typically more numerically stable than estimates of (23).

2.5 Classical Coefficient Test Statistics

Consider a composite hypothesis defined by $H_0 : q(\theta) = \eta_0$ for a smooth function q from ^{p} to $\mathbb{R}^{p_1}$. Wald-type tests based on EMM and II estimators can be constructed using the asymptotic variances in (18) and (17). As described in Gouriéroux *et al.* (1993) and Gallant and Tauchen (1996) likelihood ratio-type (LR-type) test statistics can be derived for EMM and II, based on optimal values of the objective functions in (7), (10), (12), (14) and (16). These tests are invariant to reparameterization of the null hypothesis and tend to have better finite sample performance than Wald-type tests (e.g., see Hansen *et. al.*, 1996). In addition, they do not require evaluation of the derivative of the binding function $\mu(\theta)$ which is computationally expensive and often numerically unstable. The LR-type test statistics have the form:

$$\mathrm{LR}_S^i(\eta_0) = \frac{Sn}{S+1}\left[\mathrm{J}_S^i(\hat{\theta}_S^i(\eta_0)) - \mathrm{J}_S^i(\hat{\theta}_S^i)\right], \tag{25}$$

for $i =$ EL, EA, IL, IM, IA and where $\hat{\theta}_S^i(\eta_0)$ denotes the constrained estimator defined by

$$\hat{\theta}_S^i(\eta_0) = \arg\min_{\theta}\ \mathrm{J}_S^i(\theta) \text{ s.t. } q(\theta) = \eta_0 \tag{26}$$

Under $H_0 : q(\theta) = \eta_0$ it can be shown that for fixed S, $\mathrm{LR}_{Sn}^i \xrightarrow{d} \chi^2(p_1)$ as $n \to \infty$.

Confidence sets for individual elements θ_j may be constructed by defining $q(\theta) = \theta_{j,0}$ and inverting (25) using a $\chi^2(1)$ critical value. Specifically, a $(1-\alpha)\cdot 100\%$ confidence set for θ_i may be formed from

$$\{\theta_{j,0} : \mathrm{LR}_S^i(\theta_{j,0}) \leq \chi^2_{1-\alpha}(1)\} \tag{27}$$

For the EL estimator, Gallant and Tauchen (1996) provide a computationally efficient method for computing (27).

2.6 Overidentification Tests

When the auxiliary model $\tilde{F}_{\mu}$ has more parameters than the true model F_{θ}, the following scaled optimized value of the EMM and II objective function

$$\frac{Sn}{S+1}\mathrm{J}_S^i(\hat{\theta}_S^i), i = \text{EL, EA, IL, IM, and IA}. \tag{28}$$

can be used as a general specification test. Under the null hypothesis that F_θ is correctly specified and $\tilde{F}_\mu$ is a good approximation to F_θ then as $n \to \infty$, for fixed S, (28) has a limiting chi-squared distribution with $r - p$ degrees of freedom. See Gouriéroux *et al.* (1993) and Gallant and Tauchen (1996) for technical details.

2.7 Choice of S

In empirical applications, one has to choose S. Gouriéroux and Monfort (1996) show that the asymptotic bias of the II estimator does not depend on S whereas the asymptotic variance matrices of II and EMM estimators (18) and (17) are proportional to $(1+\frac{1}{S})$. Hence, the choice of S impacts more the variability of the EMM and II estimators than it does the bias. Gallant and Tauchen (1996) suggest choosing $S \cdot n$ sufficiently large so that the simulation noise is asymptotically negligible. In practice, however, the computation time of EMM and II estimators is linearly increasing with S and the elimination of the simulation noise using large simulated sample sizes can be computationally very expensive especially for II estimators.

We propose a simple method for choosing S that is motivated by efficiency considerations used in robust estimation based on bounded influence functions. Such robust estimators depend on a tuning parameter that controls the asymptotic efficiency of the estimator relative to the ML estimator in a non-contaminated model. Typically, the tuning parameter is set such that the relative efficiency loss of the robust estimator is less than some specified level such as five percent. To see how this idea can be carried over to EMM and II estimators, consider the case in which the auxiliary model $\widetilde{F}_\mu$ and the true model F_θ are the same, and the auxiliary ML estimator is a consistent estimator of the parameter θ. Then, $\mu(\theta) = \theta$, $\Sigma^* = \mathcal{I}^{-1} = -M_\mu^{-1}$ and Ω^* corresponds to the inverse of the asymptotic covariance matrix of the ML estimator. In addition,

$$\frac{\partial \mu(\theta)}{\partial \theta'} = I_p\,,\;\; M_\theta = -M_\mu \frac{\partial \mu(\theta)}{\partial \theta'} = \mathcal{I}\,,$$

with I_p the $(p \times p)$ identity matrix. Hence, from (18) and (17) the asymptotic variances of the II and EMM estimators are

$$W_{\text{II}} = W_{\text{EMM}} = \left(1 + \frac{1}{S}\right) \Omega^{*-1}\,,$$

and the asymptotic efficiency of the EMM and II estimators relative to the ML estimator is

$$\text{Eff}(\hat{\theta}_S) = \frac{\text{tr}(\Omega^{*-1})}{\left(1+\frac{1}{S}\right)\text{tr}(\Omega^{*-1})} = \left(1 + \frac{1}{S}\right)^{-1}$$

Then, for some $\varepsilon \in (0,1)$, the constant S can be chosen such that the asymptotic efficiency of the EMM/II estimator is bigger than $1 - \varepsilon$ when compared to the auxiliary ML estimator

$$S^* = \min\{S \in \mathbb{N} \,|\, \text{Eff}(\hat{\theta}_S) > 1 - \varepsilon\} = \left[\frac{1}{\varepsilon} - 1\right] + 1\,,$$

where $[x]$ denotes the integer part of $x \in \mathbb{R}$. For example, to reach an efficiency greater than 95%, one should set $S^* = 20$.

3 Finite Sample Properties of II and EMM

To illustrate the finite sample behavior of the EMM and II estimators and test statistics, we follow Gouriéroux *et. al.* (1993), Chumacero (1997, 2001), Michaelides and Ng (2000), de Luna and Genton (2001), Genton and Ronchetti (2003), Ghysels *et. al.* (2003), and consider estimation of a simple first order moving average (MA(1)) process. The MA(1) model is useful for analysis because simulations are easy to generate, auxiliary autoregressive models are simple to estimate, an analytic binding function exists, and comparisons with exact maximum likelihood (ML) are possible.

3.1 Monte Carlo Set-up

We consider the simple MA(1) model

$$F_\theta : y_t = \epsilon_t + \theta\epsilon_{t-1}, \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2) \tag{29}$$

for $t = 1, \ldots, n$ with $|\theta| < 1$ and $\sigma_\epsilon^2 > 0$. Since the MA(1) model is assumed to be invertible, it has an infinite order stationary autoregressive representation. As a result, EMM and II procedures can be based on auxiliary finite order autoregressive models which can be estimated efficiently by least squares. We utilize an auxiliary $AR(m)$ model

$$\widetilde{F}_\mu : y_t = \mu_1 y_{t-1} + \cdots + \mu_m y_{t-m} + \xi_t,\ \xi_t \sim \mathcal{N}(0, \sigma_\xi^2)\,. \tag{30}$$

As shown in Chumacero (2001) and Ghysels et. al. (2003), the binding function associated with (30) is

$$\begin{aligned} \mu_i(\theta) &= (-1)^{i-1}\theta^i \left[\frac{1-\theta^{2(m-i+1)}}{1-\theta^{2(m+1)}}\right],\ i = 1, \ldots, m,\ |\theta|\ \neq 1 && (31) \\ &= (-1)^{i-1}\theta^i(m-i+1)/(m+1), i = 1, \ldots, m,\ \theta = 1 \\ \sigma_\xi^2(\theta, \sigma_\epsilon^2) &= \frac{\sigma_\epsilon^2(1-\theta^{2(m+2)})}{1-\theta^{2(m+1)}},\ |\theta|\ \neq 1 && (32) \\ &= \sigma_\epsilon^2(m+2)/(m+1) \end{aligned}$$

As a result, non-simulation based versions of the EMM and II estimators based on (4) and (5) are also available[2].

For the Monte Carlo experiments, we consider two cases. In the first case, $\sigma_\epsilon^2 = \sigma_\xi^2 = 1$ is assumed to be known and hence $\mu = (\mu_1', \ldots, \mu_m')'$ and $r = m$. In the second case, $\sigma_\epsilon^2 = \sigma_\xi^2 = 1$

[2] Chumacero (2001) studied the finite sample behavior of non-simulation based EMM estimates of MA(1) and ARMA(1,1) models.

but σ_ϵ^2 and σ_ξ^2 are free parameters to be estimated and so $\mu = (\mu_1, \ldots, \mu_m, \sigma_\xi^2)'$ and $r = m+1$. We simulate 10,000 samples of size $n = 50$ and 200 from (29) with $\theta = 0,\ 0.1, \ldots,\ 0.9,\ 0.99$ to evaluate the finite sample biases and mean squared errors of the estimators, and we simulate 1,000 samples of size $n = 50$ and 200 with $\theta = 0.5, \ldots,\ 0.9$ to evaluate the empirical rejection frequencies of the test statistics[3].

We follow Ghysels *et. al.* (2003) and compute the EL, EA, IL, IM and IA estimators of θ using (30) with $m = 3$ and $m = 8$. We also compute the non-simulation based EMM estimator (denoted EN) based on (4). We do not use model selection criteria to select m for a given sample. Results from previous studies (see below) have shown that EMM and II perform poorly for the MA(1) model if m is set too small especially if θ is close to the noninvertible boundary. In computing the EMM and II estimators, we use estimates of the optimal weight matrices Σ^* and Ω^* and set $S = 20$. We compute finite sample biases and root mean squared errors (RMSEs) of the EMM and II estimators, as well as the LR-type test statistic (25) for testing the hypothesis $H_0 : \theta = \theta_0$ and the overidentification test statistic (28). Where appropriate, we also compare results to those computed from exact ML estimation based on the Kalman filter[4].

3.2 Previous Studies

Several papers studied the performance of EMM and II for the MA(1) model, but no paper conducted a comprehensive comparative analysis of the two types of EMM and three types of II estimators with respect to both estimation and inference. Gouriéroux *et. al.* (1993) and Gouriéroux and Monfort (1996) considered only the MA(1) model with $\theta = -0.5$ and $n = 250$, and computed the II estimator with $S = 1$ and $\tilde{\Omega} = I_m$ using (30) with $m = 1, 2$ and 3. For $m = 3$, they found that II performed comparably to exact ML in terms of bias and RMSE. Ghysels *et. al.* (2003) focused on the performance of EMM (EA and EL) and II (IM and IL) estimators near the noninvertible boundary ($\theta = 1$) of the parameter space. They considered models with $\theta = 0.1,\ 0.5,\ 0.9,\ 0.99$, $n = 50$ and 200, and computed estimators using (30) with $m = 8$, $\tilde{\Omega} = \tilde{\Sigma} = I_m$, and $S = 1$ and 3. They found that EMM and II performed similarly in terms of bias and RMSE for $\theta \leq 0.5$, but that II performs much better and is more stable than EMM near the boundary.

Chumacero (1997) studied both estimation and inference performance of EMM (EL) for the MA(1) model with $\theta = 0.5$ and $n = 250$ using (30) with $m = 2$ and 3. In contrast to the previous studies, Chumacero treated $\sigma^2 = \text{var}(\epsilon_t)$ as a free parameter to be estimated and he used the optimal weight matrix for $\tilde{\Sigma}$. He found that EMM performed similarly to exact ML in terms of bias and RMSE. In terms of inference, he found that LR-type tests for the individual hypotheses $H_0 : \theta = 0.5$ and $H_0 : \sigma^2 = 1$ had little size distortion but that LR-type tests for the joint hypothesis

[3] We only use 1,000 Monte Carlo trials for the test statistics because we compare them to the saddlepoint tests later on.

[4] For computational efficiency, the results for the EMM and II estimators are based on custom C code available upon request from the first author. The exact ML estimation was performed using S-PLUS 8.0 and S+FinMetrics 3.0. The restriction $|\theta| < 1$ was imposed for the exact ML estimator but not for the EMM or II estimators. The EMM and II estimates with $|\theta| < 1$ were almost identical to those with θ free.

$H_0 : \theta = 0.5$ and $\sigma^2 = 1$ was moderately size distorted. He also found that the overidentification test was slightly oversized. Michaelides and Ng (2000) compared EMM (EL) and II (IL) for the MA(1) model with $\theta = -0.5$, $n = 100$, 200 and 1,000 using (30) with $m = 3$. They computed the EMM and II estimators using the optimal weight matrices and considered $S = 10$ and 50. They found that II was slightly more accurate than EMM in terms of bias and RMSE, and that overidentification test for II was substantially less size distorted than for EMM. Chumacero (2001) also compared the EL and IL estimators for the MA(1) model with $\theta = -0.5$, -0.95 and $n = 100$, 200. Instead of fixing m in (30), he chose m using various model selection criteria and found that the Akaike information criterion (AIC) performed the best. He found that $m = 3$ was typically selected for $\theta = -0.5$ and $m = 8$ was often chosen for $\theta = -0.95$. He found that II was less biased than EMM with $\theta = -0.95$ but that II was numerically more unstable than EMM. He also found size distortions in the overidentification tests and these distortions increased as θ approached -1.

To summarize the results from previous studies, EMM and II perform similarly to exact ML in terms of bias and RMSE for $|\theta| < 0.5$, and that II is more accurate than EMM for $|\theta| \approx 1$. There is some evidence that coefficient tests based on EMM and II show some size distortion in small samples. Overidentification tests based on EMM and II are size distorted for small samples and that EMM is more size distorted than II.

3.3 Results

The results from our Monte Carlo analysis of the MA(1) model with $\sigma_\epsilon^2 = 1$ and known (case I) are presented in Figures 1 and 2 and Table 1. Figures 1 and 2 present the biases (upper panels) and the RMSE (lower panels) values of the different estimators based on AR(3) and AR(8) auxiliary models, respectively, for $n = 50$ (on the left) and $n = 200$ (on the right). As expected, the ML estimator shows the best performance. No simulation-based estimator uniformly dominates in terms of bias or RMSE. For $n = 50$, the II estimators perform the best in terms of bias for $\theta < 0.5$, the IL estimator has the best overall performance and the EL estimator has the worst performance. Interestingly, the biases of the II estimators display a hump-shaped pattern for $0.5 < \theta < 1$, peaking at θ near 0.8. The IA estimator for the MA(1) model does not seem to exhibit the bias reduction qualities shown by Gouriéroux *et. al.* (2001) and Duffee and Stanton (2007) for the AR(1) model. When θ is near the noninvertibility region, we obtain similar results to Ghysels *et al.* (2003); that is, the EMM estimators exhibit larger biases than the II estimators. However, our results show that the IA and IM estimators show greater biases than the EMM estimators for θ near 0.8. In terms of RMSE, the IL and EL estimators outperform the other estimators. For $n = 200$, all estimators except EA perform similarly for $\theta < 0.5$ and the II estimators again display hump-shaped behavior for θ values near 0.8. The EL shows the lowest RMSE except for θ at the boundary. Overall, the IL estimator performs the best in terms of bias and the EL estimator performs best in terms of RMSE. Interestingly, the IA and IM estimators are nearly identical in terms of bias and RMSE. Given the computational advantage of the IA estimator, it is to be preferred over the IM estimator.

Table 1 presents the case I ($\sigma_\epsilon^2 = 1$ and known) empirical rejection frequencies of nominal 5%

		$\alpha = 0.05$							$\alpha = 0.01$						
θ	m	ML	EN	EL	EA	IL	IA	IM	ML	EN	EL	EA	IL	IA	IM
		$n = 50$							$n = 50$						
0.5	3	0.05	0.10	0.17	0.17	0.08	0.09	0.09	0.01	0.03	0.09	0.09	0.03	0.03	0.03
	8		0.20	0.44	0.43	0.22	0.24	0.20		0.11	0.33	0.32	0.11	0.13	0.10
0.6	3	0.05	0.08	0.17	0.17	0.08	0.06	0.09	0.01	0.02	0.09	0.08	0.02	0.02	0.02
	8		0.21	0.43	0.43	0.20	0.21	0.22		0.10	0.33	0.32	0.10	0.11	0.11
0.7	3	0.06	0.06	0.16	0.15	0.06	0.04	0.05	0.01	0.01	0.08	0.08	0.01	0.01	0.01
	8		0.20	0.43	0.44	0.20	0.19	0.23		0.10	0.32	0.32	0.08	0.08	0.11
0.8	3	0.04	0.05	0.16	0.15	0.03	0.05	0.03	0.01	0.02	0.09	0.08	0.01	0.01	0.01
	8		0.16	0.43	0.45	0.15	0.16	0.15		0.06	0.32	0.32	0.06	0.07	0.06
0.9	3	0.02	0.05	0.18	0.17	0.04	0.05	0.03	0.01	0.02	0.11	0.08	0.02	0.01	0.01
	8		0.09	0.42	0.43	0.08	0.09	0.09		0.04	0.31	0.32	0.04	0.04	0.04
		$n = 200$							$n = 200$						
0.5	3	0.06	0.05	0.09	0.06	0.05	0.06	0.04	0.01	0.02	0.03	0.01	0.02	0.01	0.01
	8		0.08	0.15	0.12	0.08	0.09	0.07		0.03	0.08	0.06	0.02	0.02	0.02
0.6	3	0.05	0.05	0.09	0.06	0.04	0.05	0.05	0.01	0.01	0.03	0.01	0.00	0.02	0.00
	8		0.07	0.14	0.12	0.07	0.07	0.08		0.02	0.07	0.05	0.02	0.02	0.02
0.7	3	0.06	0.05	0.07	0.05	0.04	0.04	0.04	0.01	0.01	0.03	0.01	0.00	0.00	0.01
	8		0.07	0.13	0.11	0.09	0.08	0.07		0.02	0.06	0.04	0.03	0.02	0.02
0.8	3	0.04	0.02	0.07	0.04	0.02	0.02	0.03	0.01	0.00	0.03	0.01	0.00	0.01	0.00
	8		0.07	0.12	0.11	0.08	0.07	0.08		0.01	0.05	0.04	0.02	0.02	0.01
0.9	3	0.06	0.02	0.07	0.05	0.02	0.03	0.02	0.02	0.00	0.04	0.02	0.01	0.00	0.00
	8		0.03	0.13	0.12	0.03	0.03	0.0		0.01	0.05	0.03	0.01	0.01	0.00

Table 1: Empirical rejection frequencies of LR-type coefficient tests.

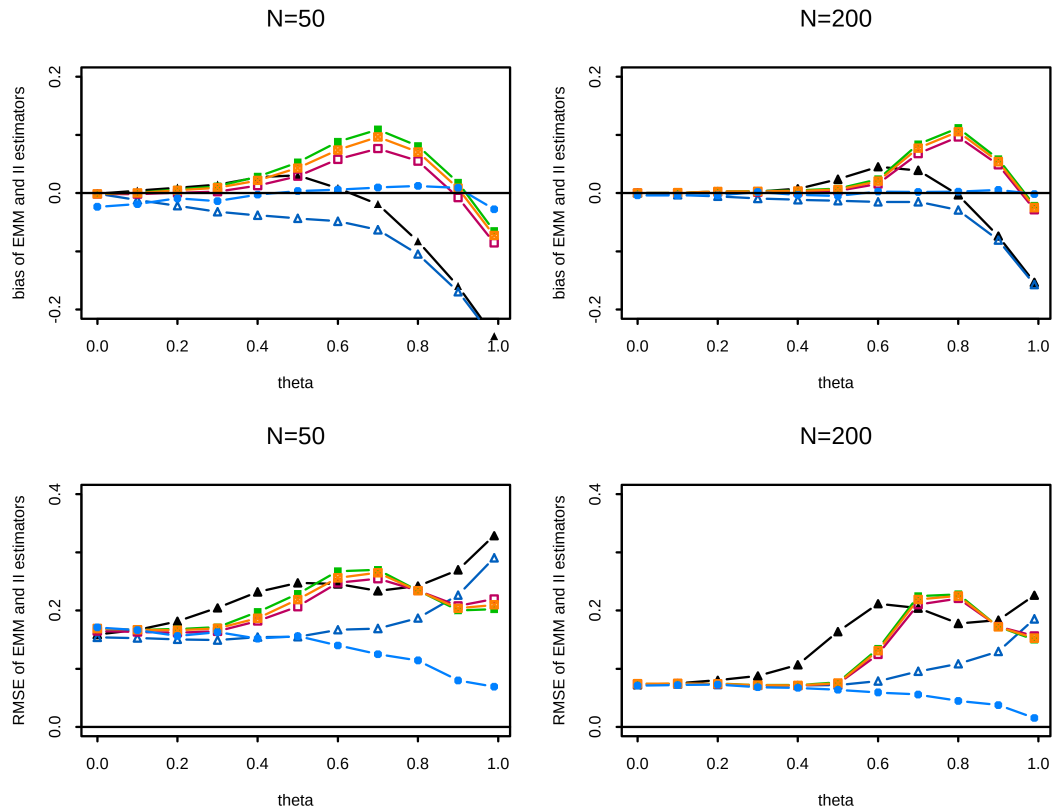

Figure 1: Biases (on the top) and RMSE (on the bottom) of the EL (empty triangles), EA (filled triangles), IL (empty squares), IA (squares with crosses), IM (filled squares), and ML (filled circles) estimators of a MA(1) process with $\theta = 0,\ 0.1, \ldots, 0.9,\ 0.99$ based on AR(3) auxiliary models.

and 1% LR-type tests for the null hypothesis $H_0 : \theta = 0.5,\ 0.6, \ldots, 0.9$ based on AR(3) and AR(8) auxiliary models for samples of size $n = 50$ and $n = 200$. The empirical rejection frequencies of the overidentification tests mirror those of the coefficient tests and are therefore omitted. For both sample sizes and all values of θ, the LR test based on the ML estimator is correctly sized. In general, the size distortions for the EMM and II estimators are larger with the AR(8) model and the tests based on the II estimators have about half the size distortion as the tests based on the EMM estimators. Interestingly, the results based on the non-simulation based EMM estimator are very similar to the simulation-based II estimators. The empirical rejection frequencies of the tests based on the EL and EA estimators are stable across θ whereas the tests based on the EN and II estimators tend to decline as θ approaches unity. The poor performance of the EN and II tests for θ near unity occurs because the binding function becomes discontinuous at $\theta = 1$. For the AR(8) model with $n = 50$, the tests based on the EMM estimators are substantially size distorted for all value of $\theta :$ the 5% tests have empirical sizes around 44%. These large size distortions imply highly inaccurate confidence intervals formed by inverting the LR-type tests. The larger size distortions of the tests based on the AR(8) auxiliary model with $n = 50$ arises because the efficient weight

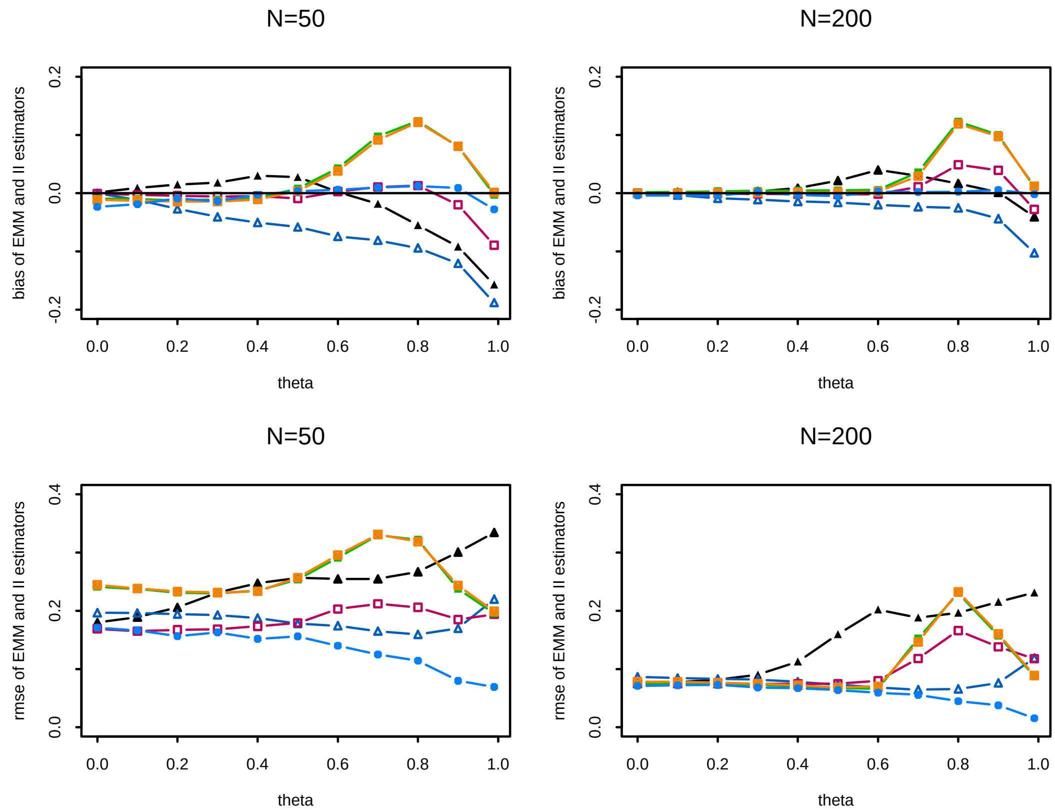

Figure 2: Biases (on the top) and RMSE (on the bottom) of the EL (empty triangles), EA (filled triangles), IL (empty squares), IA (squares with crosses), IM (filled squares), and ML (filled circles) estimators of a MA(1) process with $\theta = 0,\, 0.1, \ldots, 0.9,\, 0.99$ based on AR(8) auxiliary models.

matrices are not accurately estimated in small samples. For $n = 200$, the size distortions of all tests drop substantially but the EMM tests still exhibit a moderate amount of size distortion. The 5% tests based on the EMM estimators have empirical rejection rates around 15% for all values of θ.

4 Saddlepoint Coefficient Tests for EMM and II Estimators

The LR-type test statistics described in subsection 2.2 are based on first order asymptotic theory. Asymptotic normality of EMM and II estimators imply that the LR-type statistics defined in (25) are asymptotically χ^2-distributed. The χ^2 approximation is then used to compute p-values for a hypothesis test. However, as illustrated by the Monte Carlo results of the previous section, p-values obtained using the χ^2 approximation can be highly inaccurate for small to moderate sample sizes. The aim of this section is to construct coefficient test statistics based on EMM and II such that the distribution of the statistics can be approximated with a relative error of order n^{-1} or $n^{-1/2}$.

To improve the accuracy of the asymptotic approximation, high-order approximation methods have been proposed. The most frequently used are Edgeworth expansions (cf., for instance, Feller,

1971) and saddlepoint approximations (Daniels, 1954). The superiority of the saddlepoint approximation for tail probabilities, which are the quantities of interest for inference tests and confidence intervals, when compared to the Edgeworth expansion is shown, for instance, in Barndorff-Nielsen and Cox (1989) and Field and Ronchetti (1990).

Our proposed saddlepoint tests of hypotheses on coefficients estimated by EMM/II estimators is based on the fact that we can associate an EMM/II estimator with an asymptotically equivalent M-estimator defined by the influence function of the EMM/II estimator, which allows us to use the saddlepoint coefficient test for M-estimators introduced by Robinson, Ronchetti and Young (2003), hereafter referred to as RRY.

4.1 RRY Coefficient Tests for M-Estimators

Let $\{X_i\}_{i=1,\ldots,n}$ be an i.i.d. sample of random vectors with common range $\mathcal{X}$ and distribution F_θ with $\theta \in \mathbb{R}^p$. Define the M-functional T that for a given distribution F associates the parameter $T(F) \in \mathbb{R}^p$ defined by

$$E_F\big[\psi(X,T)\big] = 0\,,$$

where the score ψ is assumed to be a smooth function from $\mathcal{X} \times \mathbb{R}^p$ to $\mathbb{R}^p$. Consider a sample of observations $\{x_i\}_{i=1,\ldots,n}$ and define the empirical distribution F_n as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{x_i}(x)\,, \tag{33}$$

where Δ_{x_i} denotes the pointmass distribution, the probability measure which gives mass 1 to x_i. The M-functional evaluated at the empirical distribution F_n is $T_n = T(F_n)$ defined by

$$E_{F_n}\big[\psi(X,T)\big] = \frac{1}{n}\sum_{i=1}^{n}\int \psi(x,T)d\Delta_{x_i}(x) = \frac{1}{n}\sum_{i=1}^{n}\psi(x_i,T)\,.$$

Hence, the M-estimator of θ is $T_n(X_1,\ldots,X_n) = T(F_n)(X_1,\ldots,X_n)$ defined by

$$\sum_{i=1}^{n}\psi(X_i,T_n) = 0\,. \tag{34}$$

To simplify the notation, hereafter we denote the statistic $T_n \equiv T_n(X_1,\ldots,X_n)$ and its observed value by $t_n \equiv T_n(x_1,\ldots,x_n)$.

Using the M-estimator T_n, we would like to test hypotheses of the form $H_0 : q(\theta) = \eta_0$, where q is a smooth function from $\mathbb{R}^p$ to $\mathbb{R}^{p1}$.

4.1.1 Analytical RRY Test

RRY consider the case where the cumulant generating function of the vector of scores exists and under $H_0 : q(\theta) = \eta_0$ is defined by

$$K_\psi(\lambda, \theta) = \log E_{F_{\theta(\eta_0)}} \left[e^{\lambda' \psi(X,\theta)} \right] . \tag{35}$$

Under the assumption that the density of the M-estimator exists and has a saddlepoint approximation as given in Field (1982), RRY derive a saddlepoint approximation to the density of $q(T_n)$ and propose the test statistic

$$2nh(q(T_n)) , \tag{36}$$

where

$$h(y) = \inf_{\{\theta \,:\, q(\theta) = y\}} \sup_\lambda \{ -K_\psi(\lambda; \theta) \} . \tag{37}$$

Using the saddlepoint approximation to the density of $q(T_n)$, RRY show that the p-value is of the form

$$p = P(h(q(T_n)) \geq h(q(t_n))) = \big(1 - Q_{p_1}(2nh(q(t_n)))\big)(1 + \mathcal{O}(n^{-1} + 2\, h(q(t_n)))) , \tag{38}$$

where Q_{p_1} is the cumulative distribution function of $\chi^2(p_1)$. This approximation has a relative error $\mathcal{O}(1/n)$ in the normal region $\sqrt{2h(q(t_n))} = \mathcal{O}(1/\sqrt{n})$:

$$p = P(h(q(T_n)) \geq h(q(t_n))) = \big(1 - Q_{p_1}(2nh(q(t_n)))\big)\big(1 + \mathcal{O}(n^{-1})\big) . \tag{39}$$

As a result, p-values for the saddlepoint test (36) based on the chi-square distribution are expected to be more accurate in finite samples than those for the Wald or LR-type tests whose approximations have absolute error $\mathcal{O}(1/\sqrt{n})$. According to RRY, results in (38) and (39) can be extended to the case when X_i are not identically distributed.

Simple Hypothesis

Consider the simple case when $q(\theta) = \theta$ and we are interested in testing the hypothesis $H_0 : \theta = \theta_0$. In this case, the analytical RRY test statistic is

$$2nh(T_n) ,$$

where the function h is simply

$$h(y) = \sup_\lambda \{ -K_\psi(\lambda, y) \} \tag{40}$$

with

$$K_\psi(\lambda, \theta) = \log E_{F_{\theta_0}} \left[e^{\lambda' \psi(X,\theta)} \right] .$$

In the case when the model belongs to the exponential family $f_\theta(x) = c(\theta) e^{\theta' t(x)}$ and T_n is an ML estimator it is straightforward to show that (40) is equivalent to the log-likelihood ratio statistic.

Composite Hypothesis

Suppose that $\theta = (\theta_1', \theta_2')'$, $\theta_1 \in \mathbb{R}^{p_1}$ and $\theta_2 \in \mathbb{R}^{p_2}$. Consider the case when $q(\theta) = \theta_1$ and we are interested in testing the hypothesis $H_0 : \theta_1 = \theta_{1\,0}$. In this case, the analytical RRY test statistic is

$$2nh(T_{n1}),$$

where the function h is defined by

$$h(y) = \inf_{\theta_2} \sup_{\lambda} \{-K_\psi(\lambda, (y, \theta_2))\} \tag{41}$$

with

$$K_\psi(\lambda, \theta) = \log E_{F_{(\theta_{1\,0}, \theta_2)}} \left[e^{\lambda' \psi(X, \theta)} \right].$$

Essentially, the nuisance parameter θ_2 is concentrated out of $K_\psi(\lambda, y)$ when forming the test statistic for $H_0 : \theta_1 = \theta_{1\,0}$.

4.1.2 Empirical RRY Test

In practice, the distribution F_θ may be unknown and even when it is known, the cumulant generating function may not exist. For these purposes, RRY defined an empirical version of the test based on an exponentially weighted empirical cumulant generating function that imposes $H_0 : q(\theta) = \eta_0$. Define an empirical distribution $\hat{F}_0$ consistent with the null hypothesis:

$$\hat{F}_0(x) = \frac{\sum_{i=1}^n e^{\beta(\eta_0)' \psi(x_i, \theta(\eta_0))} \Delta_{x_i}(x)}{\sum_{i=1}^n e^{\beta(\eta_0)' \psi(x_i, \theta(\eta_0))}},$$

where $\beta = \beta(\eta_0)$, $\theta = \theta(\eta_0)$ are chosen to minimize the backward Kullback–Leibler distance between the empirical distribution and the tilted empirical distribution subject to

$$E_F[\psi(X, \theta)] = 0,$$

see DiCiccio and Romano (1990). The solutions to this minimization are solutions of the equations

$$\frac{\partial \kappa}{\partial \beta}(\beta, \theta) = 0, \tag{42}$$

$$q(\theta) = \eta_0, \tag{43}$$

$$\frac{\partial \kappa}{\partial \theta}(\beta, \theta) = \frac{\partial q(\theta)'}{\partial \theta} \gamma, \tag{44}$$

where

$$\kappa(\beta, \theta) = \log \left[\frac{1}{n} \sum_{i=1}^n e^{\beta' \psi(x_i, \theta)} \right]$$

and $\gamma = \gamma(\eta_0)$ is the Lagrange multiplier of the optimization problem. Notice that the saddlepoint $\beta(\theta)$ in equation (42) is solution to the following optimization problem:

$$\beta(\theta) = \arg\max_{\beta}(-\kappa(\beta, \theta))\,. \tag{45}$$

The empirical cumulant generating function is then defined by

$$K_{\psi}^{\omega}(\lambda,\theta) = \log E_{\hat{F}_0}[e^{\lambda'\psi(X,\theta)}] = \log \sum_{i=1}^{n} \underbrace{\frac{e^{\beta(\eta_0)'\psi(x_i,\theta(\eta_0))}}{\sum_{i=1}^{n} e^{\beta(\eta_0)'\psi(x_i,\theta(\eta_0))}}}_{\omega_i} \int e^{\lambda'\psi(x,\theta)}\Delta_{x_i}(x) = \log \sum_{i=1}^{n} \omega_i e^{\lambda'\psi(x_i,\theta)}\,,$$

which corresponds to a weighted empirical cumulant generating function where the weights are consistent with the null hypothesis. The empirical RRY test is then defined by

$$2n\hat{h}(q(T_n))\,, \tag{46}$$

with

$$\hat{h}(y) = \inf_{\{\theta\,:\,q(\theta)=y\}} \sup_{\lambda}\{-K_{\psi}^{\omega}(\lambda,\theta)\}\,. \tag{47}$$

For this empirical version of the test RRY did not prove that the relative error of the χ^2 approximation is of order given in (38)-(39). Given that the empirical saddlepoint approximation to the density of an M-estimator is of relative error of order $\mathcal{O}(1/\sqrt{n})$ (see Sowell, 2007) we should expect the same relative error of the χ^2 approximation to the distribution of the empirical RRY test statistic (36). They do not consider cases where X_i are not independent However, our simulations in Section 5 show that even in the case when the X_i are not independent but the scores follow a stationary and ergodic martingale difference sequence the χ^2 approximation to the distribution of the empirical RRY statistic (36) lead to more accurate inference than those obtained using the χ^2 approximation to the distribution of an LR-typestatistic.

Weights For the Simple Hypothesis

In the simple case when $q(\theta) = \theta$ and the test hypothesis is $H_0 : \theta = \theta_0$, the empirical distribution $\hat{F}_0$ is determined by the equations (42) and (43). The weights are then defined by

$$\omega_i = \frac{e^{\beta(\theta_0)'\psi(x_i,\theta_0)}}{\sum_{i=1}^{n} e^{\beta(\theta_0)'\psi(x_i,\theta_0)}}\,,$$

where $\beta(\theta_0)$ is defined by (45).

Weights For the Composite Hypothesis

Consider the case when $q(\theta) = \theta_1$ and the test hypothesis is $H_0 : \theta_1 = \theta_{1\,0}$. Equation (43) fixes the p_1 parameters in θ_1 and the parameters in θ_2, determined by equations (42)-(44), are solutions to the following minimization problem

$$\theta_2^* = \min_{\theta_2} \Big(-\kappa\big(\beta(\theta_{1\,0}, \theta_2), (\theta_{1\,0}, \theta_2)\big)\Big)\,,$$

where $\beta(\theta)$ is defined by (45). Indeed, the Lagrangian to this minimization problem is

$$\mathcal{L}(\theta, \gamma) = -\kappa(\beta(\theta), \theta) + \gamma'(q(\theta) - \theta_{1\,0})\,,$$

and the first order conditions are

$$\frac{\partial \mathcal{L}}{\partial \theta}(\theta, \gamma) = -\frac{\partial \beta(\theta)'}{\partial \theta}\underbrace{\frac{\partial \kappa}{\partial \beta}(\beta(\theta), \theta)}_{0} + \frac{\partial q(\theta)'}{\partial \theta}\gamma = 0\,,$$

$$\frac{\partial \mathcal{L}}{\partial \gamma}(\theta, \gamma) = q(\theta) - \theta_{1\,0} = 0\,.$$

4.2 Asymptotically Equivalent M-Estimators for II and EMM Estimators

The EMM and II estimators considered in this paper are not defined as M-estimators of the form (34), so it is not obvious that the RRY tests can be applied. However, using properties of the influence functions (Hampel, 1974; Hampel *et al.*, 1986) for EMM and II estimators we can associate asymptotically equivalent M-estimators and use the score function from these M-estimators to form the RRY tests evaluated at the EMM/II estimates.

The influence function for non-simulation-based II estimators defined in (**??**) has been derived by Genton and de Luna (2000), and for non-simulation-based EMM estimators (**??**) by Ortelli and Trojani (2005). To describe these influence functions, let $\widetilde{T}$ be the functional associated with the auxiliary ML estimator $\tilde{\mu}$ defined in (1). That is, $\tilde{\mu} = \widetilde{T}(F_n)$ where F_n is the empirical distribution (**??**). Denote by T_{II} the functional associated with the II estimators and T_{EMM} the functional associated with the EMM estimators. Assume that Fisher consistency holds; that is, $\widetilde{T}(F_\theta) = \mu$ and $T_{\text{II}}(F_\theta) = T_{\text{EMM}}(F_\theta) = \theta$. To economize on notation in what follows, let y and x denote y_0 and x_{-1}, respectively. The influence functions of the EMM and II estimators based on arbitrary fixed weight matrices Ω and Σ are, respectively:

$$\text{IF}(y, T_{\text{EMM}}, F_\theta) = -\left(M_\theta'\Sigma M_\theta\right)^{-1} M_\theta'\Sigma M_\mu \text{IF}(y, \widetilde{T}, F_\theta) \tag{48}$$

$$\text{IF}(y, T_{\text{II}}, F_\theta) = \left(\frac{\partial \mu(\theta)'}{\partial \theta'}\Omega\frac{\partial \mu(\theta)}{\partial \theta'}\right)^{-1}\frac{\partial \mu(\theta)'}{\partial \theta'}\Omega\,\text{IF}(y, \widetilde{T}, F_\theta)\,, \tag{49}$$

where $\text{IF}(y, \widetilde{T}, F_\theta)$ is the influence function of the auxiliary estimator $\tilde{\mu}$. For the case in which the auxiliary estimator $\tilde{\mu}$ is an M-estimator defined by $\sum_t \frac{\partial \tilde{f}}{\partial \mu}(y_t; x_{t-1}, \tilde{\mu}) = 0$, the influence function of the auxiliary estimator $\tilde{\mu}$ is given by

$$\mathrm{IF}(y, \widetilde{T}, F_\theta) = -M_\mu^{-1} \frac{\partial \tilde{f}}{\partial \mu}(y; x, \mu(\theta)) \tag{50}$$

Since the EMM and II estimators based on the optimal weight matrices Σ^* and Ω^* are asymptotically equivalent under suitable regularity conditions, it turns out that influence functions of the optimal EMM and II estimators are also equivalent. This result follows directly from (19) and (24).

A Fisher consistent estimator is asymptotically equivalent to the M-estimator defined by the influence function of the Fisher consistent estimator (see Hampel *et. al.*, 1986, p. 231; Newey and McFadden, 1993; Czellar and Ronchetti, 2008). Denote by $\hat{\theta}^{\mathrm{ae}}_{\mathrm{II}}(\Omega)$ and $\hat{\theta}^{\mathrm{ae}}_{\mathrm{EMM}}(\Sigma)$ the M-estimators that are asymptotically equivalent to the II and EMM estimators using the weight matrices Ω and Σ, respectively. Using (48)-(50), $\hat{\theta}^{\mathrm{ae}}_{\mathrm{II}}(\Omega)$ and $\hat{\theta}^{\mathrm{ae}}_{\mathrm{EMM}}(\Sigma)$ are defined by the first order conditions:

$$\sum_t \zeta_{\mathrm{II}}(y_t, \hat{\theta}^{\mathrm{ae}}_{\mathrm{II}}(\Omega), \Omega) = 0 \text{ and } \sum_t \zeta_{\mathrm{EMM}}(y_t, \hat{\theta}^{\mathrm{ae}}_{\mathrm{EMM}}(\Sigma), \Sigma) = 0\,, \tag{51}$$

where

$$\zeta_{\mathrm{II}}(y, \theta, \Omega) = \frac{\partial \mu(\theta)'}{\partial \theta} \Omega\, M_\mu^{-1} \frac{\partial \tilde{f}}{\partial \mu}(y; x, \mu(\theta))\,, \tag{52}$$

$$\zeta_{\mathrm{EMM}}(y, \theta, \Sigma) = M_\theta'\, \Sigma\, \frac{\partial \tilde{f}}{\partial \mu}(y; x, \mu(\theta))\,. \tag{53}$$

The score functions (52) and (53) are identical when evaluated at the optimal weight matrices Ω^* and Σ^* given by (19), respectively, and become

$$\zeta^*(y, \theta, \mathcal{I}^{-1}) = M_\theta'\, \mathcal{I}^{-1}\, \frac{\partial \tilde{f}}{\partial \mu}(y; x, \mu(\theta))\,. \tag{54}$$

In most applications the quantities $\mu(\theta)$, M_μ, M_θ, and $\mathcal{I}$ in the score functions (52) - (54) are unknown and must be estimated. The asymptotically equivalent M-estimator based on an estimated version of the optimal score (54) is equivalent to the non-simulation based EMM estimator defined in (4) with $\tilde{\Sigma} = \widetilde{\mathcal{I}}^{-1}$. Indeed, the first order conditions from (4) are

$$\widehat{M}'_{\hat{\theta}^{\mathrm{ae}}} \widetilde{\mathcal{I}}_n^{-1} \tilde{g}_n(\hat{\theta}^{\mathrm{ae}}) = 0$$

where

$$\widehat{M}_{\hat{\theta}^{\mathrm{ae}}} = \frac{\partial}{\partial \theta'} \left(\frac{1}{n-m} \sum_{t=m+1}^{n} \frac{\partial \tilde{f}}{\partial \mu}(y_t; x_{t-1}, \mu(\hat{\theta}^{\mathrm{ae}})) \right)$$

which is of the form$\sum_{t=m+1}^{n} \hat{\zeta}^*(y_t, \hat{\theta}^{\mathrm{ae}}, \widetilde{\mathcal{I}}_n^{-1}) = 0$ with $\hat{\zeta}^*(y_t, \hat{\theta}^{\mathrm{ae}}, \widetilde{\mathcal{I}}_n^{-1}) = \widehat{M}'_{\hat{\theta}^{\mathrm{ae}}} \widetilde{\mathcal{I}}_n^{-1} \frac{\partial \tilde{f}}{\partial \mu}(y_t; x_{t-1}, \mu(\hat{\theta}^{\mathrm{ae}}))$.

4.3 Empirical RRY Tests Based on Asymptotically Equivalent M-Estimators

Using the M-estimator score functions (52) and (53), RRY tests for the hypothesis $H_0 : q(\theta) = \eta_0 \in \mathbb{R}^{p_1}$ based on the asymptotically equivalent M-estimates defined by (51) are given by (36) with T_n replaced by $\hat{\theta}^{\text{ae}}_{\text{II}}(\Omega)$ or $\hat{\theta}^{\text{ae}}_{\text{EMM}}(\Sigma)$, respectively. Under the conditions stated in RRY, these tests are then $\chi^2_{p_1}$-distributed with a relative error of order n^{-1} under H_0. Given that we associate the asymptotically equivalent M-estimators with the simulation-based EMM and II estimators, we propose the following empirical RRY tests for EMM and II:

$$2\frac{nS}{S+1}\hat{h}(q(\hat{\theta}^i_S(\tilde{\Sigma})),\ i = \text{EL}, \text{EA} \tag{55}$$

$$2\frac{nS}{S+1}\hat{h}(q(\hat{\theta}^i_S(\tilde{\Omega})),\ i = \text{IL}, \text{IM}, \text{IA} \tag{56}$$

where the function $\hat{h}$ is defined in (47) and uses the score functions (53) and (52) for EMM and II, respectively. If the efficient EMM or II estimators are computed then the score function (54) may be used in the construction of the test for either estimator. We expect that inference based on the empirical RRY tests (55)-(55) to be more accurate in finite samples than inference based on the classical LR-type statistics (25). We provide Monte Carlo evidence to back up this claim in the next section.

We make the following remarks regarding the implementation of the RRY tests for EMM and II estimators**:**

1. When the empirical RRY tests are evaluated at the simulation-based EMM or II estimates, they must by scaled by $\frac{S}{S+1}$ to account for the increase in variability due to simulations.
2. The tests in (55) and (56) are defined for arbitrary weight matrices, not just the efficient weights matrices required for computing the LR-type tests.
3. The binding function, $\mu(\theta)$, and the limit quantities, M_θ and M_μ, needed for the computation of the empirical RRY tests are typically unknown, but can be approximated using the pseudo-data generated for the computation of the EMM and II estimators as follows. For the RRY tests using EL estimates, replace $\mu(\theta)$ by $\tilde{\mu}^L_S$, and replace $E_{F_\theta}\left[\frac{\partial \tilde{f}}{\partial \mu}(y; x, \mu)\right]$ in M_θ by $\tilde{g}_{Sn}(\theta, \mu)$. For the RRY test using EA estimates, replace $\mu(\theta)$ by $\tilde{\mu}^{\text{A}}(\theta)$ and replace $E_{F_\theta}\left[\frac{\partial f}{\partial \mu}(y; x, \mu)\right]$ by $S^{-1}\sum_{s=1}^S \tilde{g}^s_n(\theta, \mu)$. For the RRY test using IL, IM and IA estimates, replace $\mu(\theta)$ by $\tilde{\mu}^{\text{L}}_S(\theta)$, $\tilde{\mu}^{\text{M}}_S(\theta)$,and $\tilde{\mu}^{\text{A}}_S(\theta)$ defined by (13), (14) and (19), respectively. The quantity $E_{F_\theta}\left[\frac{\partial^2 \tilde{f}}{\partial \mu \partial \mu'}(y; x, \mu)\right]$ can be replaced by $\frac{\partial}{\partial \mu'}\tilde{g}_{Sn}(\theta, \mu)$ for IL, and by $\frac{\partial}{\partial \mu'}S^{-1}\sum_{s=1}^S \tilde{g}^s_n(\theta, \mu)$ for IA and IM. These approximations are summarized in Table 2.
4. The derivative of the binding function, $\frac{\partial \mu(\theta)}{\partial \theta}$, can be calculated numerically, for example using Ridders method of polynomial extrapolation.

$\hat{\theta}^i_{Sn}$	$\zeta(y,\theta)$	$\mu(\theta)$	M_θ	M_μ
EL	$M_\theta' \, \Sigma^* \frac{\partial \tilde{f}}{\partial \mu}(y;x,\mu(\theta))$	$\tilde{\mu}^{\mathrm{L}}_S(\theta)$	$\frac{\partial}{\partial \theta'} \tilde{g}_{Sn}(\theta,\mu)$	–
EA	$M_\theta' \, \Sigma^* \frac{\partial \tilde{f}}{\partial \mu}(y;x,\mu(\theta))$	$\tilde{\mu}^{\mathrm{A}}_S(\theta)$	$\frac{\partial}{\partial \theta'} \frac{1}{S} \sum_{s=1}^S \tilde{g}^s_n(\theta,\mu)$.	–
IL	$\frac{\partial \mu(\theta)'}{\partial \theta'} \Omega^* \, M_\mu^{-1} \frac{\partial \tilde{f}}{\partial \mu}(y;x,\mu(\theta))$	$\tilde{\mu}^{\mathrm{L}}_S(\theta)$	–	$\frac{\partial}{\partial \mu'} \tilde{g}_{Sn}(\theta,\mu)$
IA	$\frac{\partial \mu(\theta)'}{\partial \theta'} \Omega^* \, M_\mu^{-1} \frac{\partial \tilde{f}}{\partial \mu}(y;x,\mu(\theta))$	$\tilde{\mu}^{\mathrm{M}}_S(\theta)$	–	$\frac{\partial}{\partial \mu'} S^{-1} \sum_{s=1}^S \tilde{g}^s_n(\theta,\mu)$
IM	$\frac{\partial \mu(\theta)'}{\partial \theta'} \Omega^* \, M_\mu^{-1} \frac{\partial \tilde{f}}{\partial \mu}(y;x,\mu(\theta))$	$\tilde{\mu}^{\mathrm{A}}_S(\theta)$	–	$\frac{\partial}{\partial \mu'} S^{-1} \sum_{s=1}^S \tilde{g}^s_n(\theta,\mu)$

Table 2: Approximations of limit quanities needed for saddlepoint tests.

5 Finite Sample Performance of Empirical RRY Tests for EMM and II

To illustrate the finite sample performance of the empirical RRY tests based on the optimal EMM and II estimators, we utilize the Monte Carlo set-up from subsection 3.1. Table 3 presents the empirical rejection frequencies of nominal 5% and 1% tests for the null hypothesis $H_0 : \theta = 0.5, 0.6, \ldots, 0.9$ based on $1,000$ Monte Carlo replications of (29) with $\sigma^2_\epsilon = 1$ and known. To compute the tests, the M-estimator score functions (53) and (52) are approximated using the formulas described in Table 2.[5] In contrast to the classical LR-type tests reported in Table 1, the empirical RRY tests show much less size distortion especially for the AR(8) auxiliary model. For the AR(3) auxiliary model, the tests based on the EMM estimators show almost no size distortion whereas the tests based on the II estimators are slightly undersized and the size distortion tends to increase as θ approaches unity. For the AR(8) auxiliary model, the tests based on the EMM estimates display more size distortion than the tests based on the II estimators. The tests based on the EMM estimators are numerically more stable than those based on the II estimators. The numerical instability mainly occurs in the computation of the derivative of the binding function when evaluating the score function (52).

Figure 3 shows finite sample power of the 5% LR-type and empirical RRY tests, based on the EL estimator with AR(3) and AR(8) auxiliary models, of the null hypothesis $H_0 : \theta = 0.5$ against the alternatives $H_1 : \theta = 0.5 + \delta$ ($\delta = \pm 0.1, 0.2, 0.3, 0.4$) using $1,000$ Monte Carlo replications. We present results only for the EL estimator due to numerical problems associated with computing the power for the II estimators. For $n = 50$, the size distortions of the LR-type tests are apparent and it is clear that the empirical RRY test has higher size-adjusted power than the LR-type test for $\theta < 0.5$ and slightly smaller power for $\theta > 0.5$. For $n = 200$, the size distortion of the LR-type tests are much lower and we see that the empirical RRY test has nearly the same size-adjusted power as the LR-type test.

[5] We also computed the ESTs using analytic values for $\mu(\theta)$, M_θ and M_μ. This greatly reduced the computation time of the tests and the results were almost identical.

		$\alpha = 0.05$					$\alpha = 0.01$				
θ	m	EL	EA	IL	IA	IM	EL	EA	IL	IA	IM
		$n = 50$					$n = 50$				
0.5	3	0.05	0.04	0.04	0.04	0.04	0.01	0.01	0.01	0.01	0.01
	8	0.11	0.10	0.08	0.07	0.08	0.03	0.03	0.01	0.01	0.01
0.6	3	0.05	0.05	0.02	0.02	0.02	0.01	0.01	0.00	0.01	0.01
	8	0.11	0.09	0.07	0.08	0.07	0.03	0.02	0.01	0.01	0.01
0.7	3	0.05	0.05	0.02	0.02	0.02	0.01	0.01	0.00	0.00	0.00
	8	0.10	0.10	0.04	0.03	0.03	0.02	0.02	0.00	0.00	0.00
0.8	3	0.05	0.06	0.02	0.02	0.02	0.01	0.01	0.00	0.00	0.00
	8	0.11	0.09	0.03	0.03	0.02	0.02	0.02	0.00	0.00	0.00
0.9	3	0.05	0.07	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.00
	8	0.10	0.11	0.04	0.03	0.03	0.03	0.02	0.01	0.01	0.01
		$n = 200$					$n = 200$				
0.5	3	0.06	0.04	0.05	0.03	0.04	0.02	0.00	0.02	0.01	0.01
	8	0.08	0.07	0.06	0.05	0.05	0.02	0.02	0.01	0.01	0.01
0.6	3	0.05	0.04	0.05	0.03	0.02	0.01	0.00	0.01	0.00	0.00
	8	0.08	0.06	0.06	0.05	0.05	0.01	0.01	0.02	0.01	0.01
0.7	3	0.04	0.04	0.02	0.01	0.01	0.01	0.00	0.01	0.00	0.00
	8	0.08	0.06	0.06	0.05	0.04	0.01	0.01	0.01	0.01	0.01
0.8	3	0.05	0.04	0.02	0.01	0.01	0.01	0.01	0.01	0.00	0.00
	8	0.07	0.06	0.03	0.02	0.02	0.01	0.01	0.00	0.00	0.00
0.9	3	0.05	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.01
	8	0.08	0.06	0.02	0.02	0.02	0.01	0.01	0.00	0.00	0.00

Table 3: Empirical rejection frequencies of empirical saddlepoint coefficient tests.

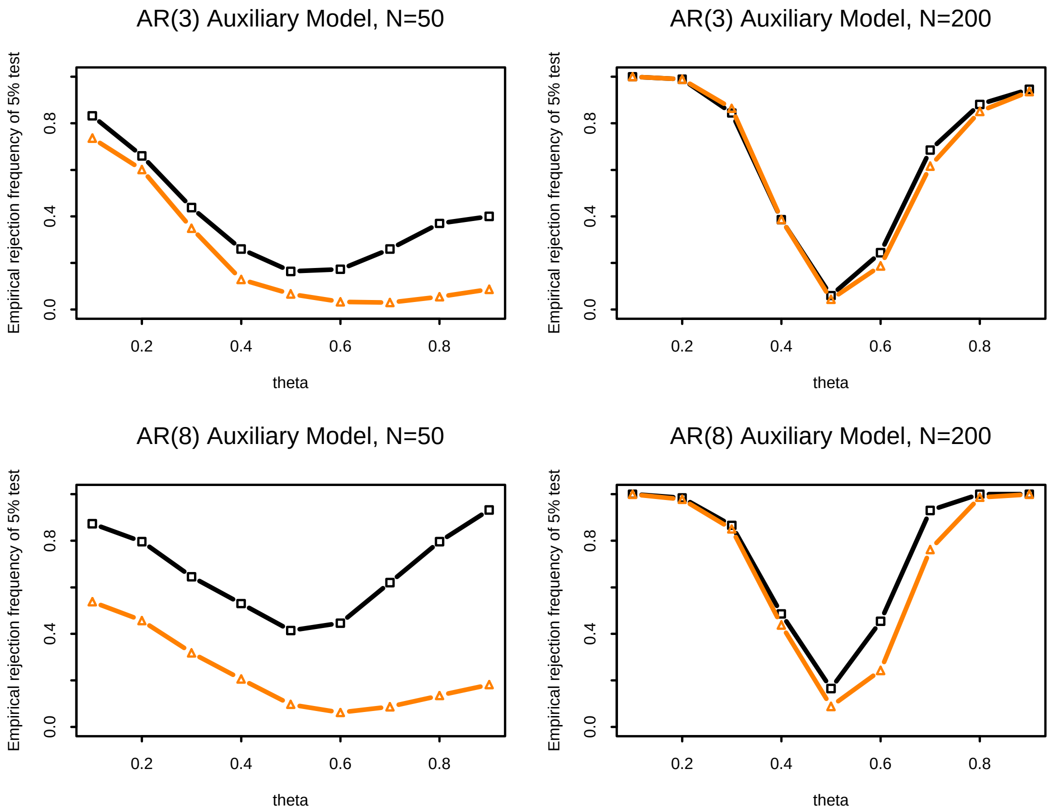

Figure 3: Size-adjusted power of nominal 5% EL LR-type (squares) and EL empirical RRY (triangles) tests.

6 Conclusion

EMM and II are widely used simulation-based estimation techniques. Finite sample comparisons have shown that point estimates based on EMM and II are often similar, but that interval estimates and empirical rejection frequencies of coefficient and overidentification tests based on asymptotic theory can be substantially different with II giving more reliable results than EMM. This has led several researchers (e.g., Ghysels et. al., 2003, Duffee and Stanton, 2007) to advocate the use of II over EMM. Our proposed empirical saddlepoint tests provide improved inference for EMM and II. For the MA(1) model, we show that the large size distortions of the classical LR-type tests for EMM are greatly reduced by the empirical saddlepoint tests without suffering a loss in power. Our results have important practical relevance as they show that accurate finite sample inference can be obtained from EMM estimates.

Our finite sample results were illustrated using a simple MA(1) model. In future research we plan to apply the empirical saddlepoint tests to more complicated models used in finance including stochastic volatility and continuous-time diffusion models estimated by EMM using seminonparametric auxiliary models.

We have shown that the saddlepoint tests can give accurate finite sample inference for EMM and II based on correctly specified models. If the observed data is contaminated in some way or if the model is slightly misspecified, then the EMM and II estimators can give misleading results and the saddlepoint tests may not provide improved inference. However, Genton and Ronchetti (2003) and Ortelli and Trojani (2005) show that II and EMM can be made robust to contamination and certain types of misspecification. It is natural then to consider applying the saddlepoint tests for EMM and II to the robust versions over these estimators. This extension is explored in Czellar and Ronchetti (2007).

References

[1] Bansal, R., Gallant, A.R., Hussey, R., and Tauchen, G. (1993), Computational Aspects of Nonparametric Simulation Estimation, in D.A. Belsley, ed., *Computational Techniques for Econometrics and Economic Analysis*, Boston: Kluwer.

[2] Bansal, R., Gallant, A.R., Hussey, R., and Tauchen, G. (1995), "Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data", *Journal of Econometrics*, **66**, 251-287.

[3] Barndorff-Nielsen, O. E. and Cox, D. R. (1989), *Asymptotic Techniques for Use in Statistics*, London: Chapman & Hall.

[4] Chumacero, R. (1997), "Finite sample properties of the Efficient Method of Moments", *Studies in Nonlinear Dynamics and Econometrics*, **2**, 35–51.

[5] Chumacero, R. (2001), "Estimating ARMA Models Efficiently", *Studies in Nonlinear Dynamics and Econometrics*, **5**, 103–114.

[6] Czellar, V. and Ronchetti, E. (2007), "Second-order Accurate and Robust Indirect Inference", unpublished manuscript, Department of Econometrics, University of Geneva.

[7] Daniels, H.E. (1954), "Saddlepoint Approximations in Statistics", *The Annals of Mathematical Statistics*, **25**, 631–650.

[8] de Luna, X. and Genton, M. G. (2000), "Robust Simulation-Based Estimation", *Statistics & Probability Letters*, **48**, 253–259.

[9] de Luna, X. and Genton, M. G. (2001), "Robust Simulation-Based Estimation of ARMA Models", *Journal of Computational and Graphical Statistics*, **10**, 370–387.

[10] DiCiccio, T.J., and Romano, J.P. (1990), "Nonparametric Confidence Limits by Resampling Methods and Least Favorable Families," *International Statistical Review*, 58, 59-76.

[11] Duffee, G. R. and Stanton, R. H. (2007), "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation", *forthcoming in the Journal of Financial Econometrics.*

[12] Feller, W. (1971), *An Introduction to Probability Theory and its Applications*, vol. 2, Wiley, New York.

[13] Field, C. and Ronchetti, E. (1990), *Small Sample Asymptotics*, Institute of Mathematical Statistics - Monograph Series, Hayward (CA).

[14] Gallant, A. R. and Tauchen, G. (1996), "Which Moments to Match?", *Econometric Theory*, **12**, 657–681.

[15] Gallant, A. R. and Tauchen, G. (2001), "EMM: A Program for Efficient Method of Moments Estimations", *manuscript*, University of North Carolina.

[16] Genton, M. G. and Ronchetti, E. (2003), "Robust Indirect Inference", *Journal of the American Statistical Association*, **98**, 67–76.

[17] Ghysels, E., Khalaf, L. and Vodounou, C. (2003), "Simulation Based Inference in Moving Average Models", *Annales D'Économie et de Statistique*, 69, 85-99.

[18] Gouriéroux, C., Monfort, A. and Renault, E. (1993), "Indirect Inference", *Journal of Applied Econometrics*, **8**, S85–S118.

[19] Gouriéroux, C., Renault, E. and Touzi, N. (2000), "Calibration by Simulation for Small Sample Bias Correction," in Mariano, R., Schuerman, T. and Weeks, M.J. (eds), *Simulation-Based Inference in Econometrics*, Cambridge University Press.

[20] Hampel, F. R. (1974), "The Influence Curve and its Role in Robust Estimation", *Journal of the American Statistical Association*, **69**, 383–393.

[21] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, John Wiley, New York.

[22] Hansen, L.P., Heaton, J.C. and Yaron, A. (1996), "Finite-Sample Properties of Some Alternative GMM Estimators", *Journal of Business and Economic Statistics*, **14**, 262-280.

[23] Lô, S.N. and Ronchetti, E. (2006), "Robust Small Sample Accurate Inference in Moment Condition Models", Working Paper, Department of Econometrics, University of Geneva.

[24] Michaelides, A. and Ng, S. (2000), "Estimating the Rational Expectations Model of Speculative Storage: A Monte Carlo Comparison of Three Simulation Estimators", *Journal of Econometrics*, **96**, 231-266.

[25] Newey, W., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," Chapter 36 in Engle, R.F., McFadden, D.L (eds), *Handbook of Econometrics, Volume IV*, Elsevier Science B.V.

[26] Ortelli, C. and Trojani, F. (2005), "Robust Efficient Method of Moments", *Journal of Econometrics*, **128**, 69–97.

[27] Robinson, J., Ronchetti, E. and Young, G.A. (2003), "Saddlepoint Approximations and Tests Based on Multivariate M-estimators", *The Annals of Statistics*, 4, 1154-1169.

[28] Smith, A. (1993), "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics*, **8**, S63-S84.

[29] Sowell, F. (2007), “The Empirical Saddlepoint Approximation for GMM Estimators,” unpublished manuscript, Department of Economics, Carnegie-Mellon University.

[30] Zhou, H. (2001), “Finite Sample Properties of EMM, GMM, QMLE, and MLE for a Square-Root Diffusion Model”, *Journal of Computational Finance*, **5**, 89-122.

[31] Zivot, E. and Wang, J. (2005). *Modeling Financial Time Series with S-PLUS, Second Edition*, Springer-Verlag, New York.