UHERO
THE ECONOMIC RESEARCH ORGANIZATION
AT THE UNIVERSITY OF HAWAI'I

Indirect Inference Based on the Score

by

Peter Fuleky and Eric Zivot

Working Paper No. 2011-12

April 26, 2011

UNIVERSITY OF HAWAI'I AT MANOA
2424 MAILE WAY, ROOM 540 • HONOLULU, HAWAI'I 96822
WWW.UHERO.HAWAII.EDU

INDIRECT INFERENCE BASED ON THE SCORE

Peter Fuleky[*]and Eric Zivot[†]

April 26, 2011

Abstract

The Efficient Method of Moments (EMM) estimator popularized by Gallant and Tauchen (1996) is an indirect inference estimator based on the simulated auxiliary score evaluated at the sample estimate of the auxiliary parameters. We study an alternative estimator that uses the sample auxiliary score evaluated at the simulated binding function which maps the structural parameters of interest to the auxiliary parameters. We show that the alternative estimator has the same asymptotic properties as the EMM estimator but in finite samples behaves more like the distance-based indirect inference estimator of Gouriéroux, Monfort and Renault (1993).

Keywords: simulation based estimation, indirect inference, efficient method of moments.

JEL Codes: C13, C15, C22.

[*]University of Hawaii Economic Research Organization, 2424 Maile Way, Saunders Hall 540, Honolulu, HI 96822; Email: fuleky@hawaii.edu

[†]Department of Economics, Box 353330, University of Washington, Seattle, WA 98195; Email: ezivot@u.washington.edu. Support from the Robert Richards chair is greatly appreciated.

1 Introduction

Indirect inference estimators take advantage of a simplified auxiliary model that is easier to estimate than a proposed structural model. The estimation consists of two stages. First, an auxiliary statistic is calculated from the observed data. Then an analytical or simulated mapping of the structural parameters to the auxiliary statistic is used to calibrate an estimate of the structural parameters. The simulation-based indirect inference estimators are typically placed into one of two categories: score-based estimators made popular by Gallant and Tauchen (1996), or distance-based estimators proposed by Smith (1993) and refined by Gouriéroux, Monfort and Renault (1993). The simulated score-based estimators have the computational advantage that the auxiliary parameters are estimated from the observed data only once, whereas the distance-based estimators must re-estimate the auxiliary parameters from simulated data as part of the optimization algorithm to estimate the structural parameters. However, many studies have shown (e.g., Michaelides and Ng, 2000; Ghysels, Khalaf and Vodounou, 2003; Duffee and Stanton, 2008) that the computational advantage of the simulated score-based estimators is often offset by poor finite sample properties relative to the distance-based estimators. In this paper we study an alternative score-based estimator that utilizes the sample auxiliary score evaluated with the auxiliary parameters estimated from simulated data. We show that this alternative estimator is asymptotically equivalent to the Gallant and Tauchen (1996) score-based estimator but has finite sample properties that are very close to the distance-based estimators.

The paper is structured as follows. In Section 2, we give an overview of indirect inference estimation. In Section 3, we introduce the alternative score-based estimators and derive their asymptotic properties. In Section 4, we use the framework of Duffee and Stanton (2008) to compare the finite sample properties of various indirect inference estimators for the parameters of a highly persistent AR(1) process via Monte Carlo. Section 5 concludes. Proofs of all results are given in the Appendix.

2 Review of Indirect Inference

Indirect inference (II) techniques were introduced into the econometrics literature by Smith (1993), Gouriéroux, Monfort, and Renault (1993), Bansal, Gallant, Hussey, and Tauchen (1994, 1995) and Gallant and Tauchen (1996), and are surveyed in Gouriéroux and Monfort (1996) and Jiang and Turnbull (2004). There are four components present in simulation-based II: (1) a true structural model whose parameters θ are one's ultimate interest but are difficult to directly estimate; (2) simulated observations from the structural model for a given θ; (3) an auxiliary approximation to the structural model whose parameters μ are easy to estimate; and (4) the binding function, a mapping from μ to θ uniquely connecting the parameters of these two models.

To be more specific, 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.[1] Typical examples are continuous time diffusion models and dynamic stochastic general equilibrium models. Define an auxiliary model $\widetilde{F}_\mu$ in which the parameter $\mu \in \mathbb{R}^r$, with $r \geq p$, is easier to estimate than θ. For ease of exposition, the auxiliary estimator of μ is defined as the quasi-maximum likelihood estimator (QMLE) of the model $\widetilde{F}_\mu$

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

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

where $\tilde{f}(y_t; x_{t-1}, \mu)$ is the log density of y_t for the model $\widetilde{F}_\mu$ conditioned on $x_{t-1} = \{y_i\}_{i=t-m,\ldots,t-1}$, $m \in \mathbb{N}$. We define $\tilde{g}(y_t; x_{t-1}, \mu) = \frac{\partial \tilde{f}(y_t; x_{t-1}, \mu)}{\partial \mu}$ as the $r \times 1$ auxiliary score vector. For more general $\widetilde{Q}_n$, we refer the reader to Gouriéroux and Monfort (1996).

II estimators use the auxiliary model information to obtain estimates of the structural

[1] For simplicity, we do not consider structural models with additional exogenous variables z_t.

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

$$\mu(\theta) = \arg\max_{\mu} E_{F_\theta}[\tilde{f}(y_t; x_{t-1}, \mu)], \tag{3}$$

where $\lim_{n\to\infty} \widetilde{Q}_n\left(\{y_t\}_{t=1,\ldots,n}, \mu\right) = E_{F_\theta}[\tilde{f}(y_t; x_{t-1}, \mu)]$, 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.

II estimators differ in how they use (3) to define an estimating equation. The distance-based II estimator finds θ to minimize the (weighted) distance between $\mu(\theta)$ and $\tilde{\mu}_n$. The score-based II estimator finds θ by solving $E_{F_\theta}[\tilde{g}(y_t; x_{t-1}, \tilde{\mu}_n)] = 0$, the first order condition associated with (3).[2] Typically, the analytic forms of $\mu(\theta)$ and $E_{F_\theta}[\tilde{g}(y_t; x_{t-1}, \mu)]$ are not known and simulation-based techniques are used to compute the two types of II estimators.

For simulation-based II, it is necessary to be able to easily generate simulated observations from F_θ for a given θ. These simulated observations are typically drawn in two ways. First, a long pseudo-data series of size $S \cdot n$ is simulated giving

$$\{y_t(\theta)\}_{t=1,\ldots,Sn}, \;\; S \geq 1. \tag{4}$$

Alternatively, S pseudo-data series of size n are simulated giving

$$\{y_t^s(\theta)\}_{t=1,\ldots,n}, \;\; s = 1, \ldots, S, \;\; S \geq 1. \tag{5}$$

Using the simulated observations (4) or (5), the distance-based II estimators are minimum

[2]Gallant and Tauchen (1996a) call the score-based II estimator the efficient method of moments (EMM) estimator. Efficiency in the context of EMM refers to the efficiency of the auxiliary model in approximating the structural model, and Gallant and Tauchen (1996, 2004) advocated the use of a particular seminonparametric auxiliary model to achieve such efficiency.

distance estimators defined as

$$\hat{\theta}_S^{\mathrm{Dj}}(\widetilde{\Omega}_n) = \arg\min_{\theta} \ \mathrm{J}_S^{\mathrm{Dj}}(\theta, \widetilde{\Omega}_n) = \arg\min_{\theta} \ \left(\tilde{\mu}_n - \tilde{\mu}_S^{\mathrm{j}}(\theta)\right)' \widetilde{\Omega}_n \left(\tilde{\mu}_n - \tilde{\mu}_S^{\mathrm{j}}(\theta)\right) \ , \ \mathrm{j} = \mathrm{A}, \mathrm{L}, \mathrm{M}, \quad (6)$$

where $\widetilde{\Omega}_n$ is a positive definite and symmetric weight matrix which may depend on the data through the auxiliary model, and the simulated binding functions are given by

$$\tilde{\mu}_S^{\mathrm{A}}(\theta) = \arg\max_{\mu} \ S^{-1} \sum_{s=1}^{S} \widetilde{Q}_n \left(\{y_t^s(\theta)\}_{t=1,\ldots,n}, \mu\right), \quad (7)$$

$$\tilde{\mu}_S^{\mathrm{L}}(\theta) = \operatorname{argmax}_{\mu} \ \widetilde{Q}_{Sn} \left(\{y_t(\theta)\}_{t=1,\ldots,Sn}, \mu\right), \quad (8)$$

$$\tilde{\mu}_S^{\mathrm{M}}(\theta) = S^{-1} \sum_{s=1}^{S} \arg\max_{\mu} \ \widetilde{Q}_n \left(\{y_t^s(\theta)\}_{t=1,\ldots,n}, \mu\right). \quad (9)$$

The superscripts A, L, and M indicate how the binding function is computed from the simulated data: "A" denotes maximizing an aggregation of S objective functions using (5); "L" denotes use of long simulations (4) in the objective function; "M" denotes use of the mean of S estimated binding functions based on (5). The M-type estimator is more computationally intensive than the A and L-type estimators, but exhibits superior finite sample properties in certain circumstances, as shown by Gouriéroux, Renault, and Touzi (2000).

Using (4) or (5), the score-based II estimators are one-step GMM estimators defined as

$$\hat{\theta}_S^{\mathrm{Sj1}}(\widetilde{\Sigma}_n) = \arg\min_{\theta} \ J_S^{\mathrm{Sj1}}(\theta) = \arg\min_{\theta} \tilde{g}_S^{\mathrm{j}}(\theta, \tilde{\mu}_n)' \widetilde{\Sigma}_n \tilde{g}_S^{\mathrm{j}}(\theta, \tilde{\mu}_n), \ \mathrm{j} = \mathrm{A}, \mathrm{L}, \quad (10)$$

where $\widetilde{\Sigma}_n$ is a positive definite (pd) and symmetric weight matrix which may depend on the data through the auxiliary model, and the simulated scores are given by

$$\tilde{g}_S^{\mathrm{A}}(\theta, \tilde{\mu}_n) = S^{-1} \sum_{s=1}^{S} \frac{1}{n-m} \sum_{t=m+1}^{n} \tilde{g}(y_t^s(\theta); x_{t-1}^s(\theta), \tilde{\mu}_n), \tag{11}$$

$$\tilde{g}_S^{\mathrm{L}}(\theta, \tilde{\mu}_n) = \frac{1}{Sn-m} \sum_{t=m+1}^{Sn} \tilde{g}(y_t(\theta); x_{t-1}(\theta), \tilde{\mu}_n). \tag{12}$$

Because (10) fixes the binding function at the sample estimate $\tilde{\mu}_n$ no M-type estimator is available.

Under regularity conditions described in Gouriéroux and Monfort (1996), the distance-based estimators (6) and score-based estimators (10) are consistent for θ_0 (true parameter vector) and asymptotically normally distributed. The limiting weight matrices that minimize the asymptotic variances of these estimators are $\widetilde{\Omega}^* = M_\mu \widetilde{\mathcal{I}}^{-1} M_\mu$ and $\widetilde{\Sigma}^* = \widetilde{\mathcal{I}}^{-1}$, where $\widetilde{\mathcal{I}} = \lim_{n\to\infty} \mathrm{var}_{F_\theta}(\sqrt{n}\tilde{g}_n(y_n, \mu(\theta_0)))$, $M_\mu = E_{F_\theta}[\widetilde{H}(y_t; x_{t-1}, \mu(\theta_0))]$, $\tilde{g}_n(y_n, \mu(\theta)) = \frac{1}{n-m}\sum_{t=m+1}^{n} \tilde{g}(y_t; x_{t-1}, \mu(\theta))$ and $\widetilde{H}(y_t; x_{t-1}, \mu) = \frac{\partial^2 \tilde{f}(y_t; x_{t-1}, \mu)}{\partial\mu\partial\mu'}$. Using consistent estimates of these optimal weight matrices, the distance-based and score-based estimators are asymptotically equivalent with asymptotic variance matrix given by[3]

$$V_S^* = \left(1 + \frac{1}{S}\right)\left(M_\theta' \widetilde{\mathcal{I}}^{-1} M_\theta\right)^{-1} = \left(1 + \frac{1}{S}\right)\left(\frac{\partial\mu(\theta_0)'}{\partial\theta} M_\mu' \widetilde{\mathcal{I}}^{-1} M_\mu \frac{\partial\mu(\theta_0)}{\partial\theta'}\right)^{-1}, \tag{13}$$

where

$$M_\theta = \left\{\frac{\partial}{\partial\theta'} E_{F_\theta}\left[\tilde{g}(y_t; x_{t-1}, \mu)\right]\right\}\Big|_{\mu=\mu(\theta_0)}.$$

3 Alternative Score-Based II Estimator

Gouriéroux and Monfort (1996, pg. 71) mentioned two alternative II estimators that they claimed are less efficient than the optimal estimators described in the previous section, and referred the reader to Smith (1993) for details. The first one is the simulated quasi-maximum

[3]The equality of the left hand side and right hand side of (13) follows from the result $\frac{\partial\mu(\theta_0)}{\partial\theta'} = M_\mu^{-1} M_\theta$.

likelihood (SQML) estimator defined as

$$\hat{\theta}_S^{\mathrm{SQMLj}} = \arg\max_{\theta} \widetilde{Q}_n \left(\{y_t\}_{t=1,\ldots,n}, \tilde{\mu}_S^{\mathrm{j}}(\theta) \right), \mathrm{j} = \mathrm{A}, \mathrm{L}, \mathrm{M}. \tag{14}$$

Smith (1993) showed that (14) is consistent and asymptotically normal with asymptotic variance matrix given by

$$V_S^{SQML} = \left(1 + \frac{1}{S}\right) \left[\frac{\partial \mu(\theta_0)'}{\partial \theta} M_\mu \frac{\partial \mu(\theta_0)}{\partial \theta'} \right]^{-1} \frac{\partial \mu(\theta_0)'}{\partial \theta} \widetilde{\mathcal{I}} \frac{\partial \mu(\theta_0)}{\partial \theta'} \left[\frac{\partial \mu(\theta_0)'}{\partial \theta} M_\mu \frac{\partial \mu(\theta_0)}{\partial \theta'} \right]^{-1}, \tag{15}$$

which he showed is strictly greater than (in a matrix sense) the asymptotic variance (13) of the efficient II estimators. As noted by Gouriéroux, Monfort, and Renault (1993), using the result $\frac{\partial \mu(\theta_0)}{\partial \theta'} = M_\mu^{-1} M_\theta$, the asymptotic variance of the SQML estimator is efficient only when $\widetilde{\mathcal{I}} = -M_\mu$.

The second alternative II estimator mentioned by Gouriéroux and Monfort (1996, pg. 71), which we call the S2 estimator, is an alternative score-based estimator of the form

$$\hat{\theta}_S^{\mathrm{Sj2}}(\widetilde{\Sigma}_n) = \arg\min_{\theta} J^{\mathrm{Sj2}}(\theta, \widetilde{\Sigma}_n) = \arg\min_{\theta} \; \tilde{g}_n^{\mathrm{j}}(\theta)' \widetilde{\Sigma}_n \tilde{g}_n^{\mathrm{j}}(\theta), \tag{16}$$

where

$$\tilde{g}_n^{\mathrm{j}}(\theta) = \frac{1}{n-m} \sum_{t=m+1}^{n} \tilde{g}(y_t; x_{t-1}, \tilde{\mu}_S^{\mathrm{j}}(\theta)), \mathrm{j} = \mathrm{A}, \mathrm{L}, \mathrm{M}. \tag{17}$$

The S2 estimator was not explicitly considered in Smith (1993). In contrast to the simulated scores (11) and (12), the score in (17) is evaluated with the observed data and the simulated binding function. The following Proposition gives the asymptotic properties of (16).

Proposition 1 *Under the regularity conditions in Gouriéroux and Monfort (1996), the score-based II estimators* $\hat{\theta}_S^{Sj2}(\widetilde{\Sigma}_n)$ *(j=A,L,M) defined in (16) are consistent and asymptoti-*

cally normal, when S is fixed and $n \to \infty$:

$$\sqrt{n}(\hat{\theta}_S^{Sj2}(\widetilde{\Sigma}_n) - \theta_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right)[M_\theta'\Sigma M_\theta]^{-1}\left[M_\theta'\Sigma\widetilde{\mathcal{I}}\Sigma M_\theta\right][M_\theta'\Sigma M_\theta]^{-1}\right) \ . \quad (18)$$

The proof is given in Appendix A. We make the following remarks:

1. The asymptotic variance of $\hat{\theta}_S^{\mathrm{Sj2}}(\widetilde{\Sigma}_n)$ in (18) is equivalent to the asymptotic variance of Gallant and Tauchen's score-based estimator $\hat{\theta}_S^{\mathrm{Sj1}}(\widetilde{\Sigma}_n)$, and is equivalent to (13) when $\widetilde{\Sigma}_n$ is a consistent estimator of $\widetilde{\mathcal{I}}^{-1}$. Contrary to the claim in Gouriéroux and Monfort (1996), the alternative score-based II estimator is not less efficient than the optimal II estimators.[4]

2. To see the relationship between the two score-based estimators, (10) and (16), note that the first order conditions (FOCs) of the optimization problem (3) defining $\mu(\theta)$ are
$$0 = E_{F_\theta}\left[\frac{\partial \tilde{f}(y_t; x_{t-1}, \mu)}{\partial \mu}\right]\Bigg|_{\mu=\mu(\theta)} \equiv \tilde{g}_E(y_t(\theta), \mu(\theta)) \equiv \tilde{g}_E(\theta, \mu(\theta)) \ . \quad (19)$$
This expression depends on θ through $y_t(\theta)$ and $\mu(\theta)$, and both score-based II estimators make use of this population moment condition. The S1 and S2 estimators differ in how sample information and simulations are used. For the S1 estimator, $\mu(\theta)$ is estimated from the sample and simulated values of $y_t(\theta)$ are used to approximate $E_{F_\theta}[\cdot]$. For the S2 estimator, $y_t(\theta)$ is obtained from the sample and simulated values of $\mu(\theta)$ are used for calibration to minimize the objective function. Because the S2 estimator (16) evaluates the sample auxiliary score with a simulated binding function, it is more like the distance-based II estimator (6).

3. To see why the S1 and S2 estimators are asymptotically equivalent and efficient, and the SQML estimator is generally inefficient, consider the first order conditions (FOCs)

[4] "Efficiency" is used to mean the optimal use of the information provided by the binding function for *a given auxiliary model.*

defining these estimators. From (10), the FOCs for the optimal S1 estimator are

$$0 = \frac{\partial \tilde{g}_S(\hat{\theta}_S, \tilde{\mu}_n)'}{\partial \theta} \widetilde{\mathcal{I}}_n^{-1} \tilde{g}_n(y_t; x_{t-1}, \tilde{\mu}_n) \ , \tag{20}$$

and, from (16), the FOCs for the optimal S2 estimator are

$$0 = \frac{\partial \tilde{\mu}_S(\hat{\theta}_S)'}{\partial \theta} \frac{\partial \tilde{g}_n(y_t; x_{t-1}, \tilde{\mu}_S(\hat{\theta}_S))'}{\partial \mu} \widetilde{\mathcal{I}}_n^{-1} \tilde{g}_n(y_t; x_{t-1}, \tilde{\mu}_S(\hat{\theta}_S)) \ , \tag{21}$$

where $\widetilde{\mathcal{I}}_n$ is a consistent estimate of $\widetilde{\mathcal{I}}$. When n and S are large enough, $\tilde{\mu}_S(\hat{\theta}_S) \approx \tilde{\mu}_n \approx \mu(\theta_0)$, $\frac{\partial \tilde{g}_S(\hat{\theta}_S, \tilde{\mu}_n)}{\partial \theta'} \approx M_\theta$, $\frac{\partial \tilde{g}_n(y_t; x_{t-1}, \tilde{\mu}_S(\hat{\theta}_S))}{\partial \mu'} \approx M_\mu$, and $\widetilde{\mathcal{I}}_n \approx \widetilde{\mathcal{I}}$. It follows that (20) and (21) can be re-expressed as

$$0 = M_\theta' \widetilde{\mathcal{I}}^{-1} \tilde{g}_n(y_t; x_{t-1}, \mu(\theta_0)) + o_p(1) \ , \tag{22}$$

and

$$0 = \frac{\partial \mu(\theta_0)'}{\partial \theta} M_\mu' \widetilde{\mathcal{I}}^{-1} \tilde{g}_n(y_t; x_{t-1}, \mu(\theta_0)) + o_p(1) \ . \tag{23}$$

Using the result $M_\theta = M_\mu \frac{\partial \mu(\theta_0)}{\partial \theta'}$ it follows that the FOCs for the S1 and S2 estimators pick out the optimal linear combinations of the overidentified auxiliary score that produces the efficient II estimator. In contrast, from (14) the FOCs for the SQML are

$$0 = \frac{\partial \mu_S(\theta_0)'}{\partial \theta} \tilde{g}_n(y_t; x_{t-1}, \mu(\theta_0)) + o_p(1) \ . \tag{24}$$

Here, the muliplication of the auxiliary score (17) by $\frac{\partial \mu_S(\theta_0)'}{\partial \theta}$ does not pick out the optimal linear combinations of the auxiliary score unless $\widetilde{\mathcal{I}} = -M_\mu$.[5]

[5]We are thankful to an anonymous referee for pointing out this intuitive explanation for the inefficiency of the SQML estimator.

4 Finite Sample Comparison of II Estimators

In this section, we use Monte Carlo methods to compare the finite sample performance of the alternative score-based estimator (16) to the traditional II estimators (6) and (10) using a simple first order autoregressive, AR(1), process. Our Monte Carlo design is motivated by Duffee and Stanton (2008) (hereafter, DS). They compared the finite sample properties of the S1 and D estimators using highly persistent AR(1) models calibrated to interest rate data and found that S1 is severely biased, has wide confidence intervals, and performs poorly in coefficient and overidentification tests. The analytically tractable AR(1) process also gives us the opportunity to compute non-simulation-based analogues of the simulation-based estimators, and to directly compare the performance of the II estimators to the benchmark conditional maximum likelihood estimator of the the structural parameter.[6]

4.1 Model Setup

Assume that the true data generating process is an AR(1) process of the form

$$F_\theta : y_t = \theta_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iid\ N(0,1),\ |\theta_1| < 1\ , \tag{25}$$

with θ_1 values close to unity, which is motivated by the observed highly persistent behavior of interest rate data. In accordance with DS, we calibrate our model to mimic interest rate processes sampled at the weekly frequency with three different half-lives for shocks: $\theta_1 = 0.8522$ (one month half-life), $\theta_1 = 0.9868$ (one year half-life), and 0.9978 (six year half-life). We analyze samples of size $T = 200, 1000, 2000$ and 10000. As in DS, we use the

[6] A similar analysis based on a continuous-time Ornstein-Uhlenbeck process is given in Fuleky (2009).

overidentified auxiliary model

$$\widetilde{F}_\mu : y_t = \mu_0 + \mu_1 y_{t-1} + \xi_t \ , \quad \xi_t \sim iidN(0, \mu_2) \ , \tag{26}$$

$$= \beta' x_{t-1} + \xi_t \ , \ \text{ where } \ x_{t-1} = (1, y_{t-1})' \ \text{ and } \ \beta = (\mu_0, \mu_1)' \ . \tag{27}$$

The auxiliary estimator, $\tilde{\mu}_n$, found by maximizing (2) is the least squares estimator. Given (25) and (26), the binding function has the simple form

$$\mu(\theta_1) = \operatorname*{plim}_{n \to \infty} \tilde{\mu} = (0, \theta_1, 1)' \ . \tag{28}$$

Because the binding function and the expected score vector $E_{F_{\theta_1}}[\tilde{g}(y_t; x_{t-1}, \mu)]$ have closed form expressions, non-simulation-based versions of the distance-based and score-based II estimators are available.[7] We denote these estimators DN, SN1 and SN2, respectively.

For the simulations (4) and (5), we set $S = 20$ so that the simulation-based estimators have a 95% asymptotic efficiency relative to the non-simulation-based estimators (see (13)), and use the same random number seed for all values of θ_1 during the optimizations. When simulating from (25), the stability constraint $|\theta_1| < 1$ is imposed and simulations are started from $y_0 = 0$, the long run mean of the process.

4.2 Objective Functions and Confidence Intervals

Figure 1 illustrates the LR-type statistics for testing $H_0 : \theta_1 = \theta_1^0$ as functions of θ_1^0 for the II estimators based on a single representative sample of size $n = 1000$ generated from (25) with $\theta_1 = \{0.8522, 0.9868, 0.9978\}$. The 95% confidence intervals for θ_1 are obtained by inverting the LR statistics, which have asymptotic chi-square distributions with one degree of freedom. Table 1 summarizes the point estimates and confidence intervals for each of the estimators.

[7] Expressions for the conditional log-density, score, Hessian, etc., are summarized in Appendix B and DS.

The distance-based (D) and the alternative score-based (S2) estimates and LR-type statistics are very similar. The D and S2 estimates are slightly greater than the Gallant-Tauchen score-based (S1) estimates, and the D and S2 LR-type statistics are more symmetric than the corresponding S1 statistics. The M-type LR-type statistics are shifted toward unity reflecting the different finite sample properties of the M-type estimators in comparison to the N, L and A-type ones. As noted by DS, the shape of LR^{S1} is highly asymmetric due to the scaling of some sample moments by the population variance. In contrast, the shapes of the LR functions for the S2 and D estimators are almost identical and are roughly symmetric in θ_1. This occurs because they are scaled by the variance of the observed sample which is constant for any θ_1.

4.3 Computational Issues

The S2 estimator can be considered a hybrid estimator consisting of two steps. In the first step the simulation-based binding function $\tilde{\mu}_S(\theta_1)$ is calculated. In distance-based II this simulated binding function is directly compared to the auxiliary estimate $\tilde{\mu}$. In the S2 estimator the mean score evaluated with $\tilde{\mu}_S(\theta_1)$ is compared to the mean score evaluated with $\tilde{\mu}$, where the latter is equal to zero by construction. Because the score function is evaluated with the observed data, a fixed input, all the variability of the S2 objective function can be attributed to the simulated binding function $\tilde{\mu}_S(\theta_1)$, just like in the case of the D estimators' objective function. Therefore the objective functions of the simulation-based S2 and D estimators will also look similar.[8]

Gallant and Tauchen (2002) criticize distance-based II for its computational inefficiency, because it potentially involves two nested optimizations: the estimator of the simulated binding function is embedded within the D estimator. This, they argued, may lead to numerical instability if the auxiliary estimator does not have a closed form analytical expression but instead relies on an optimizer. The inner (binding function) optimization, which is computed

[8]The shape of the objective function is equivalent to the shape of the LR statistic except for a level shift.

within a tolerance, will cause some jitter, and render the outer (structural) optimization problem non-smooth. Because the S2 estimator also uses the simulated binding function, similar issues have to be weighed when the auxiliary model for S2 is chosen. However, if a simple auxiliary model is chosen such that the auxiliary estimator has a closed form analytical solution, the speed and stability of the S2 and D estimator becomes much improved.

The current Monte Carlo study is a case in point: Table 2 indicates that the average computation time associated with the S2 and D estimators that use a simulated binding function (including the M-type!) is actually lower than that of the simulation based S1 estimators.[9] The binding function in the S2 and D estimator does not involve a nested optimization, only the analytical expression for the least squares estimator is evaluated. However, in each iteration pseudo-series are simulated according to (4) or (5). The time required to generate a simulated sequence dominates the time required to evaluate the sequence in an analytical expression. Consequently, the computational efficiency of the estimators is heavily influenced by the algorithm's speed of convergence, and the number of times simulations have to be generated. Because the S1 objective functions have irregular shapes, a higher number of iterations is required for convergence of these estimators, and this explains their relative computational inefficiency in the current setup.

4.4 Bias and RMSE

Table 3 summarizes the bias and dispersion of the Conditional MLE (CMLE) and II estimators of θ_1. The CMLE estimator is based on the least squares estimation of the structural model (25) that does not include a constant. The results are based on 1000 Monte Carlo simulations. Gouriéroux and Monfort (1996, pg. 66) note that the score-based and distance-based II estimates should be very close in a just identified setting. However, Table 3 indicates that the distributions of these estimators in an over-identified setting can be very different.

[9] Optimization was performed by the `R` function `optimize`, which uses a combination of golden section search and successive parabolic interpolation.

The S1 estimators are extremely biased in comparison to CMLE, a confirmation of DS's finding. In contrast, the bias and dispersion of the corresponding S2 and D-type estimators (N, L and A-type) are comparable to those of CMLE.

In general, the CMLE and II estimators are affected by a finite sample bias (Hurwitz, 1950; Mariott and Pope, 1954; Kendall, 1954) due to the highly persistent nature of the adopted parameterization of the AR(1) process in (25). The M-type estimator has been shown to correct this finite sample bias in a just-identified setting (Gouriéroux, Renault, and Touzi, 2000; Gouriéroux, Phillips, and Yu, 2008), but the results of Table 3 show that this is not the case in an over-identified setting. While the N, A and L-type estimators show a negative bias, the M-type estimator shows a positive finite sample bias.[10]

4.5 Test Statistics

Tables 4 and 5 show the empirical rejection rates of nominal 5% over-identification tests and LR-type coefficient tests[11] of $\theta_1 = \theta_1^0$, respectively, based on 1000 Monte Carlo simulations. Our results are similar to those of DS for the S1 and D estimators. In addition, the newly considered S2 estimator shows an improved testing performance over the S1 estimator.

The LR and over-identification test statistics for Sj1 (j = N, L, A) show substantial size distortion in smaller samples of the highly persistent AR(1) process. The high rejection rate of these tests is caused by the finite sample bias of the S1 estimators combined with the asymmetry of the S1 objective functions. For the Sj2 and Dj (j = N,L,A,M) estimators, the rejection rates are approximately equal and closer to the nominal level. Here, the finite sample behavior of the test statistics depends on how the binding function is approximated: the LR statistics are approximately correctly sized for the N, L and A-type estimators. Fuleky (2009) shows that the higher rejection rates of the LR-type tests based on the M-type esti-

[10]Fuleky (2009) shows that in a just identified setting where a constant and the variance of ε_t in (25) are assumed to be unknown, and are being estimated along with θ_1, the SM2 and DM estimators exhibit a 90% reduction in mean bias compared to CMLE, and these results confirm the finite sample bias correcting properties of the M-type estimators in just identified models.

[11]See Gouriéroux and Monfort (1996) for the expressions of these statistics.

mators is caused by the over-identification restrictions in conjunction with the nonlinearity of the binding function in small samples.[12]

5 Conclusion

In this paper we study the asymptotic and finite sample properties of an alternative score-based II estimator that uses the sample auxiliary score evaluated at the simulated binding function. We show that this estimator is asymptotically equivalent to Gallant and Tauchen's simulated score estimator, but in finite samples behaves much more like the distance-based II estimators. For estimating the autoregressive parameter of a highly persistent AR(1) process, we show that the alternative score-based estimator does not exhibit the poor finite sample properties of the simulated score estimator, and that the former is more computationally efficient than the latter. Our results counter some of the criticisms of the score-based II estimators raised by Duffee and Stanton (2008).

[12]The M-type estimators have improved inference properties in just identified models. In a just identified setting where the θ_0 and θ_2 parameters are assumed to be unknown, and are being estimated along with θ_1, the empirical size of the LR-type test for testing $H_0 : \theta_1 = \theta_1^0$ is close to the nominal size and approximately equal for the SN2, SM2, DN and DM estimators.

References

Bansal, R. and Gallant, A.R. and Hussey, R. and Tauchen, G. (1994). Computational aspects of nonparametric simulation estimation. *Computational techniques for econometrics and economic analysis.*

Bansal, R. and Gallant, A.R. and Hussey, R. and Tauchen, G. (1995). Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics.* 66(1-2):251-287.

Duffee, G. and Stanton, R. (2008). Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation. *Journal of Financial Econometrics*, 6(1):108-150.

Fuleky, P. (2009). Simulation-based estimation of continuous-time models. Phd Thesis, Department of Economics, University of Washington.

Gallant, A. and Tauchen, G. (1996). Which Moments to Match? *Econometric Theory*, 12(4):657-81.

Gallant, A. and Tauchen, G. (1996a). Specification Analysis of Continuous Time Models in Finance. *Modelling Stock Market Volatility: Bridging the Gap to Continuous Time*, Academic Press.

Gallant, A. and Tauchen, G. (2002). Simulated Score Methods and Indirect Inference for Continuous Time Models. *Handbook of Financial Econometrics*, North Holland, Amsterdam.

Gallant, A. and Tauchen, G. (2004). EMM: A Program for Effcient Method ofMoments Estimation. Unpublished manuscript, Department of Economics, Duke University.

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

Gouriéroux, C. and Monfort, A. (1996). Simulation-Based Econometric Methods. Oxford University Press, USA.

Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, 8:S85-S118.

Gouriéroux, C., Phillips, P., and Yu, J. (2008). Indirect Inference for Dynamic Panel Models. *Journal of Econometrics*.

Gouriéroux, C., Renault, E., and Touzi, N. (2000). Calibration by simulation for small sample bias correction. *Simulation-Based Inference in Econometrics, Methods and Applications*. Cambridge University Press, Cambridge.

Hurwicz, L. (1950). Least Squares Bias in Time Series. *Statistical Inference in Dynamic Economic Models*.

Jiang, W. and Turnbull, B. (2004). The Indirect Method: Inference Based on Intermediate Statistics|A Synthesis and Examples. *Statistical Science*, 19(2):239-263.

Kendall, MG. (1954). Note on Bias in the Estimation of Autocorrelation. *Biometrika*.

Marriott, FHC. and Pope, JA. (1954). Least Squares Bias in Time Series. *Biometrika*.

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.

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

6 Appendix A: Proof of Proposition 1

The regularity conditions from Gouriéroux and Monfort (1996, Appendix 4A) are:

(A1) $\widetilde{Q}_n\left(\{y_t\}_{t=1,\ldots,n},\mu\right)=\tilde{f}_n\left(y_n,\mu\right)=\frac{1}{n-m}\sum_{t=m+1}^{n}\tilde{f}(y_t;x_{t-1},\mu)\xrightarrow{p}\tilde{f}_E(\theta,\mu)=E_{F_\theta}[\tilde{f}(y_t;x_{t-1},\mu)]$ uniformly in (θ,μ) as $n\to\infty$.

(A2) $\tilde{f}_E(\theta,\mu)$ has a unique maximum with respect to $\mu:\mu(\theta)=\arg\max_\mu\tilde{f}_E(\theta,\mu)$.

(A3) $\tilde{f}_n\left(y_n,\mu\right)$ and $\tilde{f}_E(\theta,\mu)$ are differentiable with respect to μ, and $\tilde{g}_E(\theta,\mu)=\frac{\partial\tilde{f}_E(\theta,\mu)}{\partial\mu}$ $=\lim_{n\to\infty}\frac{\partial\tilde{f}_n(y_n,\mu)}{\partial\mu}$.

(A4) The only solution to the asymptotic first order conditions is $\mu(\theta):\tilde{g}_E(\theta,\mu)=0\Rightarrow\mu=\mu(\theta)$.

(A5) The equation $\mu=\mu(\theta)$ admits a unique solution in θ.

(A6) $p\lim_{n\to\infty}\frac{\partial^2\tilde{f}_n(y_n,\mu(\theta))}{\partial\mu\partial\mu'}=E_{F_\theta}[\widetilde{H}(y_t;x_{t-1},\mu(\theta_0))]=M_\mu$

(A7) $\sqrt{n}\tilde{g}_n(y_n,\mu(\theta_0))=\sqrt{n}\frac{\partial\tilde{f}_n(y_n,\mu(\theta_0))}{\partial\mu}\xrightarrow{d}N(0,\widetilde{\mathcal{I}})$ as $n\to\infty$.

For ease of exposition, we only give the proof for $\hat{\theta}_S^{\mathrm{SL2}}(\widetilde{\Sigma}_n)=\hat{\theta}_S^{\mathrm{L}}$ which follows closely the proof from Gouriéroux and Monfort (1996, Appendix 4A). The results for the other estimators are similar. For consistency, first note that for fixed S and as $n\to\infty$

$$\tilde{g}_n(y_n,\mu(\theta))\xrightarrow{p}\tilde{g}_E(\theta_0,\mu(\theta)),$$
$$\tilde{\mu}_S^{\mathrm{L}}(\theta)=\arg\max_\mu\tilde{f}_{Sn}\left(y_{Sn}^{p}(\theta),\mu\right)\xrightarrow{p}\arg\max_\mu S\tilde{f}_E(\theta,\mu)=\mu(\theta).$$

Then $\hat{\theta}_S^{\mathrm{L}}\xrightarrow{p}\arg\min_\theta\tilde{g}_E(\theta_0,\mu(\theta))'\Sigma\tilde{g}_E(\theta_0,\mu(\theta))$ which, by A4, is uniquely minimized at $\theta=\theta_0$. Hence, $\hat{\theta}_S^{\mathrm{L}}\xrightarrow{p}\theta_0$.

For asymptotic normality, the first order condition of the optimization problem in (16) is

$$\frac{\partial\tilde{g}_n(y_n,\tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))'}{\partial\theta}\widetilde{\Sigma}_n\tilde{g}_n(y_n,\tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))=0.\tag{29}$$

Taking a mean value expansion (MVE) of $\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))$ around θ_0 and plugging it into (29) gives

$$\frac{\partial \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))'}{\partial \theta} \widetilde{\Sigma}_n \left[\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) + \frac{\partial \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\bar{\theta}))}{\partial \mu'} \frac{\partial \tilde{\mu}_S^{\mathrm{L}}(\bar{\theta})}{\partial \theta'} (\hat{\theta}_S^{\mathrm{L}} - \theta_0) \right] = 0 \ , \tag{30}$$

where $\bar{\theta}$ represents the vector of intermediate values. Using the results

$$\frac{\partial \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))'}{\partial \theta} = \frac{\partial \tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))'}{\partial \theta} \frac{\partial \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))'}{\partial \mu} \xrightarrow{p} \frac{\partial \mu(\theta_0)'}{\partial \theta} \frac{\partial \tilde{g}_E(\theta_0, \mu(\theta_0))'}{\partial \mu} = M_\theta' \ ,$$
$$\frac{\partial \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\bar{\theta}))}{\partial \mu'} \frac{\partial \tilde{\mu}_S^{\mathrm{L}}(\bar{\theta}))}{\partial \theta'} \xrightarrow{p} \frac{\partial \tilde{g}_E(\theta_0, \mu(\theta_0))}{\partial \mu'} \frac{\partial \mu(\theta_0)}{\partial \theta'} = M_\theta,$$

and re-arranging (30) then gives

$$\sqrt{n}(\hat{\theta}_S^{\mathrm{L}} - \theta_0) = - \left[M_\theta' \Sigma M_\theta \right]^{-1} M_\theta' \Sigma \sqrt{n} \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) + o_p(1). \tag{31}$$

Next, use a MVE of $\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0))$ around $\tilde{\mu}$ to give

$$\begin{aligned} \sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) &= \sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}) + \frac{\partial \tilde{g}_n(y_n, \bar{\mu})}{\partial \mu'} \sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \tilde{\mu}) \\ &= \sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}) + M_\mu \sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \tilde{\mu}) + o_p(1), \end{aligned} \tag{32}$$

and another MVE of $\tilde{g}_n(y_n, \tilde{\mu}) = 0$ around $\mu(\theta_0)$ to give

$$\sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}) = \sqrt{n}\tilde{g}_n(y_n, \mu(\theta_0)) + \frac{\partial \tilde{g}_n(y_n, \bar{\bar{\mu}})}{\partial \mu'} \sqrt{n}(\tilde{\mu} - \mu(\theta_0)) = 0,$$

so that

$$\sqrt{n}(\tilde{\mu} - \mu(\theta_0)) = -M_\mu^{-1} \sqrt{n} \tilde{g}_n(y_n, \mu(\theta_0)) + o_p(1). \tag{33}$$

In addition, use a MVE of the simulated score $\tilde{g}_{Sn}(y_{Sn}(\theta_0), \tilde{\mu}_S^{\mathrm{L}}(\theta_0))$ around $\mu(\theta_0)$

$$\sqrt{n}\tilde{g}_{Sn}(y_{Sn}(\theta_0), \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) = \sqrt{n}\tilde{g}_{Sn}(y_{Sn}(\theta_0), \mu(\theta_0)) + \frac{\partial \tilde{g}_{Sn}(y_{Sn}(\theta_0), \bar{\bar{\mu}})}{\partial \mu'}\sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \mu(\theta_0)) = 0,$$

so that

$$\begin{aligned}\sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \mu(\theta_0)) &= -\left[\frac{\partial \tilde{g}_{Sn}(y_{Sn}(\theta_0), \bar{\bar{\mu}})}{\partial \mu'}\right]^{-1}\sqrt{n}\tilde{g}_{Sn}(y_{Sn}(\theta_0), \mu(\theta_0)) \\ &= -S^{-1}M_\mu^{-1}\sqrt{n}\sum_{s=1}^{S}\tilde{g}_n(y_n^s(\theta_0), \mu(\theta_0)) + o_p(1),\end{aligned} \tag{34}$$

since $\tilde{g}_{Sn}(y_{Sn}(\theta_0), \mu(\theta_0)) = \sum_{s=1}^{S}\tilde{g}_n(y_n^s(\theta_0), \mu(\theta_0))$ and so

$$\frac{\partial \tilde{g}_{Sn}(y_{Sn}(\theta_0), \bar{\bar{\mu}})}{\partial \mu'} = \sum_{s=1}^{S}\frac{\partial \tilde{g}_n(y_n^s(\theta_0), \bar{\bar{\mu}})}{\partial \mu'} \xrightarrow{p} S \cdot M_\mu.$$

By subtracting (33) from (34) we get

$$\sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \tilde{\mu}) = M_\mu^{-1}\sqrt{n}\left[\tilde{g}_n(y_n, \mu(\theta_0)) - S^{-1}\sum_{s=1}^{S}\tilde{g}_n(y_n^s(\theta_0), \mu(\theta_0))\right]. \tag{35}$$

Using (35) and $\tilde{g}_n(y_n, \tilde{\mu}) = 0$, (32) can be rewritten as

$$\sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) = \sqrt{n}\left[\tilde{g}_n(y_n, \mu(\theta_0)) - S^{-1}\sum_{s=1}^{S}\tilde{g}_n(y_n^s(\theta_0), \mu(\theta_0))\right], \tag{36}$$

Because y_n and $y_n^s(\theta_0)$ $(s = 1, \ldots, S)$ are independent it follows that

$$\mathrm{AsyVar}[\sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0))] =$$
$$\mathrm{AsyVar}[\sqrt{n}\tilde{g}_n(y_n, \mu(\theta_0))] + S^{-2}\sum_{s=1}^{S}\mathrm{AsyVar}[\sqrt{n}\tilde{g}_n(y_n, \mu(\theta_0))] = \left(1 + \frac{1}{S}\right)\mathcal{I},$$

so that

$$\sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right)\mathcal{I}\right). \tag{37}$$

Plugging (37) into (31) gives the desired result.

Estimates of θ_1 and 95% confidence intervals for θ_1											
	$SN1$	$SL1$	$SA1$	$SN2$	$SL2$	$SA2$	$SM2$	DN	DL	DA	DM
$\theta_1^0 = 0.8522$											
$\hat{\theta}$	0.8364	0.8380	0.8380	0.8365	0.8382	0.8383	0.8415	0.8365	0.8382	0.8383	0.8415
L	0.8015	0.8022	0.8023	0.7989	0.8004	0.8005	0.8038	0.8031	0.8045	0.8046	0.8078
U	0.8618	0.8637	0.8638	0.8740	0.8758	0.8759	0.8790	0.8699	0.8718	0.8719	0.8750
CI	0.0604	0.0615	0.0615	0.0751	0.0754	0.0753	0.0751	0.0668	0.0673	0.0673	0.0673
$\theta_1^0 = 0.9868$											
$\hat{\theta}$	0.9798	0.9802	0.9804	0.9798	0.9809	0.9812	0.9850	0.9798	0.9809	0.9812	0.9850
L	0.9525	0.9509	0.9501	0.9659	0.9672	0.9674	0.9713	0.9666	0.9678	0.9681	0.9719
U	0.9874	0.9878	0.9882	0.9936	0.9941	0.9944	0.9977	0.9929	0.9935	0.9939	0.9974
CI	0.0349	0.0369	0.0381	0.0277	0.0269	0.0270	0.0264	0.0263	0.0257	0.0258	0.0255
$\theta_1^0 = 0.9978$											
$\hat{\theta}$	0.9860	0.9856	0.9857	0.9864	0.9875	0.9877	0.9912	0.9864	0.9875	0.9877	0.9912
L	0.9555	0.9496	0.9484	0.9746	0.9759	0.9761	0.9799	0.9753	0.9765	0.9767	0.9804
U	0.9920	0.9918	0.9922	0.9981	0.9984	0.9988	0.9993	0.9975	0.9979	0.9982	0.9991
CI	0.0365	0.0422	0.0438	0.0235	0.0225	0.0227	0.0195	0.0222	0.0214	0.0215	0.0186

Table 1: Point estimate, and asymptotic 95% confidence interval for θ_1 from a representative simulation of the AR(1) process with sample size 1000.

Average estimation time											
n	$SN1$	$SL1$	$SA1$	$SN2$	$SL2$	$SA2$	$SM2$	DN	DL	DA	DM
$\theta_1^0 = 0.8522$											
200	0.00	0.06	0.14	0.00	0.03	0.09	0.12	0.00	0.03	0.08	0.10
1000	0.00	0.30	0.31	0.00	0.15	0.17	0.20	0.00	0.12	0.14	0.16
2000	0.01	0.59	0.73	0.00	0.29	0.37	0.45	0.00	0.23	0.30	0.36
10000	0.03	2.86	2.91	0.00	1.41	1.60	1.84	0.00	1.25	1.33	1.52
$\theta_1^0 = 0.9868$											
200	0.00	0.06	0.14	0.00	0.05	0.13	0.21	0.00	0.04	0.11	0.20
1000	0.00	0.38	0.39	0.00	0.22	0.26	0.31	0.00	0.17	0.20	0.26
2000	0.01	0.75	0.92	0.00	0.41	0.54	0.65	0.00	0.32	0.41	0.51
10000	0.04	3.58	3.70	0.00	2.08	2.30	2.66	0.00	1.65	1.74	2.03
$\theta_1^0 = 0.9978$											
200	0.00	0.05	0.13	0.00	0.06	0.15	0.23	0.00	0.05	0.13	0.21
1000	0.00	0.35	0.35	0.00	0.27	0.31	0.42	0.00	0.22	0.25	0.36
2000	0.01	0.71	0.88	0.00	0.50	0.66	0.88	0.00	0.40	0.52	0.75
10000	0.05	3.83	3.98	0.00	2.52	2.75	3.22	0.00	2.07	2.18	2.61

Table 2: Average estimation time in seconds based on 1000 Monte Carlo experiments. Estimation was performed in R 2.10 (32 bit) on an iMac with 2.66 GHz Intel Core 2 Duo and 2 GB 800 MHz DDR2 SDRAM.

Bias and dispersion of estimates												
n	$CMLE$	$SN1$	$SL1$	$SA1$	$SN2$	$SL2$	$SA2$	$SM2$	DN	DL	DA	DM
$\theta_1^0 \times 1000 = 8522$												
200	-87	-322	-323	-324	-89	-79	-79	99	-89	-79	-79	100
	(402)	(643)	(654)	(655)	(407)	(413)	(413)	(426)	(408)	(413)	(413)	(426)
1000	-24	-64	-64	-64	-24	-22	-22	12	-24	-22	-22	12
	(172)	(191)	(195)	(195)	(173)	(178)	(178)	(178)	(173)	(178)	(178)	(177)
2000	-15	-33	-33	-33	-15	-14	-14	3	-15	-14	-14	3
	(120)	(128)	(130)	(130)	(121)	(123)	(123)	(122)	(121)	(123)	(123)	(122)
10000	-4	-8	-7	-7	-4	-3	-4	-0	-4	-4	-4	-0
	(53)	(54)	(55)	(55)	(53)	(55)	(55)	(55)	(53)	(55)	(55)	(55)
$\theta_1^0 \times 1000 = 9868$												
200	-108	-1223	-1207	-1224	-110	-103	-103	34	-110	-103	-103	36
	(227)	(1786)	(1800)	(1814)	(228)	(228)	(228)	(164)	(228)	(228)	(228)	(164)
1000	-21	-176	-181	-181	-21	-20	-20	22	-21	-20	-20	22
	(63)	(392)	(403)	(402)	(63)	(64)	(64)	(66)	(63)	(64)	(64)	(66)
2000	-11	-50	-56	-56	-11	-10	-10	10	-11	-10	-10	10
	(41)	(130)	(157)	(158)	(41)	(42)	(42)	(42)	(41)	(42)	(42)	(42)
10000	-3	-7	-7	-7	-3	-3	-3	1	-3	-3	-3	1
	(17)	(19)	(20)	(20)	(17)	(18)	(18)	(18)	(17)	(18)	(18)	(18)
$\theta_1^0 \times 1000 = 9978$												
200	-100	-1491	-1449	-1480	-108	-105	-104	-25	-108	-105	-104	-24
	(197)	(2000)	(1991)	(2016)	(198)	(196)	(197)	(112)	(197)	(196)	(196)	(111)
1000	-20	-510	-514	-519	-20	-19	-19	7	-20	-19	-19	7
	(42)	(771)	(782)	(786)	(42)	(41)	(42)	(28)	(42)	(41)	(42)	(28)
2000	-10	-253	-255	-256	-10	-9	-9	8	-10	-9	-9	7
	(25)	(435)	(444)	(446)	(25)	(25)	(25)	(21)	(25)	(25)	(25)	(21)
10000	-2	-29	-31	-30	-2	-2	-2	2	-2	-2	-2	2
	(8)	(103)	(107)	(106)	(8)	(8)	(8)	(8)	(8)	(8)	(8)	(8)

Table 3: Mean empirical bias × 1000 and (RMSE × 1000) of estimates for 1000 Monte Carlo simulations.

Rejection frequencies of overidentification tests											
n	$SN1$	$SL1$	$SA1$	$SN2$	$SL2$	$SA2$	$SM2$	DN	DL	DA	DM
$\theta_1^0 = 0.8522$											
200	0.102	0.092	0.093	0.068	0.074	0.074	0.084	0.077	0.072	0.071	0.065
1000	0.067	0.054	0.054	0.057	0.058	0.058	0.062	0.064	0.050	0.051	0.053
2000	0.058	0.053	0.053	0.048	0.047	0.047	0.049	0.056	0.052	0.052	0.051
10000	0.053	0.059	0.059	0.053	0.052	0.052	0.052	0.053	0.059	0.059	0.055
$\theta_1^0 = 0.9868$											
200	0.349	0.336	0.336	0.120	0.130	0.126	0.302	0.153	0.148	0.146	0.301
1000	0.142	0.124	0.124	0.079	0.080	0.081	0.088	0.086	0.072	0.072	0.079
2000	0.089	0.087	0.088	0.066	0.062	0.062	0.065	0.066	0.063	0.063	0.067
10000	0.062	0.069	0.069	0.059	0.054	0.054	0.054	0.056	0.062	0.062	0.061
$\theta_1^0 = 0.9978$											
200	0.547	0.536	0.537	0.200	0.203	0.204	0.578	0.229	0.229	0.218	0.579
1000	0.350	0.345	0.343	0.147	0.150	0.145	0.356	0.140	0.132	0.127	0.351
2000	0.243	0.234	0.234	0.105	0.104	0.101	0.182	0.107	0.102	0.101	0.190
10000	0.130	0.128	0.127	0.067	0.067	0.067	0.069	0.072	0.071	0.071	0.076

Table 4: Empirical rejection frequencies of overidentification tests at 5% nominal level for 1000 Monte Carlo simulations. The test has an asymptotic $\chi^2(2)$ distribution.

Rejection frequencies of likelihood ratio type tests											
n	$SN1$	$SL1$	$SA1$	$SN2$	$SL2$	$SA2$	$SM2$	DN	DL	DA	DM
$\theta_1^0 = 0.8522$											
200	0.215	0.203	0.201	0.062	0.063	0.064	0.096	0.061	0.064	0.065	0.091
1000	0.098	0.101	0.101	0.063	0.056	0.057	0.062	0.063	0.060	0.059	0.066
2000	0.084	0.077	0.078	0.059	0.060	0.060	0.067	0.055	0.060	0.061	0.064
10000	0.064	0.068	0.068	0.055	0.056	0.056	0.052	0.054	0.057	0.057	0.056
$\theta_1^0 = 0.9868$											
200	0.752	0.727	0.708	0.070	0.076	0.075	0.276	0.070	0.065	0.069	0.264
1000	0.383	0.371	0.360	0.054	0.054	0.053	0.120	0.051	0.053	0.053	0.117
2000	0.248	0.245	0.242	0.048	0.052	0.053	0.089	0.052	0.048	0.049	0.091
10000	0.108	0.113	0.111	0.057	0.060	0.060	0.072	0.054	0.061	0.061	0.072
$\theta_1^0 = 0.9978$											
200	0.929	0.913	0.858	0.070	0.068	0.069	0.198	0.073	0.073	0.067	0.172
1000	0.792	0.770	0.741	0.055	0.058	0.056	0.259	0.050	0.059	0.054	0.263
2000	0.629	0.615	0.604	0.052	0.051	0.052	0.236	0.054	0.058	0.057	0.235
10000	0.271	0.267	0.259	0.049	0.048	0.051	0.096	0.051	0.048	0.050	0.095

Table 5: Empirical rejection frequencies of likelihood ratio type tests at 5% nominal level for 1000 Monte Carlo simulations. The test has an asymptotic $\chi^2(1)$ distribution.

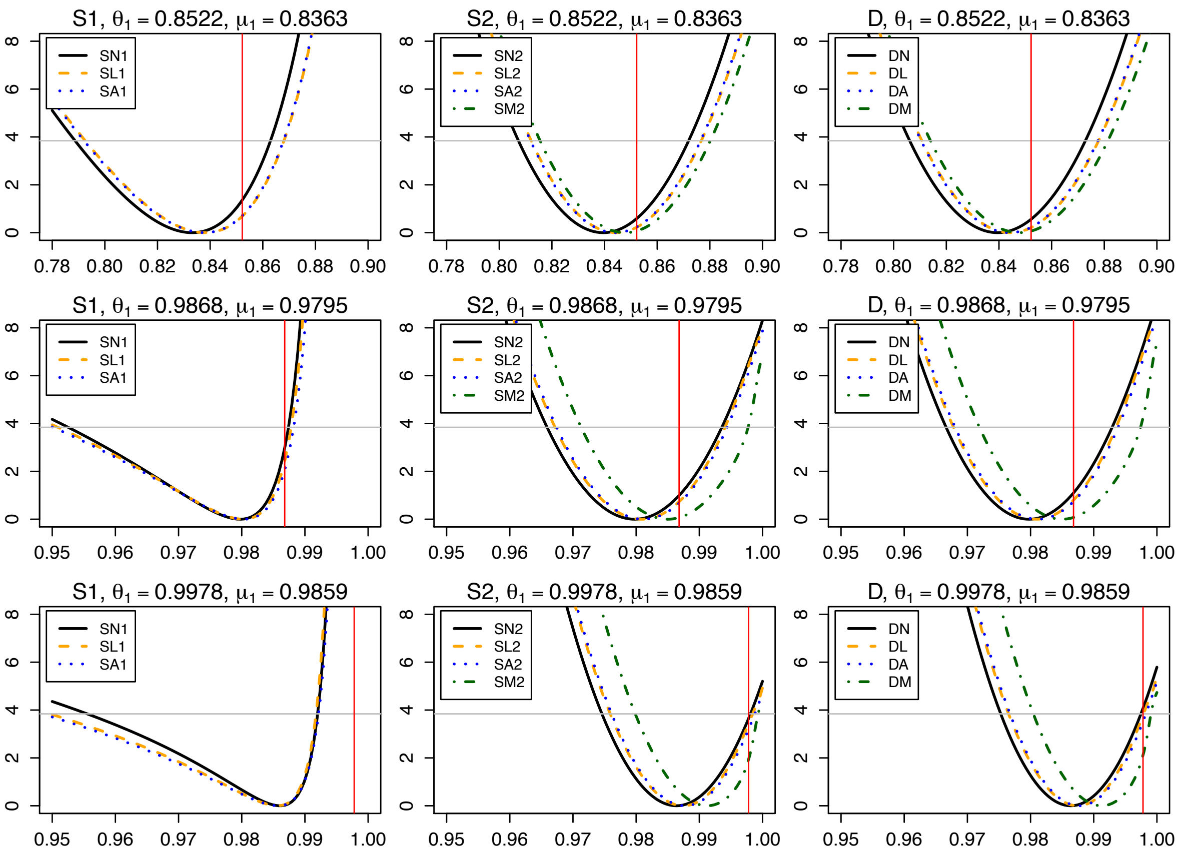

Figure 1: LR-type statistics for testing $H_0 : \theta_1 = \theta_1^0$ as functions of θ_1^0 from a representative sample of size $n = 1000$. The underlying model is described in Section 4.1. The horizontal grey line and the vertical red line represent the 95% $\chi^2(1)$ critical value and the true value of θ_1 respectively.