Econometrics Journal (2014), volume **17**, pp. 1–16.
Article No. ectj??????

Indirect Inference Based on the Score

Peter Fuleky† and Eric Zivot‡

†*UHERO and Department of Economics, University of Hawaii, Honolulu, HI 96822, USA.*
E-mail: fuleky@hawaii.edu
‡*Department of Economics, University of Washington, Seattle, WA 98195, USA.*
E-mail: ezivot@u.washington.edu
Received: August 2012

Summary 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.*

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 (1996b), or distance-based estimators proposed by Smith (1993) and refined by Gouriéroux et al. (1993). However, many studies have shown (e.g. Michaelides and Ng, 2000; Ghysels et al., 2003; Duffee and Stanton, 2008) that the score-based estimators often have 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 (1996b) score-based estimator but has finite sample properties that are very close to the distance-based estimators.

2. REVIEW OF INDIRECT INFERENCE

Indirect inference (II) techniques were introduced into the econometrics literature by Smith (1993), Gouriéroux et al. (1993), Bansal et al. (1994), Bansal et al. (1995) and Gallant and Tauchen (1996b), 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 $\widetilde{F}_\mu$

$$\tilde{\mu}_n = \arg\max_{\mu} \widetilde{Q}_n \left(\{y_t\}_{t=1,\dots,n}, \mu\right) \ , \tag{2.1}$$

$$\widetilde{Q}_n \left(\{y_t\}_{t=1,\dots,n}, \mu\right) = \frac{1}{n-m} \sum_{t=m+1}^{n} \tilde{f}(y_t; x_{t-1}, \mu) \ , \tag{2.2}$$

where $\tilde{f}(y_t; x_{t-1}, \mu)$ is the log density of y_t for the model $\widetilde{F}_\mu$ conditioned on $x_{t-1} = \{y_i\}_{i=t-m,\dots,t-1}$, $m \in \mathbb{N}$. 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 parameters θ. The link between the auxiliary model parameters and the structural parameters is given by the binding function $\mu(\theta)$, which is the functional solution of the asymptotic optimization problem

$$\mu(\theta) = \arg\max_{\mu} E_{F_\theta}[\tilde{f}(y_t; x_{t-1}, \mu)], \tag{2.3}$$

where $\lim_{n\to\infty} \widetilde{Q}_n \left(\{y_t\}_{t=1,\dots,n}, \mu\right) = E_{F_\theta}[\tilde{f}(y_t; x_{t-1}, \mu)]$, 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 (2.3) 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 (2.3).[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

$$\{y_t(\theta)\}_{t=1,\dots,Sn}, \ S \geq 1. \tag{2.4}$$

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

$$\{y_t^s(\theta)\}_{t=1,\dots,n}, \ s = 1, \dots, S, \ S \geq 1. \tag{2.5}$$

Using the simulated observations (2.4) or (2.5), the distance-based II estimators (subsequently also referred to as D estimators) are minimum distance estimators defined as

$$\hat{\theta}_S^{\mathrm{Dj}}(\widetilde{\Omega}_n) = \arg\min_{\theta} \ \mathrm{J}_S^{\mathrm{Dj}}(\theta, \widetilde{\Omega}_n) = \arg\min_{\theta} \ \left(\tilde{\mu}_n - \tilde{\mu}_S^{\mathrm{j}}(\theta)\right)' \widetilde{\Omega}_n \left(\tilde{\mu}_n - \tilde{\mu}_S^{\mathrm{j}}(\theta)\right), \ \mathrm{j} = \mathrm{L}, \mathrm{A}, \mathrm{M}, \tag{2.6}$$

where $\widetilde{\Omega}_n$ is a positive definite and symmetric weight matrix which may depend on the data

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

[2] Gallant and Tauchen (1996a) call the score-based II estimator the efficient method of moments (EMM) estimator.

through the auxiliary model, and the simulated binding functions are given by

$$\tilde{\mu}_S^{\mathrm{L}}(\theta) = \operatorname{argmax}_{\mu} \widetilde{Q}_{Sn}\left(\{y_t(\theta)\}_{t=1,\dots,Sn}, \mu\right), \tag{2.7}$$

$$\tilde{\mu}_S^{\mathrm{A}}(\theta) = \arg\max_{\mu} S^{-1} \sum_{s=1}^{S} \widetilde{Q}_n\left(\{y_t^s(\theta)\}_{t=1,\dots,n}, \mu\right), \tag{2.8}$$

$$\tilde{\mu}_S^{\mathrm{M}}(\theta) = S^{-1} \sum_{s=1}^{S} \arg\max_{\mu} \widetilde{Q}_n\left(\{y_t^s(\theta)\}_{t=1,\dots,n}, \mu\right). \tag{2.9}$$

The superscripts L, A, and M indicate how the binding function is computed from the simulated data: "L" denotes use of long simulations (2.4) in the objective function; "A" denotes maximizing an aggregation of S objective functions using (2.5); "M" denotes use of the mean of S estimated binding functions based on (2.5). 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 et al. (2000).

Using (2.4) or (2.5), the score-based II estimators (subsequently also referred to as S1 estimators) are one-step GMM estimators defined as

$$\hat{\theta}_S^{\mathrm{Sj1}}(\widetilde{\Sigma}_n) = \arg\min_{\theta} \ J_S^{\mathrm{Sj1}}(\theta) = \arg\min_{\theta} \tilde{g}_S^{\mathrm{j}}(\theta, \tilde{\mu}_n)' \widetilde{\Sigma}_n \tilde{g}_S^{\mathrm{j}}(\theta, \tilde{\mu}_n), \ \mathrm{j} = \mathrm{L}, \mathrm{A}, \tag{2.10}$$

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

$$\tilde{g}_S^{\mathrm{L}}(\theta, \tilde{\mu}_n) = \frac{1}{Sn - m} \sum_{t=m+1}^{Sn} \tilde{g}(y_t(\theta); x_{t-1}(\theta), \tilde{\mu}_n) \tag{2.11}$$

$$\tilde{g}_S^{\mathrm{A}}(\theta, \tilde{\mu}_n) = S^{-1} \sum_{s=1}^{S} \frac{1}{n - m} \sum_{t=m+1}^{n} \tilde{g}(y_t^s(\theta); x_{t-1}^s(\theta), \tilde{\mu}_n), . \tag{2.12}$$

Because (2.10) 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 (2.6) and score-based estimators (2.10) 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 $\widetilde{\mathcal{I}} = \lim_{n\to\infty} \operatorname{var}_{F_\theta}(\sqrt{n}\tilde{g}_n(y_n, \mu(\theta_0)))$ with $\tilde{g}_n(y_n, \mu(\theta)) = \frac{1}{n-m}\sum_{t=m+1}^{n} \tilde{g}(y_t; x_{t-1}, \mu(\theta))$, and $M_\mu = E_{F_\theta}[\widetilde{H}(y_t; x_{t-1}, \mu(\theta_0))]$ with $\widetilde{H}(y_t; x_{t-1}, \mu) = \frac{\partial^2 \tilde{f}(y_t; x_{t-1}, \mu)}{\partial\mu\partial\mu'}$. Using consistent estimates of these optimal weight matrices, the distance-based and score-based estimators are asymptotically equivalent with asymptotic variance matrix given by

$$V_S^* = \left(1 + \frac{1}{S}\right)\left(M_\theta' \widetilde{\mathcal{I}}^{-1} M_\theta\right)^{-1} = \left(1 + \frac{1}{S}\right)\left(\frac{\partial\mu(\theta_0)'}{\partial\theta} M_\mu' \widetilde{\mathcal{I}}^{-1} M_\mu \frac{\partial\mu(\theta_0)}{\partial\theta'}\right)^{-1}, \tag{2.13}$$

where

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

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 likelihood (SQML) estimator defined as

$$\hat{\theta}_S^{\mathrm{SQMLj}} = \arg\max_\theta \widetilde{Q}_n\left(\{y_t\}_{t=1,\ldots,n}, \tilde{\mu}_S^{\mathrm{j}}(\theta)\right),\ \mathrm{j} = \mathrm{L, A, M}. \tag{3.1}$$

Smith (1993) showed that (3.1) is consistent and asymptotically normal with asymptotic variance matrix given by

$$V_S^{SQML} = \left(1 + \frac{1}{S}\right)\left[\frac{\partial\mu(\theta_0)'}{\partial\theta} M_\mu \frac{\partial\mu(\theta_0)}{\partial\theta'}\right]^{-1} \frac{\partial\mu(\theta_0)'}{\partial\theta} \widetilde{\mathcal{I}} \frac{\partial\mu(\theta_0)}{\partial\theta'} \left[\frac{\partial\mu(\theta_0)'}{\partial\theta} M_\mu \frac{\partial\mu(\theta_0)}{\partial\theta'}\right]^{-1}, \tag{3.2}$$

which he showed is strictly greater (in a matrix sense) than the asymptotic variance (2.13) of the efficient II estimators. As noted by Gouriéroux et al. (1993), 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

$$\hat{\theta}_S^{\mathrm{Sj2}}(\widetilde{\Sigma}_n) = \arg\min_\theta J^{\mathrm{Sj2}}(\theta, \widetilde{\Sigma}_n) = \arg\min_\theta\ \tilde{g}_n^{\mathrm{j}}(\theta)'\widetilde{\Sigma}_n \tilde{g}_n^{\mathrm{j}}(\theta), \tag{3.3}$$

where

$$\tilde{g}_n^{\mathrm{j}}(\theta) = \frac{1}{n-m}\sum_{t=m+1}^{n} \tilde{g}(y_t; x_{t-1}, \tilde{\mu}_S^{\mathrm{j}}(\theta)),\ \mathrm{j} = \mathrm{L, A, M}. \tag{3.4}$$

The S2 estimator was not explicitly considered in Smith (1993). In contrast to the simulated scores (2.11) and (2.12), the score in (3.4) is evaluated with the observed data and the simulated binding function. The following Proposition gives the asymptotic properties of (3.3).

Proposition 3.1. *Under the regularity conditions in Gouriéroux and Monfort (1996), the score-based II estimators $\hat{\theta}_S^{Sj2}(\widetilde{\Sigma}_n)$ (j=L,A,M) defined in (3.3) are consistent and asymptotically normal, when S is fixed and $n \to \infty$:*

$$\sqrt{n}(\hat{\theta}_S^{Sj2}(\widetilde{\Sigma}_n) - \theta_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right)[M_\theta'\Sigma M_\theta]^{-1}\left[M_\theta'\Sigma\widetilde{\mathcal{I}}\Sigma M_\theta\right][M_\theta'\Sigma M_\theta]^{-1}\right). \tag{3.5}$$

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

Remark 3.2. When $\widetilde{\Sigma}_n$ is a consistent estimator of $\widetilde{\mathcal{I}}^{-1}$, the asymptotic variance of $\hat{\theta}_S^{\mathrm{Sj2}}(\widetilde{\Sigma}_n)$ in (3.5) 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 (2.13). Contrary to the claim in Gouriéroux and Monfort (1996), for a given auxiliary model the alternative score-based II estimator is not less efficient than the optimal traditional II estimators.

Remark 3.3. To see the relationship between the two score-based estimators, (2.10) and (3.3), note that the first order conditions (FOCs) of the optimization problem (2.3) defining $\mu(\theta)$ are

$$0 = E_{F_\theta}\left[\frac{\partial\tilde{f}(y_t; x_{t-1}, \mu)}{\partial\mu}\right]\Bigg|_{\mu=\mu(\theta)} \equiv \tilde{g}_E(y_t(\theta), \mu(\theta)) \equiv \tilde{g}_E(\theta, \mu(\theta)). \tag{3.6}$$

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 (3.3) evaluates the sample auxiliary score with a simulated binding function, it has certain properties that make it similar to the distance-based II estimator (2.6).

Remark 3.4. To see why the S1 and S2 estimators are asymptotically equivalent and efficient, and the SQML estimator is generally inefficient, consider the FOCs defining these estimators. From (2.10), the FOCs for the optimal S1 estimator are

$$0 = \frac{\partial \tilde{g}_S(\hat{\theta}_S, \tilde{\mu}_n)'}{\partial \theta} \widetilde{\mathcal{I}}_n^{-1} \tilde{g}_n(y_t; x_{t-1}, \tilde{\mu}_n) \ , \tag{3.7}$$

and, from (3.3), the FOCs for the optimal S2 estimator are

$$0 = \frac{\partial \tilde{\mu}_S(\hat{\theta}_S)'}{\partial \theta} \frac{\partial \tilde{g}_n(y_t; x_{t-1}, \tilde{\mu}_S(\hat{\theta}_S))'}{\partial \mu} \widetilde{\mathcal{I}}_n^{-1} \tilde{g}_n(y_t; x_{t-1}, \tilde{\mu}_S(\hat{\theta}_S)) \ , \tag{3.8}$$

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 (3.7) and (3.8) can be re-expressed as

$$0 = M_\theta' \widetilde{\mathcal{I}}^{-1} \tilde{g}_n(y_t; x_{t-1}, \mu(\theta_0)) + o_p(1) \ , \tag{3.9}$$

and

$$0 = \frac{\partial \mu(\theta_0)'}{\partial \theta} M_\mu' \widetilde{\mathcal{I}}^{-1} \tilde{g}_n(y_t; x_{t-1}, \mu(\theta_0)) + o_p(1) \ . \tag{3.10}$$

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 over-identified auxiliary score and produce efficient II estimators. In contrast, from (3.1) the FOCs for the SQML are

$$0 = \frac{\partial \mu_S(\theta_0)'}{\partial \theta} \tilde{g}_n(y_t; x_{t-1}, \mu(\theta_0)) + o_p(1) \ . \tag{3.11}$$

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

4. FINITE SAMPLE COMPARISON OF II ESTIMATORS

We compare the finite sample performance of the alternative score-based S2 estimator to the traditional S1 and D estimators using an Ornstein-Uhlenbeck (OU) process. Our analysis is motivated by Duffee and Stanton (2008), who compared the finite sample properties of traditional indirect estimators using highly persistent AR(1) models and found that the S1 estimator is severely biased, has wide confidence intervals, and performs poorly in coefficient and over-identification tests. We show that the alternative formulation of the score-based estimator leads to a remarkable improvement in its finite sample performance.

4.1. Model Setup

The true data generating process is an OU process of the form

$$F_\theta : dy = (\theta_0 - \theta_1 y)dt + \theta_2 dW, \quad dW \sim \text{iid N}(0, dt) \ , \tag{4.1}$$

and the auxiliary model is its Euler discretization

$$\widetilde{F}_\mu : y_t = \mu_0 \Delta + (1 - \mu_1 \Delta) y_{t-\Delta} + \mu_2 \sqrt{\Delta} \xi_{t-\Delta}, \quad \xi_{t-\Delta} \sim \text{iid N}(0, 1) \ . \tag{4.2}$$

Table 1. Just-Identified Estimation of the θ Parameters

Empirical Size of Likelihood Ratio Tests $H_0^{LR_1}: \theta_1 = \theta_{10}$

	n	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	0.869	0.206	0.285	0.800	0.162	0.251	0.724	0.171	0.252	0.063	0.098
	1000	0.321	0.082	0.083	0.314	0.076	0.080	0.288	0.078	0.082	0.058	0.060
θ^B	100	0.286	0.068	0.106	0.265	0.064	0.088	0.248	0.061	0.088	0.061	0.085
	1000	0.090	0.053	0.057	0.097	0.051	0.052	0.095	0.051	0.055	0.047	0.048

Empirical Size of Likelihood Ratio Tests $H_0^{LR}: (\theta_0, \theta_1, \theta_2) = (\theta_{00}, \theta_{10}, \theta_{20})$

	n	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	0.911	0.146	0.189	0.868	0.131	0.174	0.821	0.131	0.179	0.416	0.383
	1000	0.401	0.070	0.067	0.396	0.073	0.069	0.381	0.070	0.070	0.140	0.128
θ^B	100	0.404	0.088	0.112	0.379	0.087	0.104	0.375	0.083	0.104	0.176	0.164
	1000	0.128	0.058	0.057	0.117	0.063	0.057	0.117	0.064	0.057	0.068	0.055

Note: Empirical size of likelihood ratio tests for a nominal size $\alpha = 5\%$ and true value of the parameter vector $(\theta_{00}, \theta_{10}, \theta_{20})$. Results reported for just identified estimation of the OU process (4.1) with true parameter values: $\theta^A = (0, 0.01, 1)$ and $\theta^B = (0, 0.1, 1)$.

Observations are generated from the exact solution of the OU process

$$y_t = \frac{\theta_0}{\theta_1}(1 - e^{-\theta_1 \Delta}) + e^{-\theta_1 \Delta} y_{t-\Delta} + \theta_2 \sqrt{\frac{1 - e^{-2\theta_1 \Delta}}{2\theta_1}} \epsilon_t, \quad \epsilon_t \sim \text{iid N}(0, 1) \; . \tag{4.3}$$

A comparison of (4.3) and (4.2) reveals that the binding function (2.3) has the form

$$\mu_0(\theta) = \frac{\theta_0}{\theta_1 \Delta}\left(1 - e^{-\theta_1 \Delta}\right), \; \mu_1(\theta) = \frac{1}{\Delta}\left(1 - e^{-\theta_1 \Delta}\right), \; \mu_2(\theta) = \theta_2 \sqrt{\frac{1 - e^{-\theta_1 \Delta}}{2\theta_1 \Delta}} \tag{4.4}$$

and that $\tilde{\mu}_n = (\tilde{\mu}_{n0}, \tilde{\mu}_{n1}, \tilde{\mu}_{n2})$ is an asymptotically biased estimator of $\theta = (\theta_0, \theta_1, \theta_2)$ (see Lo, 1988). Without loss of generality, we set $\Delta = 1$ in equations (4.2) − (4.4) (see Fuleky, 2012). The analytically tractable OU process gives us the opportunity to compute non-simulation-based analogues (SN1, SN2, and DN) of the simulation-based estimators.

Because the finite sample performance of the estimators is mostly influenced by the speed of mean reversion, in our data generating process we vary θ_1 and consider the following two sets of true parameter values $\theta^A = (0, 0.01, 1)$ and $\theta^B = (0, 0.1, 1)$. The values $\theta_1^A = 0.01$ and $\theta_1^B = 0.1$ correspond to autoregressive coefficients equal to $e^{-0.01} = 0.99$ and $e^{-0.01} = 0.9$, respectively, in (4.3). In addition to estimating θ_0, θ_1, and θ_2, we also consider the case when θ_0 and θ_2 are assumed to be known, and the indirect estimators of θ_1 are over-identified ($r > p$). For the simulations (2.4) and (2.5), we set $S = 20$, so that the simulation-based estimators have a 95% asymptotic efficiency relative to the non-simulation-based estimators (see equation (2.13)). We analyze samples of size $n = \{100, 1000\}$, and our results are based on 1000 Monte Carlo simulations.

Table 2. Over-Identified Estimation of the θ_1 Parameter

Bias and [Root Mean Squared Error] of $\hat{\theta}_1$

	n	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	.2637	.0254	.0281	.2509	.0241	.0268	.2531	.0242	.0269	-.0094	-.0086
		[.3624]	[.0415]	[.0444]	[.3524]	[.0406]	[.0437]	[.3544]	[.0408]	[.0438]	[.0296]	[.0311]
	1000	.0154	.0022	.0023	.0151	.0021	.0021	.0152	.0021	.0021	-.0023	-.0022
		[.0268]	[.0057]	[.0057]	[.0264]	[.0057]	[.0057]	[.0265]	[.0057]	[.0058]	[.0060]	[.0060]
θ^B	100	.1409	.0175	.0242	.1430	.0149	.0218	.1430	.0149	.0219	-.0285	-.0214
		[.2339]	[.0569]	[.0607]	[.2395]	[.0566]	[.0602]	[.2391]	[.0566]	[.0602]	[.0634]	[.0622]
	1000	.0075	.0024	.0029	.0075	.0021	.0026	.0075	.0021	.0026	-.0017	-.0012
		[.0179]	[.0149]	[.0151]	[.0182]	[.0152]	[.0154]	[.0182]	[.0153]	[.0154]	[.0152]	[.0153]

Empirical Size of Over-Identification Tests

	n	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	0.394	0.160	0.181	0.382	0.165	0.165	0.383	0.167	0.164	0.194	0.160
	1000	0.149	0.085	0.089	0.148	0.082	0.076	0.147	0.080	0.076	0.091	0.088
θ^B	100	0.132	0.100	0.099	0.125	0.094	0.091	0.125	0.092	0.088	0.083	0.099
	1000	0.058	0.054	0.052	0.060	0.055	0.051	0.060	0.056	0.052	0.059	0.051

Empirical Size of Likelihood Ratio Tests

	n	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	0.933	0.048	0.105	0.884	0.041	0.092	0.846	0.040	0.091	0.404	0.383
	1000	0.451	0.039	0.038	0.432	0.045	0.041	0.418	0.045	0.042	0.132	0.127
θ^B	100	0.437	0.040	0.080	0.418	0.040	0.075	0.411	0.042	0.073	0.186	0.148
	1000	0.133	0.052	0.056	0.118	0.053	0.063	0.118	0.053	0.064	0.057	0.061

Note: Results for the $\theta^A = (0, 0.01, 1)$ and $\theta^B = (0, 0.1, 1)$ parameterizations of the OU process (4.1) when only the mean reversion parameter, θ_1, is being estimated, and θ_0 and θ_2 are being held fixed at their true values. Empirical size of tests for a nominal size $\alpha = 5\%$.

4.2. Results

In line with the proposition of Gouriéroux and Monfort (1996, pg. 66), the score-based and distance-based II estimators of a particular type (N, L, A, or M) produce equivalent results in a just identified setting. The bias and root mean squared error of the just identified estimators is summarized in Table 3 in Appendix B. Notably, each II estimator of θ_1 is biased upward, but in comparison to the others, the M-type estimators are more accurate with a tighter distribution around the true value.

Despite their equivalent distributional characteristics, the just identified II estimators don't have equal test performance. Table 1 summarizes the rejection rates of likelihood ratio tests of the hypotheses, $H_0^{LR_1} : \theta_1 = \theta_{10}$ and $H_0^{LR} : (\theta_0, \theta_1, \theta_2) = (\theta_{00}, \theta_{10}, \theta_{20})$, where $(\theta_{00}, \theta_{10}, \theta_{20})$ denotes the true value of the parameter vector. In both tests the S1 estimator is much more oversized than the S2 and D estimators. The large improvement in the performance of the S2 estimator over the

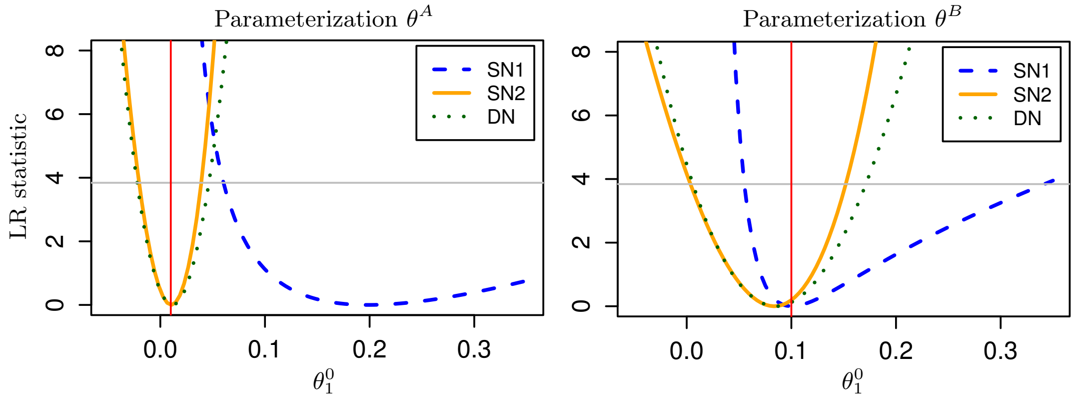

Figure 1. Plots of LR-type statistics for $H_0 : \theta_1 = \theta_{10}$ as functions of θ_{10} in the over-identified OU model with θ_0 and θ_2 being held fixed at their true values. The left and right panels display plots based on representative samples of size $n = 100$, and parameterizations $\theta^A = (0, 0.01, 1)$ and $\theta^B = (0, 0.1, 1)$, respectively. 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. The shape of the objective function is equivalent to the shape of the LR statistic except for a level shift.

S1 estimator can be attributed to using the simulated binding function instead of the simulated score for calibration. The shape of the S1 objective function is determined by the simulated score, $\tilde{g}_S(y(\theta), \tilde{\mu}_n)$, which depends on the variance of the simulated sample. Consequently, the S1 objective function quickly steepens as θ_1 approaches the non-stable region of the parameter space below $\hat{\theta}_1$. As a result, the confidence sets around the S1 estimates, which are upward biased in the θ_1 dimension, frequently exclude the true θ_1 parameter value. In contrast, the shape of the S2 and D objective functions is determined by the simulated binding function, $\tilde{\mu}_S(\theta)$, which is approximately linear around $\hat{\theta}_1$, and the roughly symmetric confidence sets around the estimates contain the true parameter value with higher frequency. As $\theta_1 \to 0$, the binding function slightly steepens and the confidence set tightens, which affects the rejection rate of the least-upward-biased M-type estimators. In joint tests the shrinkage of the confidence sets dominates the bias reduction of the M-type estimators and leads to higher rejection rates.

Table 2 shows that the S2 estimator retains its superiority over the S1 estimator in an over-identified setting.[3] The S1 estimator is up to ten times more biased than the S2 estimator (N, L and A-type), which itself exhibits some bias reduction compared to the D estimator. Here, θ_0 and θ_2 are being held fixed at the true values, which in general are different from the values that minimize the just identified objective function for a given set of observations, and θ_1 has to compensate for those restrictions when minimizing the over-identified objective function. This, in conjunction with the relatively mild penalty when θ_1 moves away from the non-stable region of the parameter space, will cause the over-identified S1 estimator to have a larger upward bias than the just identified S1 estimator. In contrast, due to the approximate linearity of the binding function and near symmetry of the S2 and D objective functions, the S2 and D estimators do not suffer from this excessive bias. However, because of the interaction between the weighting matrix and the moment conditions, the over-identified M-type estimators lose their bias correcting properties (see also Altonji and Segal, 1996). Finally, the over-identified S1 estimators have the

[3]The estimation times listed in Table 4 in Appendix C demonstrate that this improvement can be achieved without much additional computational cost.

highest rejection rates in both J and LR tests. 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 (see Figure 1).

5. CONCLUSION

We study the asymptotic and finite sample properties of a score-based II estimator that uses the sample auxiliary score evaluated at the simulated binding function. This estimator is asymptotically equivalent to the original score-based II estimator of Gallant and Tauchen (1996), but in finite samples behaves much more like the distance-based II estimator of Gouriéroux, Monfort and Renault (1993). In our Monte Carlo study of a continuous time OU process, the alternative score-based estimator exhibits greatly improved finite sample properties compared to the original one. Our results indicate that estimators operating through the simulated binding function are more suitable for highly persistent time series models than estimators operating through the simulated score.

ACKNOWLEDGEMENTS

This paper summarizes the main results in the first author's Ph.D. dissertation at the University of Washington. The second author greatly appreciates support from the Robert Richards chair. We are grateful to the editor and two anonymous referees for helpful suggestions.

REFERENCES

Altonji, J. and L. Segal (1996). Small-sample bias in GMM estimation of covariance structures. *Journal of Business & Economic Statistics 14*, 353–366.

Ball, C. and W. Torous (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance 3*, 215–238.

Bansal, R., A. Gallant, R. Hussey and G. Tauchen (1994). Computational aspects of nonparametric simulation estimation. In D. A. Belsley (Ed.), *Computational Techniques for Econometrics and Economic Analysis*, 3–22. Berlin: Springer.

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

Calzolari, G., F. Mealli and C. Rampichini (2001). Alternative simulation-based estimators of logit models with random effects. Econometrics Working Paper quaderno48, Universita' degli Studi di Firenze, Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti".

Duffee, G. and R. Stanton (2008). Evidence on simulation inference for near unit-root processes with Implications for term structure estimation. *Journal of Financial Econometrics 6*, 108–142.

Fuleky, P. (2012). On the choice of the unit period in time series models. *Applied Economics Letters 19*, 1179–1182.

Gallant, A. and G. Tauchen (1996a). Specification analysis of continuous time models in finance. In P. E. Rossi (Ed.), *Modelling Stock Market Volatility: Bridging the Gap to Continuous Time*, 357–384. San Diego, CA: Academic Press.

Gallant, A. and G. Tauchen (1996b). Which moments to match? *Econometric Theory 12*, 657–81.

Gallant, A. and G. Tauchen (2010). Simulated score methods and indirect inference for continuous time models. In Y. A it-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics, Volume I*, 427–477. Amsterdam: North Holland.

Ghysels, E., L. Khalaf and C. Vodounou (2003). Simulation based inference in moving average models. *Annales D'Économie et de Statistique 69*, 85–99.

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

Gouriéroux, C., A. Monfort and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, 85–118.

Gouriéroux, C., P.C.B. Phillips and J. Yu (2010). Indirect inference for dynamic panel models. *Journal of Econometrics 157*, 68–77.

Gouriéroux, C., E. Renault and N. Touzi (2000). Calibration by simulation for small sample bias correction. In R. Mariano et al. (Eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, 328–358. Cambridge: Cambridge University Press.

Jiang, W. and B. Turnbull (2004). The indirect method: inference based on intermediate statistics—A synthesis and examples. *Statistical Science 19*, 239–263.

Li, T. (2010). Indirect inference in structural econometric models. *Journal of Econometrics 157*, 120–128.

Michaelides, A. and S. Ng (2000). Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators. *Journal of Econometrics 96*, 231–266.

Phillips, P.C.B. and J. Yu (2009). Maximum likelihood and Gaussian estimation of continuous time models in finance. In T. G. Andersen et al. (Eds.), *Handbook of Financial Time Series*, 497–530. Berlin: Springer.

Smith, A.A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, 63–84.

APPENDIX A: PROOF OF PROPOSITION 3.1

We make the same set of assumptions as Gouriéroux and Monfort (1996, Appendix 4A, pages 85-86). For completeness we list them here:

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

ASSUMPTION A.2. $\tilde{f}_E(\theta,\mu)$ *has a unique maximum with respect to* $\mu:\mu(\theta)=\arg\max_\mu \tilde{f}_E(\theta,\mu)$.

ASSUMPTION A.3. $\tilde{f}_n\left(y_n,\mu\right)$ *and* $\tilde{f}_E(\theta,\mu)$ *are twice continuously differentiable with respect to* μ.

ASSUMPTION A.4. *The only solution to the asymptotic first order conditions,* $\lim\limits_{n\to\infty}\frac{\partial\tilde{f}_n(y_n,\mu)}{\partial\mu} = \frac{\partial\tilde{f}_E(\theta,\mu)}{\partial\mu} = \tilde{g}_E(\theta,\mu)$*, is* $\mu(\theta):\tilde{g}_E(\theta,\mu)=0\Rightarrow\mu=\mu(\theta)$.

ASSUMPTION A.5. *The equation* $\mu=\mu(\theta)$ *admits a unique solution in* θ.

ASSUMPTION A.6. $\mathrm{plim}_{n\to\infty}\frac{\partial^2\tilde{f}_n(y_n,\mu(\theta))}{\partial\mu\partial\mu'} = E_{F_\theta}[\widetilde{H}(y_t;x_{t-1},\mu(\theta_0))] = M_\mu$

ASSUMPTION A.7. $\sqrt{n}\tilde{g}_n(y_n,\mu(\theta_0)) = \sqrt{n}\frac{\partial\tilde{f}_n(y_n,\mu(\theta_0))}{\partial\mu} \xrightarrow{d} N(0,\widetilde{\mathcal{I}})$ *as* $n\to\infty$.

In addition:

ASSUMPTION A.8. M_μ *is full rank. This assumption in conjunction with the implicit function theorem ensures the first order differentiability of the binding function.*

ASSUMPTION A.9. $E_{F_\theta}\sup_{\mu;\|\mu-\mu(\theta_0)\|_1\le\varepsilon}\left\|\frac{\partial^2\tilde{f}(y_t;x_{t-1},\mu)}{\partial\mu\partial\mu'}\right\|_2<\infty$ *for some* $\varepsilon>0$ *small enough and suitable norms* $\|\cdot\|_j$, $j=1,2$. *This assumption is necessitated by the mean value expansions below.*

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 $n\to\infty$

$$\tilde{g}_n(y_n,\mu(\theta)) \xrightarrow{p} \tilde{g}_E(\theta_0,\mu(\theta)),$$

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

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 (3.3) is

$$\frac{\partial\tilde{g}_n(y_n,\tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))'}{\partial\theta}\widetilde{\Sigma}_n\tilde{g}_n(y_n,\tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))=0. \tag{A.1}$$

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 (A.1) gives

$$\frac{\partial \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\hat{\theta}_S^{\mathrm{L}}))'}{\partial \theta} \widetilde{\Sigma}_n \left[\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) + \frac{\partial \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\bar{\theta}))}{\partial \mu'} \frac{\partial \tilde{\mu}_S^{\mathrm{L}}(\bar{\theta}))}{\partial \theta'} (\hat{\theta}_S^{\mathrm{L}} - \theta_0) \right] = 0 \ , \tag{A.2}$$

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 (A.2) then gives

$$\sqrt{n}(\hat{\theta}_S^{\mathrm{L}} - \theta_0) = -\left[M_\theta' \Sigma M_\theta\right]^{-1} M_\theta' \Sigma \sqrt{n} \tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) + o_p(1). \tag{A.3}$$

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

$$\begin{aligned} \sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) &= \sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_n) + \frac{\partial \tilde{g}_n(y_n, \bar{\mu})}{\partial \mu'} \sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \tilde{\mu}_n) \\ &= \sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_n) + M_\mu \sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \tilde{\mu}_n) + o_p(1), \end{aligned} \tag{A.4}$$

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

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

so that

$$\sqrt{n}(\tilde{\mu}_n - \mu(\theta_0)) = -M_\mu^{-1} \sqrt{n}\tilde{g}_n(y_n, \mu(\theta_0)) + o_p(1). \tag{A.5}$$

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

$$\begin{aligned} \sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \mu(\theta_0)) &= -\left[\frac{\partial \tilde{g}_{Sn}(y_{Sn}(\theta_0), \bar{\bar{\mu}})}{\partial \mu'}\right]^{-1} \sqrt{n}\tilde{g}_{Sn}(y_{Sn}(\theta_0), \mu(\theta_0)) \\ &= -S^{-1} M_\mu^{-1} \sqrt{n} \sum_{s=1}^{S} \tilde{g}_n(y_n^s(\theta_0), \mu(\theta_0)) + o_p(1), \end{aligned} \tag{A.6}$$

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 (A.5) from (A.6) we get

$$\sqrt{n}(\tilde{\mu}_S^{\mathrm{L}}(\theta_0) - \tilde{\mu}_n) = M_\mu^{-1} \sqrt{n} \left[\tilde{g}_n(y_n, \mu(\theta_0)) - S^{-1} \sum_{s=1}^{S} \tilde{g}_n(y_n^s(\theta_0), \mu(\theta_0))\right]. \tag{A.7}$$

Using (A.7) and $\tilde{g}_n(y_n, \tilde{\mu}_n) = 0$, (A.4) can be rewritten as

$$\sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) = \sqrt{n} \left[\tilde{g}_n(y_n, \mu(\theta_0)) - S^{-1} \sum_{s=1}^{S} \tilde{g}_n(y_n^s(\theta_0), \mu(\theta_0))\right] \ , \tag{A.8}$$

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

$$\text{AsyVar}[\sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0))] =$$

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

so that

$$\sqrt{n}\tilde{g}_n(y_n, \tilde{\mu}_S^{\mathrm{L}}(\theta_0)) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right)\mathcal{I}\right). \tag{A.9}$$

Plugging (A.9) into (A.3) gives the desired result.

APPENDIX B: ESTIMATION RESULTS FOR JUST IDENTIFIED ESTIMATORS

We summarize the bias and root mean squared error of the just identified estimators in Table 3. The auxiliary estimates, $\tilde{\mu}$, are affected by discretization and finite sample biases, which partially offset each other. While the former bias dominates for quickly mean-reverting processes, the latter dominates for highly persistent ones (Ball and Torous, 1996; Fuleky, 2012; Phillips and Yu, 2009). In a just identified setting, the non-simulation based II estimators produce the same result as the conditional maximum likelihood (CML) estimator applied to (4.3); hence they fully correct the discretization bias of the auxiliary estimator. In line with the proposition of Gouriéroux and Monfort (1996, pg. 66), the score-based and distance-based II estimators of a particular type (L, A, or M) produce equivalent results in a just identified setting. Furthermore, our results for the just identified M-type estimator confirm its finite sample bias correcting properties previously demonstrated by Gouriéroux et al. (2000), Gouriéroux et al. (2010), and Phillips and Yu (2009) among others.

Table 3. Just-Identified Estimation of the θ Parameters

Bias and [Root Mean Squared Error] of $\hat{\theta}_0$

	n	$\tilde{\mu}_n$	CML	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	-.0011	-.0014	-.0014	-.0014	-.0014	-.0007	-.0007	-.0007	-.0002	-.0002	-.0002	.0013	.0013
		[.2785]	[.2928]	[.2928]	[.2928]	[.2928]	[.2890]	[.2891]	[.2891]	[.2896]	[.2896]	[.2896]	[.1605]	[.1605]
	1000	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0005	.0004	.0004
		[.0499]	[.0504]	[.0504]	[.0504]	[.0504]	[.0508]	[.0508]	[.0508]	[.0508]	[.0508]	[.0508]	[.0389]	[.0389]
θ^B	100	-.0005	.0000	.0000	.0000	.0000	.0002	.0002	.0002	.0003	.0003	.0003	.0009	.0009
		[.1394]	[.1543]	[.1543]	[.1543]	[.1543]	[.1551]	[.1551]	[.1551]	[.1551]	[.1551]	[.1551]	[.1174]	[.1174]
	1000	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
		[.0323]	[.0341]	[.0341]	[.0341]	[.0341]	[.0348]	[.0348]	[.0348]	[.0348]	[.0348]	[.0348]	[.0336]	[.0336]

Bias and [Root Mean Squared Error] of $\hat{\theta}_1$

	n	$\tilde{\mu}_n$	CML	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	.0579	.0613	.0613	.0613	.0613	.0596	.0596	.0596	.0597	.0597	.0597	.0133	.0133
		[.0709]	[.0762]	[.0762]	[.0762]	[.0762]	[.0753]	[.0753]	[.0753]	[.0754]	[.0754]	[.0754]	[.0466]	[.0466]
	1000	.0049	.0051	.0051	.0051	.0051	.0049	.0049	.0049	.0049	.0049	.0049	.0004	.0004
		[.0078]	[.0080]	[.0080]	[.0080]	[.0080]	[.0080]	[.0080]	[.0080]	[.0080]	[.0080]	[.0080]	[.0066]	[.0066]
θ^B	100	.0370	.0495	.0495	.0495	.0495	.0473	.0473	.0473	.0474	.0474	.0474	.0011	.0011
		[.0673]	[.0839]	[.0839]	[.0839]	[.0839]	[.0828]	[.0828]	[.0828]	[.0829]	[.0829]	[.0829]	[.0702]	[.0702]
	1000	-.0004	.0051	.0051	.0051	.0051	.0049	.0049	.0049	.0049	.0049	.0049	.0009	.0009
		[.0141]	[.0165]	[.0165]	[.0165]	[.0165]	[.0168]	[.0168]	[.0168]	[.0168]	[.0168]	[.0168]	[.0161]	[.0161]

Bias and [Root Mean Squared Error] of $\hat{\theta}_2$

	n	$\tilde{\mu}_n$	CML	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	-.0220	.0129	.0129	.0129	.0129	.0135	.0135	.0135	.0136	.0136	.0136	.0103	.0103
		[.0745]	[.0753]	[.0753]	[.0753]	[.0753]	[.0775]	[.0775]	[.0775]	[.0776]	[.0776]	[.0776]	[.0762]	[.0762]
	1000	-.0065	.0009	.0009	.0009	.0009	.0013	.0013	.0013	.0013	.0013	.0013	.0005	.0005
		[.0235]	[.0228]	[.0228]	[.0228]	[.0228]	[.0234]	[.0234]	[.0234]	[.0234]	[.0234]	[.0234]	[.0234]	[.0234]
θ^B	100	-.0605	.0105	.0105	.0105	.0105	.0110	.0110	.0110	.0110	.0110	.0110	.0036	.0036
		[.0909]	[.0783]	[.0783]	[.0783]	[.0783]	[.0802]	[.0802]	[.0802]	[.0802]	[.0802]	[.0802]	[.0787]	[.0787]
	1000	-.0494	.0009	.0009	.0009	.0009	.0012	.0012	.0012	.0012	.0012	.0012	.0005	.0005
		[.0539]	[.0238]	[.0238]	[.0238]	[.0238]	[.0243]	[.0243]	[.0243]	[.0243]	[.0243]	[.0243]	[.0243]	[.0243]

Note: Estimation results for the OU process (4.1) with the following two sets of true parameter values: $\theta^A = (0, 0.01, 1)$ and $\theta^B = (0, 0.1, 1)$. For comparison, we also list the distributional characteristics of the auxiliary estimator ($\tilde{\mu}_n$) and the conditional maximum likelihood (CML) estimator applied to (4.3).

APPENDIX C: COMPUTATIONAL EFFICIENCY OF BINDING FUNCTION BASED ESTIMATORS

Gallant and Tauchen (2010) criticize distance-based II for its computational inefficiency, because it potentially involves two nested optimizations with the estimator of the simulated binding function being embedded within the D estimator. However, if one chooses a simple auxiliary model that can be estimated by ordinary least squares, as suggested by Calzolari et al. (2001) and Li (2010), the speed disadvantage of the binding function based estimators disappears. In such a setting, the binding function in the S2 and D estimators does not involve a nested optimization; only the analytical expression for the least squares estimator of the auxiliary model is evaluated. Our results in Table 4 indicate that the speed of the simulation based S2 or D estimator of a particular type (L or A) is comparable to the speed of the S1 estimator of the same type when the auxiliary model is estimated by ordinary least squares.

Table 4. Estimation Time

Just Identified Estimation

	n	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	0.013	0.015	0.005	0.157	0.161	0.154	0.739	0.492	0.476	1.381	1.311
	1000	0.014	0.042	0.005	1.556	1.171	1.132	1.676	1.173	1.146	1.669	1.636
θ^B	100	0.013	0.014	0.005	0.160	0.161	0.151	0.730	0.486	0.460	0.906	0.904
	1000	0.013	0.040	0.005	1.433	1.140	1.097	1.606	1.153	1.100	1.571	1.479

Over-Identified Estimtation

	n	SN1	SN2	DN	SL1	SL2	DL	SA1	SA2	DA	SM2	DM
θ^A	100	0.004	0.004	0.002	0.020	0.026	0.021	0.099	0.082	0.071	0.169	0.149
	1000	0.004	0.005	0.001	0.202	0.137	0.132	0.236	0.134	0.129	0.225	0.217
θ^B	100	0.004	0.003	0.001	0.021	0.020	0.017	0.104	0.060	0.054	0.180	0.152
	1000	0.004	0.005	0.001	0.181	0.134	0.121	0.218	0.134	0.124	0.194	0.179

Note: Average estimation times (in seconds) for the following two sets of true parameter values of the OU process (4.1): $\theta^A = (0, 0.01, 1)$ and $\theta^B = (0, 0.1, 1)$. The estimation was carried out on a Mac Mini with a 2GHz Intel Core i7 processor and 8GB of memory.