Further Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation

Peter Fuleky and Eric Zivot

January 9, 2010

Abstract

We study indirect estimation methods with a special emphasis on issues related to continuous time models of the interest rate. Using a highly persistent discrete AR(1) model, Duffee and Stanton (2008) (DS) argued for the superiority of indirect inference (II) over the efficient method of moments (EMM) for estimating term structure models. We extend the work of DS by conducting a thorough comparison of analytic and simulation-based EMM and II estimators in the context of continuous time models of the interest rate, and we confront issues that are specific to this setting, such as the choice of an auxiliary model and the relative importance of discretization bias and finite sample bias. We arrive at a different conclusion than DS by considering an alternative formulation of the EMM estimator that performs better than the original one, and overall its behavior closely matches that of II. We also show that the excessive bias of the EMM estimator is particular to the over-identified setting considered by DS.

1 Introduction

Indirect estimators take advantage of a simplified auxiliary model that is easier to estimate than the true 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 parameter to the auxiliary statistic is used to calibrate the structural parameter. Depending on the formulation of the auxiliary statistic used in their objective function, the indirect estimators are usually placed into one of two categories: *efficient method of moments* (EMM) made popular by Gallant and Tauchen (1996), or *indirect inference* (II) originally proposed by Smith Jr (1993) and Gouriéroux et al. (1993). The former method is based on the auxiliary score, and the latter one is based on the binding function. In this paper, we also consider an alternative EMM-2 estimator (Fuleky and Zivot, 2010) that uses the the binding function and through it inherits some of the features of the II estimator.

II and EMM were originally developed with the intent to enable parameter estimation in models with intractable likelihood functions, such as diffusions. Both estimators have been independently analyzed in continuous time settings: Andersen and Lund (1997) and Zhou (2001) employed EMM to estimate some single factor interest rate processes, and Gouriéroux et al. (1993) and Broze et al. (1995) gave examples on estimating diffusions with II. However, Gallant and Tauchen (1996) and later Jiang and Turnbull (2004) argued that both estimators can be accommodated by a unifying framework for indirect estimation procedures. To our knowledge there has been no direct comparison of the methods in a continuous time setting, but Duffee and Stanton (2008) (DS) compared EMM and II by estimating a discrete first order autoregressive (AR(1)) model. DS found that EMM is extremely biased, and that despite its wide confidence intervals, EMM has a high empirical rejection rate in tests. They attribute this latter result to EMM having an asymmetric criterion function, which was also documented by Tauchen (1998). [1]

We extend the study of DS to the estimation of continuous time models of the interest rate, and add EMM-2 to the line-up of estimators that we compare. A one factor model of the interest rate can be written as a continuous time stochastic process of the following form

$$F_\theta : dy = \alpha(y,\theta)dt + \beta(y,\theta)dW \ , \quad dW \sim \text{iid N}(0, dt) \ , \tag{1}$$

where W is a standard Brownian motion, $\alpha(y,\theta)$ is a drift function, and $\beta(y,\theta)$ is a diffusion function. If the model is correctly specified and its likelihood function is tractable, the structural θ parameters can be estimated by maximum likelihood. However, as Ball and Torous (1996) and

[1] Both, DS and Tauchen (1998), showed that the asymmetry of the criterion function is caused by the presence of the population variance in the EMM criterion function, which puts a large penalty on the structural parameter close to the boundary of stationarity. While this property helps to avoid explosive parameter estimates, it causes the confidence interval to cover only little of the structural parameter space between the point estimate and the boundary. In contrast, the II objective function does not contain the population variance and it remains symmetric around the point estimate. The difference in the shapes of the objective functions implies that if the true parameter value falls into the interval between the point estimate and the boundary of stability, the EMM estimate is rejected with much higher probability in tests than the II estimate. This was one of DS's main critiques of EMM.

Phillips and Yu (2007) show, the ML estimates of the drift parameters have a strong finite sample bias in highly persistent processes. This finite sample bias is similar to the one observed originally by Hurwicz (1950) and Marriott and Pope (1954) in persistent autoregressive models.

In cases when the likelihood function of (1) does not have a closed-form analytic expression, its discrete approximation

$$F_\mu : y_t = y_{t-\Delta} + \alpha(y_{t-\Delta}, \mu)\Delta + \beta(y_{t-\Delta}, \mu)\sqrt{\Delta}\, \epsilon_t, \quad \epsilon_t \sim \text{iid N}(0,1) \ , \tag{2}$$

is usually estimated, see Chan et al. (1992) and Broze et al. (1995). But Lo (1988) points out that this naive estimator is mis-specified for θ and gives asymptotically biased estimates. Nevertheless, the crude Euler discretization in (2) represents a natural choice of an auxiliary model which can be estimated using least squares techniques. While the μ parameter estimates are asymptotically biased for θ, the indirect methods implicitly correct this discretization bias of the naive estimator by ultimately estimating the underlying structural model.

In addition, Gouriéroux et al. (2000) and Gouriéroux et al. (2006) indicate that the II estimator can deliver further finite sample bias correction beyond the asymptotic bias correction if it uses a finite sample binding function. Because this finite sample bias correcting property hinges on the use of the binding function, it is not present in the original EMM estimator, but it becomes available in EMM-2. Thus the EMM-2 estimator can be employed to mitigate both sources of bias that are affecting the naive estimator of diffusion parameters.

The crude Euler discretization in (2) implies a just identified indirect estimator with the same number of auxiliary parameters as there are structural ones. DS's critique of EMM vis-a-vis II was partially based on results that were specific to an over-identified setting, and we show that over-identification introduces additional bias beyond the discretization and finite sample biases observed in a just identified model. Gouriéroux and Monfort (1996) show that in just-identified models the EMM and II estimates are identical, which counters some of DS's criticism of EMM. The rest of DS's criticism can be addressed by an alternative formulation of the EMM estimator: the EMM-2 estimator is calibrated, just like II, via the binding function. Therefore, it is void of the population variance and the explosive behavior close to the non-sationary region. Consequently, the EMM-2 estimator not only gives similar point estimates as II, but performs comparably in tests too.

The main contribution of this study to the existing literature is a thorough analysis of the finite sample properties of EMM, EMM-2 and II estimators in a continuous time setting. In Section 2, I give an overview of indirect estimators and introduce the EMM-2 estimator. In Section **??**, I describe some practical considerations related to indirect estimators, and in Section 3, I illustrate some of the issues related to continuous time models. In Sections 4 and **??**, I compare the finite sample performance of EMM, EMM-2, and II via a Monte Carlo study: Section 4 deals with parameter estimation in an OU process, and Section **??** deals with the distribution of corresponding bond prices. Section 5 concludes.

2 Indirect Estimators Defined

This section is laying the groundwork for the remainder of the paper. Following Gouriéroux and Monfort (1996), we give a brief overview of the existing analytical and simulation based indirect estimation methods. In addition, we consider an alternative formulation of the EMM estimator and derive its asymptotic distribution.

2.1 Components of Indirect Estimators

There are three components present in all indirect estimators: (1) the true structural model whose parameters are one's ultimate interest, (2) its auxiliary approximation, and (3) the binding function connecting these two models.

2.1.1 Structural Model

Assume that a sample of n observations $\{y_t\}_{t=\Delta,\ldots,n\Delta=T}$ with observation interval Δ, are generated from a strictly stationary and ergodic probability model F_θ, $\theta \in \mathbb{R}^p$, with joint density $f(y_{-m\Delta}, ..., y_{-\Delta}, y_0; \theta)$. This is the *structural model.* For example, a one factor model of the interest rate can be written as a continuous time stochastic process of the form given in (1). It may be hard or impossible to directly estimate the parameters of the structural model. However, the ability to generate quasi-samples from the structural model is a prerequisite for simulation based inference. If the continuous time model does not have a closed form solution, then simulations can be generated from its discrete approximation, such as the weak order 1 fine Euler discretization, or the strong order 1 Milshtein scheme, and others.

2.1.2 Auxiliary Model

Indirect estimators were originally developed with the intent to enable parameter estimation in models with intractable likelihood functions (Gallant and Tauchen, 1996; Gouriéroux et al., 1993). When the joint density of the true process is unknown, it is assumed that an approximation, or so called *auxiliary model*, can be estimated instead. The auxiliary model $\widetilde{F}_\mu$ is indexed by the parameter $\mu \in \mathbb{R}^r$, with $r \geq p$.

The natural choice for the auxiliary model of the diffusion in (1) is its crude Euler discretization (2), which implies just-identified estimators, that is, $r = p$. Over-identification only occurs if restrictions are imposed on the structural parmeters. This choice differs from that of Andersen and Lund (1997) and Gallant and Tauchen (1997) who used highly parametrized over-identified semi non-parametric (SNP) auxiliary models. They argued for this highly parameterized auxiliary model for efficiency reasons, but Phillips and Yu (2007) show that the naive estimator of discretized diffusions compares well with the maximum likelihood estimator, and the crude Euler discretization gives a sufficiently good description of the observed data.

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

$$\tilde{\mu} = \arg\max_{\mu} \tilde{f}_n\left(\{y_t\}_{t=\Delta,\ldots,n\Delta}, \mu\right), \tag{3}$$

where $\tilde{f}_n$ is the sample objective function associated with the auxiliary model $\widetilde{F}_\mu$. Because the auxiliary model is assumed to have a tractable conditional likelihood function, $\tilde{f}_n$ can be written as

$$\tilde{f}_n\left(\{y_t\}_{t=\Delta,\ldots,n\Delta}, \mu\right) = \frac{1}{n-m} \sum_{t=(m+1)\Delta}^{n\Delta} \tilde{f}(y_t; x_{t-\Delta}, \mu), \tag{4}$$

where $\tilde{f}(y_t; x_{t-\Delta}, \mu)$ is the log density of y_t for the model $\widetilde{F}_\mu$ conditioned on $x_{t-\Delta} = \{y_{t-m\Delta}, ..., y_{t-\Delta}\}$, and μ can be estimated by conditional maximum likelihood. The derivative of the log-density $\tilde{f}(y_t; x_{t-\Delta}, \mu)$ with respect to the auxiliary parameter μ is given by the score vector

$$\tilde{g}(y_t; x_{t-\Delta}, \mu) = \frac{\partial \tilde{f}(y_t; x_{t-\Delta}, \mu)}{\partial \mu} \tag{5}$$

While the direct, or naive, estimates of the auxiliary μ parameter will be asymptotically biased for θ, the indirect methods implicitly correct this discretization bias of the naive estimator, by ultimately estimating the underlying structural model as opposed to its discrete approximation.

2.1.3 Binding Function

Indirect estimators use the auxiliary model information summarized by some auxiliary statistic, such as the auxiliary score or the auxiliary estimates, to obtain estimates of the structural parameters θ. The connection between the two sets of parameters is given by a mapping from the structural parameters to the auxiliary parameters, the so-called *binding function*, $\mu(\theta)$, which is the functional solution of the asymptotic optimization problem

$$\mu(\theta) = \arg\max_{\mu} \lim_{n\to\infty} \tilde{f}_n\left(\{y_t\}_{t=\Delta,\ldots,n\Delta}, \mu\right) = \arg\max_{\mu} E_{F_\theta}[\tilde{f}(y_0; x_{-\Delta}, \mu)], \tag{6}$$

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

As discussed in Gouriéroux and Monfort (1996), indirect estimators differ in how they use (6) to define an estimating equation. The first type, typically called the *indirect inference* (II) estimator, originally proposed by Smith Jr (1993) and Gouriéroux et al. (1993), finds θ to minimize the distance between $\mu(\theta)$ and $\tilde{\mu}$. When the model is just-identified ($r = p$), the indirect estimator solves $\tilde{\mu} - \mu(\theta) = 0$, in particular, $\hat{\theta} = \mu^{-1}(\tilde{\mu})$. Thus, $\hat{\theta}$ sets the value of the binding function equal to that of the auxiliary estimate.

The second type of indirect estimator, made popular by Gallant and Tauchen (1996) and typically called the *efficient method of moments* (EMM), finds θ to satisfy the first order conditions associated with (6) when evaluated at $\tilde{\mu}$. Gouriéroux and Monfort (1996) and Jiang and Turnbull (2004) point out that for a given auxiliary model $\widetilde{F}_\mu$, EMM can be looked at as II based on the auxiliary score. EMM gets its name from the fact that if $\widetilde{F}_\mu$ is a very good approximation to F_θ then the score associated with $\widetilde{F}_\mu$ constitutes the moments to use for efficient estimation of θ.

The EMM-2 estimator takes an alternative route for calibration, and evaluates the score of the auxiliary estimator (5) with the binding function (6). It is asymptotically equivalent to the original EMM estimator, but behaves differently in finite samples. The next subsections summarize the various analytic and simulation based indirect estimators.

2.2 Analytic Estimators

Analytic, or non-simulation based, versions of EMM and II can be defined if $\mu(\theta)$ is known, or if the expected score associated with $\widetilde{F}_\mu$ can be evaluated analytically under F_θ.

2.2.1 EMM

The analytic EMM estimators make use of the population moment condition

$$E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \mu(\theta))\right] = 0, \tag{7}$$

based on (6). The analytic EMM estimator of Gallant and Tauchen (1996), which we call EN1, requires an analytic representation for $E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \tilde{\mu})\right]$, and is defined as

$$\begin{aligned} \hat{\theta}^{\mathrm{EN1}}(\widetilde{\Sigma}_n) &= \arg\min_\theta J^{\mathrm{EN1}}(\theta, \widetilde{\Sigma}_n) \\ &= \arg\min_\theta \; E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \tilde{\mu})\right]' \widetilde{\Sigma}_n E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \tilde{\mu})\right], \end{aligned} \tag{8}$$

where $\widetilde{\Sigma}_n$ is a positive definite and symmetric weight matrix which may depend on the data through the auxiliary model. If $\tilde{\mu} \xrightarrow{p} \mu(\theta)$ then $E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \tilde{\mu})\right] \xrightarrow{p} E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \mu(\theta))\right] = 0$.

2.2.2 EMM-2

We consider a second type of analytic EMM estimator, EN2, that requires a known analytic representation for the binding function $\mu(\theta)$. This estimator is a *generalized method of moments* (GMM) estimator that makes use of the population moment condition (7) and is defined as

$$\hat{\theta}^{\mathrm{EN2}}(\widetilde{\Sigma}_n) = \arg\min_\theta J^{\mathrm{EN2}}(\theta, \widetilde{\Sigma}_n) = \arg\min_\theta \; \tilde{g}_n(\theta)' \widetilde{\Sigma}_n \tilde{g}_n(\theta), \tag{9}$$

where

$$\tilde{g}_n(\theta) = \frac{1}{n-m} \sum_{t=(m+1)\Delta}^{n\Delta} \frac{\partial \tilde{f}(y_t; x_{t-\Delta}, \mu(\theta))}{\partial \mu} \tag{10}$$

is the sample score evaluated at $\mu(\theta)$. Under F_θ it follows from the ergodic theorem that $\tilde{g}_n(\theta) \xrightarrow{p} E_{F_\theta}[\tilde{g}(y_0; x_{-\Delta}, \mu(\theta))] = 0$ so that $\hat{\theta}^{\text{EN2}}$ is asymptotically equivalent to $\hat{\theta}^{\text{EN1}}$ (Fuleky and Zivot, 2010).

The optimal EMM and EMM-2 estimators use the weight matrix $\widetilde{\Sigma}^* \xrightarrow{p} \mathcal{I}^{-1}$, where

$$\mathcal{I} = \lim_{n\to\infty} \text{var}_{F_\theta}(\sqrt{n}\tilde{g}_n(\theta)), \tag{11}$$

is the asymptotic variance of the sample score evaluated at $\mu(\theta)$. As discussed in Gallant and Tauchen (1996), if $\widetilde{F}_\mu$ is a very good approximation to F_θ, then a consistent estimate of $\mathcal{I}$ based on the auxiliary model is

$$\widetilde{\mathcal{I}}_n = \frac{1}{n-m} \sum_{t=(m+1)\Delta}^{n\Delta} \tilde{g}(y_t; x_{t-\Delta}, \tilde{\mu}) \tilde{g}(y_t; x_{t-\Delta}, \tilde{\mu})' \,. \tag{12}$$

Otherwise, a heteroskedasticity and autocorrelation consistent estimator should be used.

2.2.3 Indirect Inference

The analytic II estimator, IN, is a *classical minimum distance estimator* of the form

$$\hat{\theta}^{\text{IN}}(\tilde{\Omega}_n) = \arg\min_\theta J^{\text{IN}}(\theta, \widetilde{\Omega}_n) = \arg\min_\theta \; (\tilde{\mu} - \mu(\theta))' \widetilde{\Omega}_n (\tilde{\mu} - \mu(\theta)), \tag{13}$$

where $\tilde{\Omega}_n$ is a positive definite and symmetric weight matrix which may depend on the data through the auxiliary model. DS call (13) the asymptotic II estimator.

The optimal II estimator uses the weight matrix $\widetilde{\Omega}^* \xrightarrow{p} M_\mu \mathcal{I}^{-1} M_\mu$, where

$$M_\mu = E_{F_\theta}\left[\frac{\partial^2 \tilde{f}(y_0; x_{-\Delta}, \mu(\theta))}{\partial\mu\partial\mu'}\right] \,. \tag{14}$$

An estimate of the optimal weight matrix is $\widetilde{\Omega}^*_n = \widetilde{H}_n \widetilde{\mathcal{I}}_n^{-1} \widetilde{H}_n$, where

$$\widetilde{H}_n = \frac{1}{n-m} \sum_{t=(m+1)\Delta}^{n\Delta} \frac{\partial^2 \tilde{f}(y_t; x_{t-\Delta}, \tilde{\mu})}{\partial\mu\partial\mu'} \,, \tag{15}$$

is the sample Hessian associated with $\widetilde{F}_\mu$, and $\widetilde{\mathcal{I}}_n$ is defined in (12). With an estimate of the efficient weight matrix, the II objective function has the form

$$J^{\text{IN}}(\theta, \widetilde{\Omega}^*_n) = \left[\widetilde{H}_n(\tilde{\mu} - \mu(\theta))\right]' \widetilde{\mathcal{I}}_n^{-1} \left[\widetilde{H}_n(\tilde{\mu} - \mu(\theta))\right] \,, \tag{16}$$

which is similar in form to the efficient EMM objective functions $J^{\text{EN1}}(\theta, \widetilde{\Sigma}^*_n)$ and $J^{\text{EN2}}(\theta, \widetilde{\Sigma}^*_n)$. In fact, the EN1 and EN2 estimators can be looked at as II estimators that are calibrated by using the score, instead of the distance between the auxiliary estimate and the binding function.

2.3 Simulation Based Estimators

If the analytic forms of $\mu(\theta)$ and $E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \tilde{\mu})\right]$ are not known then the analytic EMM and II estimators are not feasible. If it is possible to simulate from F_θ for a fixed θ, then simulation-based versions of (8), (9) and (13) can be solved to obtain the simulation-based EMM and II estimators of θ.

Simulated observations $\{y_t(\theta)\}$ from F_θ to be used in estimation can be drawn in two ways (Gouriéroux and Monfort, 1996). First, a long pseudo-data series of size $S \cdot n$ is simulated giving

$$\{y_t(\theta)\}_{t=\Delta,\ldots,Sn\Delta}, \quad S \geq 1. \tag{17}$$

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

$$\{y_t^s(\theta)\}_{t=\Delta,\ldots,n\Delta}, \quad s = 1, \ldots, S, \quad S \geq 1\,. \tag{18}$$

2.3.1 EMM Estimators

Corresponding to the two non-simulation based EMM estimators (8) and (9) there are two simulation-based EMM estimators. The simulation-based EMM estimator corresponding to EN1 in (8) uses simulations to approximate the expected value of the score. Based on the two types of simulated samples (17) and (18), $E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \tilde{\mu})\right]$ can be approximated using

$$\begin{aligned}
\tilde{g}_{Sn}(\theta, \tilde{\mu}) &= \frac{1}{S(n-m)} \sum_{t=(m+1)\Delta}^{Sn\Delta} \frac{\partial \tilde{f}\left(y_t(\theta); x_{t-\Delta}(\theta), \tilde{\mu}\right)}{\partial \mu}, \\
\bar{\tilde{g}}_n^S(\theta, \tilde{\mu}) &= \frac{1}{S(n-m)} \sum_{s=1}^{S} \sum_{t=(m+1)\Delta}^{n\Delta} \frac{\partial \tilde{f}\left(y_t^s(\theta); x_{t-\Delta}^s(\theta), \tilde{\mu}\right)}{\partial \mu}.
\end{aligned}$$

The simulation-based EMM estimators corresponding to (8) are then

$$\hat{\theta}_S^{\text{EL1}}(\tilde{\Sigma}_n) = \arg\min_\theta \ \text{J}_S^{\text{EL1}}(\theta, \tilde{\Sigma}_n) = \arg\min_\theta \tilde{g}_{Sn}(\theta, \tilde{\mu})' \tilde{\Sigma}_n \tilde{g}_{Sn}(\theta, \tilde{\mu}) \ , \tag{19}$$

$$\hat{\theta}_S^{\text{EA1}}(\tilde{\Sigma}_n) = \arg\min_\theta \ \text{J}_S^{\text{EA1}}(\theta, \tilde{\Sigma}_n) = \arg\min_\theta \bar{\tilde{g}}_n^S(\theta, \tilde{\mu})' \tilde{\Sigma}_n \bar{\tilde{g}}_n^S(\theta, \tilde{\mu}) \ . \tag{20}$$

The L superscript in (19) indicates that the EMM estimator is exploiting a *long series* simulation principle, and the A superscript in (20) indicates that the estimator is exploiting an *aggregate score* simulation principle. The EL1 estimator is implemented in Gallant and Tauchen (2004) and is used in most empirical applications of EMM in macroeconomics and finance.

2.3.2 EMM-2 Estimators

The simulation-based EMM-2 estimator corresponding to EN2 in (9) uses simulations to approximate the binding function $\mu(\theta)$ in the sample score $\tilde{g}_n(\theta) = \frac{1}{n-m} \sum_{t=(m+1)\Delta}^{n\Delta} \frac{\partial \tilde{f}(y_t; x_{t-\Delta}, \mu(\theta))}{\partial \mu}$.

Gouriéroux and Monfort (1996) show how the binding function can be approximated in three different ways. In the first approximation, a long pseudo-data series as in (17) is simulated and $\mu(\theta)$ is estimated using:

$$\tilde{\mu}_S^{\mathrm{L}}(\theta) = \underset{\mu}{\operatorname{argmax}}\ \tilde{f}_{Sn}\left(\{y_t(\theta)\}_{t=\Delta,\ldots,Sn\Delta}, \mu\right) . \tag{21}$$

In the second and third approximations, S pseudo-data series of size n are simulated as in (18) and $\mu(\theta)$ is estimated using

$$\tilde{\mu}_S^{\mathrm{A}}(\theta) = \arg\max_{\mu} \frac{1}{S}\sum_{s=1}^{S} \tilde{f}_n\left(\{y_t^s(\theta)\}_{t=\Delta,\ldots,n\Delta}, \mu\right) . \tag{22}$$

$$\tilde{\mu}_S^{\mathrm{M}}(\theta) = \frac{1}{S}\sum_{s=1}^{S} \arg\max_{\mu}\ \tilde{f}_n\left(\{y_t^s(\theta)\}_{t=\Delta,\ldots,n\Delta}, \mu\right), \tag{23}$$

The M superscript indicates that $\mu(\theta)$ is approximated using the *mean of auxiliary estimators*. The approximation $\tilde{\mu}_S^{\mathrm{M}}(\theta)$ is computationally expensive as it requires S optimizations, whereas the approximations $\tilde{\mu}_S^{\mathrm{L}}(\theta)$ and $\tilde{\mu}_S^{\mathrm{A}}(\theta)$ only require a single optimization. The higher computation cost associated with $\tilde{\mu}_S^{\mathrm{M}}(\theta)$ may be justified, however, due to its bias correction properties as shown by Gouriéroux et al. (2000) and Gouriéroux et al. (2006).

The simulation-based EMM-2 estimators corresponding to (9) which use (21), (23) or (22) are defined by

$$\hat{\theta}_S^{\mathrm{Ej2}}(\widetilde{\Sigma}_n) = \arg\min_{\theta} J^{\mathrm{Ej2}}(\theta, \widetilde{\Sigma}_n) = \arg\min_{\theta}\ \tilde{g}_n^{\mathrm{j}}(\theta)'\widetilde{\Sigma}_n\tilde{g}_n^{\mathrm{j}}(\theta), \tag{24}$$

where

$$\tilde{g}_n^{\mathrm{j}}(\theta) = \frac{1}{n-m}\sum_{t=(m+1)\Delta}^{n\Delta} \tilde{g}(y_t; x_{t-\Delta}, \tilde{\mu}_S^{\mathrm{j}}(\theta)), \mathrm{j} = \mathrm{L}, \mathrm{M}, \mathrm{A}. \tag{25}$$

is the sample score evaluated at the simulated binding function. To my knowledge, these simulation-based EMM estimators have not been considered before.

The EL2 or EA2 estimators are asymptotically equivalent to but computationally more expensive than the EL1 or EA1 estimators because the binding function must be re-estimated at each iteration of the optimization in (24). As I will demonstrate in Section **??**, the finite sample behavior of the estimators defined in (24) can be substantially different from those defined in (19) and (20) especially for highly persistent data.

2.3.3 II Estimators

For simulation-based II, simulations, as in (21)-(23), are used to approximate $\mu(\theta)$ in (13). The simulation-based II estimators are defined by

$$\hat{\theta}_S^{\mathrm{Ij}}(\widetilde{\Omega}_n) = \arg\min_{\theta} \mathrm{J}_S^{\mathrm{Ij}}(\theta, \widetilde{\Omega}_n) = \arg\min_{\theta}\ \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{j}}(\theta)\right)' \widetilde{\Omega}_n \left(\tilde{\mu} - \tilde{\mu}_S^{\mathrm{j}}(\theta)\right) , \mathrm{j} = \mathrm{L}, \mathrm{M}, \mathrm{A}. \tag{26}$$

The IM estimator is most commonly used in practice, and DS called it the finite-sample II estimator.

2.4 Asymptotic Properties

The asymptotic properties of EL1, EA1, IL, IA, and IM estimators are derived in Gouriéroux et al. (1993), Gouriéroux and Monfort (1996) and Gallant and Tauchen (1996). The results for the EL2 estimators are given in the appendix to this chapter. Under regularity conditions described in Gouriéroux and Monfort (1996), the optimal analytic EMM and II estimators are consistent and asymptotically normally distributed with asymptotic variance matrices given by

$$W_{\mathrm{EN1}}^{*} = W_{\mathrm{EN2}}^{*} = \left(M_{\theta}' \widetilde{\Sigma}^{*} M_{\theta} \right)^{-1}, \tag{27}$$

$$W_{\mathrm{IN}}^{*} = \left(\frac{\partial \mu(\theta)'}{\partial \theta} \widetilde{\Omega}^{*} \frac{\partial \mu(\theta)}{\partial \theta'} \right)^{-1}, \tag{28}$$

where

$$M_{\theta} = \left\{ \frac{\partial}{\partial \theta'} E_{F_{\theta}} \left[\frac{\partial \tilde{f}(y_0; x_{-\Delta}, \mu)}{\partial \mu} \right] \right\} \Big|_{\mu=\mu(\theta)}. \tag{29}$$

For fixed S, the simulation-based EMM and II estimators are also consistent and asymptotically normally distributed with asymptotic variance matrices given (27) and (28), respectively, scaled by $(1 + 1/S)$. Gouriéroux and Monfort (1996) derived the result

$$\frac{\partial \mu(\theta)}{\partial \theta'} = -M_{\mu}^{-1} M_{\theta}, \tag{30}$$

from which it follows that (27) and (28) are equal and the optimal EMM, EMM-2, and II estimators are asymptotically equivalent.

2.5 Classical Tests

Consider a hypothesis defined by $H_0 : q(\theta) = \eta_0$ for a smooth function q from $\mathbb{R}^p$ to $\mathbb{R}^{p_1}$. The LR-type test statistics based on the analytical EMM and II estimators are

$$\mathrm{LR}^{\mathrm{ENk}}(\eta_0) = n \left[\mathrm{J}^{\mathrm{ENk}}(\hat{\theta}^{\mathrm{Ek}}(\eta_0, \widetilde{\Sigma}_n^{*}), \widetilde{\Sigma}_n^{*}) - \mathrm{J}^{\mathrm{ENk}}(\hat{\theta}^{\mathrm{Ek}}(\widetilde{\Sigma}_n^{*}), \widetilde{\Sigma}_n^{*}) \right], k = 1, 2 \tag{31}$$

$$\mathrm{LR}^{\mathrm{IN}}(\eta_0) = n \left[\mathrm{J}^{\mathrm{IN}}(\hat{\theta}^{\mathrm{IN}}(\eta_0, \widetilde{\Omega}_n^{*}), \widetilde{\Omega}_n^{*}) - \mathrm{J}^{\mathrm{IN}}(\hat{\theta}^{\mathrm{IN}} \widetilde{\Omega}_n^{*}), \widetilde{\Omega}_n^{*}) \right], \tag{32}$$

where $\hat{\theta}(\eta_0, \widetilde{\Sigma}_n^{*})$ and $\hat{\theta}(\eta_0, \widetilde{\Omega}_n^{*})$ denote the optimal EMM and II estimators constrained by $H_0 : q(\theta) = \eta_0$, respectively. Under $H_0 : q(\theta) = \eta_0$, these statistics are asymptotically distributed chi-square with p_1 degrees of freedom.

The LR-type test statistics for the simulation-based EMM and II estimators must be scaled by $S/(S + 1)$ to account for the increase in variability due to the simulations. That is, $\mathrm{LR}_S^{\mathrm{Ejk}}(\eta_0) = \frac{S}{S+1} \mathrm{LR}^{\mathrm{ENk}}(\eta_0) (j = \mathrm{L}, \mathrm{M}, \mathrm{A}; k = 1, 2)$ and $\mathrm{LR}_S^{\mathrm{Ij}}(\eta_0) = \frac{S}{S+1} \mathrm{LR}^{\mathrm{IN}}(\eta_0) (j = \mathrm{L}, \mathrm{M}, \mathrm{A})$. For fixed S, these statistics are also asymptotically distributed chi-square with p_1 degrees of freedom.

Confidence sets for individual elements θ_j may be constructed by defining $q(\theta) = \theta_j^0$ and inverting the LR-type statistics using a $\chi^2(1)$ critical value.

When the auxiliary model $\widetilde{F}_\mu$ has more parameters than the true model F_θ, the scaled optimized value of the optimal EMM and II objective functions can be used to test the validity of the overidentifying restrictions imposed by the auxiliary model. Under the null of correct specification, these statistics are asymptotically distributed chi-square with $r-p$ degrees of freedom. For the non-simulation based estimators the overidentification test statistics are

$$n\mathrm{J}^{\mathrm{ENk}}(\hat{\theta}^{\mathrm{Ek}}(\widetilde{\Sigma}_n^*), \widetilde{\Sigma}_n^*),\ n\mathrm{J}^{\mathrm{IN}}(\hat{\theta}^{\mathrm{IN}}(\widetilde{\Omega}_n^*), \widetilde{\Omega}_n^*); \mathrm{k}=1,2 \tag{33}$$

For the simulation-based estimators, the over-identification statistics are given by (33) scaled by $S/(S+1)$.

3 Illustration of Issues Related to Continuous Time Models

In the following subsections, we use an analytically tractable OU model (continuous time counterpart of a discrete AR(1) process used by Duffee and Stanton (2008)) to highlight the practical issues associated with the use of indirect methods to estimate continuous time one factor models of the interest rate.

3.1 Structural Model

Assume that a sample of n observations $\{y_t\}_{t=\Delta\ldots,n\Delta=T}$ with observation interval Δ, are generated from an OU process of the form

$$F_\theta : y = (\theta_0 - \theta_1 y)dt + \theta_2 dW, \quad dW \sim \text{iid N}(0, dt)\ , \tag{34}$$

with $\theta_0 > 0,\ \theta_1 > 0,\ \theta_2 > 0$. A positive θ_1 ensures stationarity and mean reversion of the process towards its long run mean value θ_0/θ_1. The highly persistent nature of interest rates implies that θ_1 is close to zero, and historical interest rate data imply $\theta_0/\theta_1 \approx 0.07$ (see for example Chan et al. (1992) and Broze et al. (1995)) The ability to simulate quasi-samples from the structural model is a prerequisite for simulation based inference. In the case of the OU process, observations can be generated from its exact solution (Vasicek, 1977)

$$y_t = \frac{\theta_0}{\theta_1}(1 - e^{-\theta_1\Delta}) + e^{-\theta_1\Delta} y_{t-\Delta} + \theta_2\sqrt{\frac{1-e^{-2\theta_1\Delta}}{2\theta_1}}\epsilon_t, \quad \epsilon_t \sim \text{iid N}(0,1)\ , \tag{35}$$

where Δ is the observation interval. It is equivalent to a simple AR(1) model, $y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t, \quad u_t \sim \text{iid N}(0, \alpha_2)$, with the reparameterization $\alpha_0 = \frac{\theta_0}{\theta_1}(1-e^{-\theta_1\Delta})$, $\alpha_1 = e^{-\theta_1\Delta}$, and $\alpha_2 = \theta_2\sqrt{\frac{1-e^{-2\theta_1\Delta}}{2\theta_1}}$. The interpretation of the parameter values is tied to the value of Δ. Many studies set $\Delta = 1$ for a monthly (weekly) observation interval, in which case θ_1 reflects the monthly (weekly) mean reversion and θ_2 the monthly (weekly) volatility. We choose to standardize Δ, so that $\Delta = 1$ represents a year, $\Delta = 1/12$ a month, and $\Delta = 1/50$ a week. This way the parameters always have yearly interpretations, that is, θ_1 always represents the yearly mean reversion and θ_2 the yearly volatility. Fuleky (2009) further explores the issue of choosing a unit interval in time series models in a greater detail.

3.2 Auxiliary Model

For the OU process we consider the following crude Euler discretization as the auxiliary model

$$\widetilde{F}_\mu : y_t = \mu_0\Delta + (1-\mu_1\Delta)y_{t-\Delta} + \mu_2\sqrt{\Delta}\xi_t, \quad \xi_t \sim \text{iid N}(0,1) \ . \tag{36}$$

where Δ is the observation interval. It is equivalent to a simple AR(1) model, $y_t = \beta_0 + \beta_1 y_{t-1} + v_t, \quad v_t \sim \text{iid N}(0,\beta_2)$, with the reparameterization $\beta_0 = \mu_0\Delta$, $\beta_1 = (1-\mu_1\Delta)$, and $\beta_2 = \mu_2\sqrt{\Delta}$.

For the discretized OU process, the sample objective function can be written as

$$\tilde{f}_n\left(\{y_t\}_{t=\Delta,\ldots,n\Delta},\mu\right) = \frac{1}{n-1}\sum_{t=2\Delta}^{n\Delta}\left[-\frac{1}{2}\ln(2\pi\mu_2^2\Delta) - \frac{1}{2}\frac{\xi_t^2}{\mu_2^2\Delta}\right] \ , \tag{37}$$

where $\xi_t = y_t - \mu_0\Delta - (1-\mu_1\Delta)y_{t-\Delta}$. The derivative of the log-density $\tilde{f}(y_t;y_{t-\Delta},\mu)$ with respect to the auxiliary parameter μ is given by the score vector

$$\tilde{g}(y_t;y_{t-\Delta},\mu) = \frac{\partial\tilde{f}(y_t;y_{t-\Delta},\mu)}{\partial\mu} = \begin{pmatrix} \dfrac{1}{\mu_2^2}\xi_t \\ -\dfrac{1}{\mu_2^2}\xi_t y_{t-\Delta} \\ \dfrac{1}{\mu_2}\left(\dfrac{\xi_t^2}{\mu_2^2\Delta} - 1\right) \end{pmatrix} \ , \tag{38}$$

and the analytical solution to the auxiliary estimator is equivalent to the least squares estimator

$$\tilde{\mu} = \begin{pmatrix} \dfrac{1}{\Delta}\left[\overline{y_t} - (1-\tilde{\mu}_1\Delta)\overline{y_{t-\Delta}}\right] \\ \dfrac{1}{\Delta}\left[1 - \dfrac{\overline{y_t y_{t-\Delta}} - \overline{y_t}\ \overline{y_{t-\Delta}}}{\overline{y_{t-\Delta}^2} - (\overline{y_{t-\Delta}})^2}\right] \\ \sqrt{\dfrac{\overline{\tilde{\xi}_t^2}}{\Delta}} \end{pmatrix} , \tag{39}$$

where $\tilde{\xi}_t = y_t - \tilde{\mu}_0\Delta - (1-\tilde{\mu}_1\Delta)y_{t-\Delta}$ and $\overline{(\cdot)}$ represents the mean value.

While the economic interpretation of the μ parameters in (36) is similar to that of the θ parameters in (34) or (35), they are not the same: the auxiliary model is mis-specified and the maximum likelihood estimates of the auxiliary parameters are biased and inconsistent (Lo, 1988).

3.3 Binding Function

For the OU process, $\mu(\theta)$ is given by the probability limit of (39) under F_θ. The binding function can be derived analytically, or by just comparing the coefficients of (35) and (36) as is done by Broze et al. (1998) and Phillips and Yu (2007), and is given by

$$\mu_0(\theta) = \frac{\theta_0}{\theta_1 \Delta}(1 - e^{-\theta_1 \Delta}), \quad \mu_1(\theta) = \frac{1}{\Delta}(1 - e^{-\theta_1 \Delta}), \quad \mu_2(\theta) = \theta_2 \sqrt{\frac{1 - e^{-2\theta_1 \Delta}}{2\theta_1 \Delta}} \,. \tag{40}$$

Because each θ vector is mapped to a distinct $\mu(\theta)$ vector, this binding function is bijective and hence invertible. Note that in (40) the difference $\mu(\theta) - \theta$ represents the asymptotic discretization bias of the auxiliary estimator. This bias is a function of the parameter values and the observation interval, and it increases with the latter. Thus, a higher frequency of observations will result in lower asymptotic discretization bias of the auxiliary parameter estimates: $\Delta \to 0 \Rightarrow \mu(\theta) \to \theta$. For $\theta_1 = 0.1$, increasing the observation interval from weekly ($\Delta = 1/50$) to monthly ($\Delta = 1/12$), quadruples the discretization bias of $\mu_1(\theta)$ from -10^{-4} to -4×10^{-4}.

When the model is just-identified ($r = p$), the IN estimator solves $\tilde{\mu} - \mu(\theta) = 0$. Thus, $\hat{\theta}$ sets the value of the binding function equal to that of the auxiliary estimator. Asymptotically both contain the same discretization bias, and therefore an asymptotically unbiased structural estimator of the OU parameters is given by $\hat{\theta}^{\mathrm{IN}} = \mu^{-1}(\tilde{\mu})$, that is, by inverting (40)

$$\hat{\theta}_0^{\mathrm{IN}}(\tilde{\mu}) = -\frac{\tilde{\mu}_0}{\tilde{\mu}_1 \Delta} \log(1 - \tilde{\mu}_1 \Delta), \quad \hat{\theta}_1^{\mathrm{IN}}(\tilde{\mu}) = -\frac{1}{\Delta} \log(1 - \tilde{\mu}_1 \Delta), \quad \hat{\theta}_2^{\mathrm{IN}}(\tilde{\mu}) = \tilde{\mu}_2 \sqrt{-\frac{2\log(1 - \tilde{\mu}_1 \Delta)}{1 - e^{2\log(1 - \tilde{\mu}_1 \Delta)}}} \,. \tag{41}$$

Note, in the just identified setting, there is no optimization required to obtain the value of the estimator. Furthermore, $E_{F_\theta}\left[\tilde{g}(y_0; x_{-\Delta}, \tilde{\mu})\right] = 0$ and $\tilde{g}_n(y_t; x_{t-\Delta}, \mu(\hat{\theta}^{\mathrm{EN2}})) = 0$ imply that $\tilde{\mu} = \mu(\hat{\theta}^{\mathrm{ENj}})$ for j=1,2. Therefore, $\hat{\theta}^{\mathrm{IN}}(\tilde{\mu}) = \hat{\theta}^{\mathrm{EN1}}(\tilde{\mu}) = \hat{\theta}^{\mathrm{EN2}}(\tilde{\mu})$; see also Gouriéroux and Monfort (1996). The equality of the $\hat{\theta}$ across estimators does not necessarily carry over to the over-identified case: when the number of auxiliary parameters exceeds the number of the structural ones, the finite sample results will depend on the weighting matrices $\widetilde{\Sigma}_n$ and $\widetilde{\Omega}_n$, and the functional forms of the individual auxiliary statistics. We show in Section XX that if the auxiliary statistics contain nonlinearities, over-identifying restrictions on the model parameters will cause additional bias in the estimates.

3.4 Simulation

It is rarely possible to simulate from F_θ for a fixed θ, as in the case of the OU process with its exact solution in (35). For most continuous time models there is no closed form solution of the structural model available. In such cases simulations can be generated from a discrete approximation of the continuous process, such as the fine Euler discretization (weak order 1) or the Milshtein scheme (strong order 1).

To illustrate a fine Euler discretization of the OU process (34), first a sequence $\{y_t\}_{t=\delta,\ldots,nk\delta=T}$ is obtained by dividing the observation interval Δ into k subintervals of length $\delta = \Delta/k$ and generating

$$F_{\theta,\delta} : y_t = y_{t-\delta} + \theta_0\delta - \theta_1\delta y_{t-\delta} + \theta_2\sqrt{\delta}\epsilon_{t-\delta} \tag{42}$$

$$= \theta_0\delta + (1-\theta_1\delta)y_{t-\delta} + \theta_2\sqrt{\delta}\epsilon_{t-\delta}, \quad \epsilon_{t-\delta} \sim iid\ N(0,1)\ , \tag{43}$$

where δ represents the duration of the simulation step. Then, by selecting every k-th data point, a sequence of observations $\{y_t\}_{t=\Delta,\ldots,n\Delta=T}$ is obtained. Because the unconditional mean of the OU process has an analytic form θ_0/θ_1, the simulations are started from this value. For processes that don't have closed form marginal distributions or known expected values, any reasonable (data driven) starting value can be chosen but a burn in sequence should be used to eliminate the transitory effects. Broze et al. (1998) show that for any fixed simulation step, indirect estimators of continuous time processes remain biased even in large samples. However, they show that by choosing an appropriately small simulation step the simulation bias becomes negligible.

The simulation bias analyzed by Broze et al. (1998) occurs if the model used for calibration differs from the true model. For example, if the true data is generated by (35), but the simulated data during calibration is generated by (43), the indirect estimators will contain simulation bias. If $\delta = \Delta$, then the asymptotic mapping for simulations, $\mu^\Delta(\theta)$, is given by the identity (compare (43) and (36))

$$\mu_0^\Delta(\theta) = \theta_0, \quad \mu_1^\Delta(\theta) = \theta_1, \quad \mu_2^\Delta(\theta) = \theta_2\ . \tag{44}$$

In just identified models $\hat{\theta}^\Delta$ solves $\tilde{\mu} = \mu^\Delta(\theta)$, that is $\hat{\theta}^\Delta = \tilde{\mu}$. Asymptotically, $\hat{\theta}^\Delta$ converges to a biased estimate

$$\hat{\theta}_0^\Delta \xrightarrow{p} \frac{\theta_0}{\theta_1\Delta}(1-e^{-\theta_1\Delta}), \quad \hat{\theta}_1^\Delta \xrightarrow{p} \frac{1}{\Delta}(1-e^{-\theta_1\Delta}), \quad \hat{\theta}_2^\Delta \xrightarrow{p} \theta_2\sqrt{\frac{1-e^{-2\theta_1\Delta}}{2\theta_1\Delta}}\ . \tag{45}$$

Only the bias that is present in *both* the auxiliary estimate and the mapping $\mu^\Delta(\theta)$ can be eliminated in indirect estimation. Here, the mapping $\mu^\Delta(\theta)$ is asymptotically unbiased, and as a consequence the discretization bias present in $\tilde{\mu}$ will not be eliminated: the simulation bias of the indirect estimator $\hat{\theta}^\Delta$ equals the discretization bias of the auxiliary estimator $\tilde{\mu}$. As k increases in $\delta = \Delta/k$, the discrete process approaches the continuous time process, and the simulation bias is reduced. Thus, the discretization bias represents an upper bound on the simulation bias. For $0 < k < 1$ Broze et al. (1998) derive the following analytical expressions for $\mu^\delta(\theta)$ in an OU model

$$\mu_0^\delta(\theta) = \frac{\theta_0}{\theta_1\Delta}(1-(1-\theta_1\delta)^{\Delta/\delta}), \quad \mu_1^\delta(\theta) = \frac{1}{\Delta}(1-(1-\theta_1\delta)^{\Delta/\delta}), \quad \mu_2^\delta(\theta) = \theta_2\sqrt{\frac{1-(1-\theta_1\delta)^{2\Delta/\delta}}{\theta_1\Delta(2-\theta_1\delta)}}\ . \tag{46}$$

As $\delta \to 0$, the asymptotic bias of $\mu^\delta(\theta)$ approaches the discretization bias of $\mu(\theta)$ in (40). Czellar and Ronchetti (2008) show that the inverse mapping $\hat{\theta}^\delta = (\mu^\delta)^{-1}(\tilde{\mu})$ implies a choice of δ such that the relative simulation bias is below a certain acceptable threshold. For example, noting that $\tilde{\mu}_1 \xrightarrow{p} \mu_1(\theta)$ in (40), the relative simulation bias of $\hat{\theta}_1^\delta$ can be found as a function of δ

$$\hat{\theta}_1^\delta \xrightarrow{p} \frac{1-e^{-\theta_1\delta}}{\delta} \quad \Rightarrow \quad \frac{asbias(\hat{\theta}_1^\delta)}{\theta_1} = \frac{1-e^{-\theta_1\delta}}{\delta\theta_1} - 1 \ . \tag{47}$$

Then, for known θ_1, δ can be chosen such that the relative simulation bias of the indirect estimator $\hat{\theta}_1^\delta$ is below a certain acceptable threshold.

3.5 Choice of S and its Impact on the L, A and M-type Estimators

The choice of S controls the asymptotic efficiency of simulation based estimators, and as $S \to \infty$, the simulation based estimators are asymptotically equivalent to the analytic estimators. Czellar and Zivot (2008) show that setting $S = 20$ gives a 95% asymptotic efficiency of the simulation-based estimators relative to the auxiliary estimators when the auxiliary model nests the structural model. For $S = 1$ the information contained in L,A, and M type simulations is identical, and therefore the L,A, and M type just identified estimators will give the same estimate for $S = 1$, but not for $S > 1$. In the discussion below $\mathrm{plim}_{S\to\infty} \tilde{\mu}_S^{\mathrm{i}}(\theta)$ represents uniform, as opposed to pointwise, convergence of the binding function. In addition, the binding function is assumed to be injective (Gouriéroux et al., 1993). To see the impact of varying S, note the following: [2]

(1.) The L-type binding function is based on a long data set of effective size $S \times n - 1$. The A-type binding function is based on S separate data sets of effective size $n - 1$, but it uses all the information at once and therefore the actual effective size of the data set is $S \times (n - 1)$. Thus the difference between the amounts of information used by the two estimators is equivalent to the information content of $(S - 1)$ observations. For $n = 1000$ this amounts to about 0.1%. Therefore, as $S \to \infty$, both the L and A type binding functions have features mimicking the asymptotic result: they are based on a data set of approximate size $S \times n$, so that $\mathrm{plim}_{S\to\infty} \tilde{\mu}_S^{\mathrm{i}}(\theta) = \mu(\theta) = \arg\max_\mu E_{F_\theta}[\tilde{f}(y, \mu)]$ for i = L, A. (As S becomes larger, the auxilary estimates based on the simulated samples will give a good approximation to the asymptotic result.) During the indirect inference calibration process, the distance between $\tilde{\mu}_S^{\mathrm{i}}(\theta)$ and $\tilde{\mu}$ is minimized by adjusting θ. In just identified models the first order condition for the analytical estimator is $\mu(\hat{\theta}^{\mathrm{IN}}) - \tilde{\mu} = 0$, so that $\hat{\theta}^{\mathrm{IN}} = \mu^{-1}(\tilde{\mu})$. Because $\mathrm{plim}_{S\to\infty} \tilde{\mu}_S^{\mathrm{i}}(\theta) = \mu(\theta)$, the just identified simulation based estimator will converge to the analytic one as $S \to \infty$, that is $\mathrm{plim}_{S\to\infty} \hat{\theta}^{\mathrm{Ii}} = \hat{\theta}^{\mathrm{IN}}$ for i = L, A in finite samples. Similarly, in just identified models $\mathrm{plim}_{S\to\infty} \hat{\theta}^{\mathrm{Ei1}} = \hat{\theta}^{\mathrm{EN1}}$ and $\mathrm{plim}_{S\to\infty} \hat{\theta}^{\mathrm{Ei2}} = \hat{\theta}^{\mathrm{EN2}}$ for i = L, A in finite samples. Finally, for $S \to \infty$ all these just identified estimators will be approximately equal in large samples, because $\hat{\theta}^{\mathrm{EN1}} = \hat{\theta}^{\mathrm{EN2}} = \hat{\theta}^{\mathrm{IN}}$ in large samples. However, in Section XX we show that this result does not hold in over-identified models.

(2.) The M-type binding function is based on simulated samples of the same size as the observed sample. The auxiliary estimates based on the simulated samples have the same finite

[2] also discuss the relative importance of discretization and finite sample bias at realistic parameter values, and mention the findings of Philips and Yu

sample properties as the estimate based on the observed sample. Thus, the M-type binding function is approximating the expected value of the auxiliary estimator in finite samples; that is, $\text{plim}_{S\to\infty} \tilde{\mu}_S^{\text{M}}(\theta) = E[\tilde{\mu}^{\text{M}}(\theta)] = E\tilde{\mu} = E[\arg\max_\mu \tilde{f}(y,\mu)]$. Denote $\overline{\mu}(\theta) = E[\tilde{\mu}^{\text{M}}(\theta)]$. During the indirect inference calibration process, the distance between these two auxiliary estimators with the same finite sample bias is minimized by adjusting θ. For $S \to \infty$, the first order condition in just identified models can be written as $\overline{\mu}(\hat{\theta}^{\text{IM}}) - \tilde{\mu} = 0$, so that $\text{plim}_{S\to\infty} \hat{\theta}^{\text{IM}} = \overline{\mu}^{-1}(\tilde{\mu})$. Gouriéroux et al. (2006) call this a "b_T-mean unbiased" estimator. b_T-mean unbiasedness does not in general imply mean unbiasedness, but if $\overline{\mu}(\theta)$ is linear in θ then $\hat{\theta}^{\text{IM}}$ will be mean unbiased, and significant bias reduction does not require a large S.

(3.) Because the binding function in (6) is defined in the asymptotic sense ($n \to \infty$), the M-type approximation should be treated differently than the L and A-type approximations. That is, $E[\arg\max_\mu \tilde{f}(y,\mu)] \neq \arg\max_\mu E_{F_\theta}[\tilde{f}(y,\mu)]$. The M-type estimator possesses some finite sample bias reduction qualities that have been demonstrated in just identified settings (Gouriéroux et al. (2000), Gouriéroux et al. (2006), Phillips and Yu (2007)), but we show in Section XX that they do not necessarily hold for over-identified models.

3.6 Use of Constraints

Sometimes one needs to ensure that the fine Euler discretization does not generate inadmissible values. Gallant and Tauchen (2004) discuss the need to "bullet-proof" the simulator so that the data generating process does not cause numerical exceptions on the computer, such as square roots of negative values and dividing by zero. For example the continuous time square root process (described in the Appendix) remains strictly positive for the proper choice of the parameters, but a sequence simulated from its fine Euler discretization will dip below zero with non-zero probability. In such cases, constraints can be imposed on the values generated by the process. Lord et al. (2006) give a recent survey of methods to avoid negative observations including absorption, reflection, truncation and some particular variants of these.

As noted in Section 2.1.1, the indirect estimation methods we consider rely on the assumption that the true data generating process is stationary. In the case of homoskedastic diffusion models with a finite volatility coefficient θ_2, such as the OU process, stationarity is ensured by the value of mean reversion parameter $\theta_1 > 0$. For $\theta_1 < 0$ the process becomes mean averting, that is, it diverges to infinity. During estimation, the optimizer searches across the admissible parameter space, and if the parameter is permitted to take on values that generate explosive sequences, the optimizer might run into numerical difficulties. There are several ways, to impose stability constraints on the model: constrain auxiliary model parameters only, structural model parameter only, or both sets of parameters.

Tauchen (1998) showed via an AR(1) model that the EMM objective function contains a built in penalty for explosive parameter values. His result is also applicable if the data is generated by

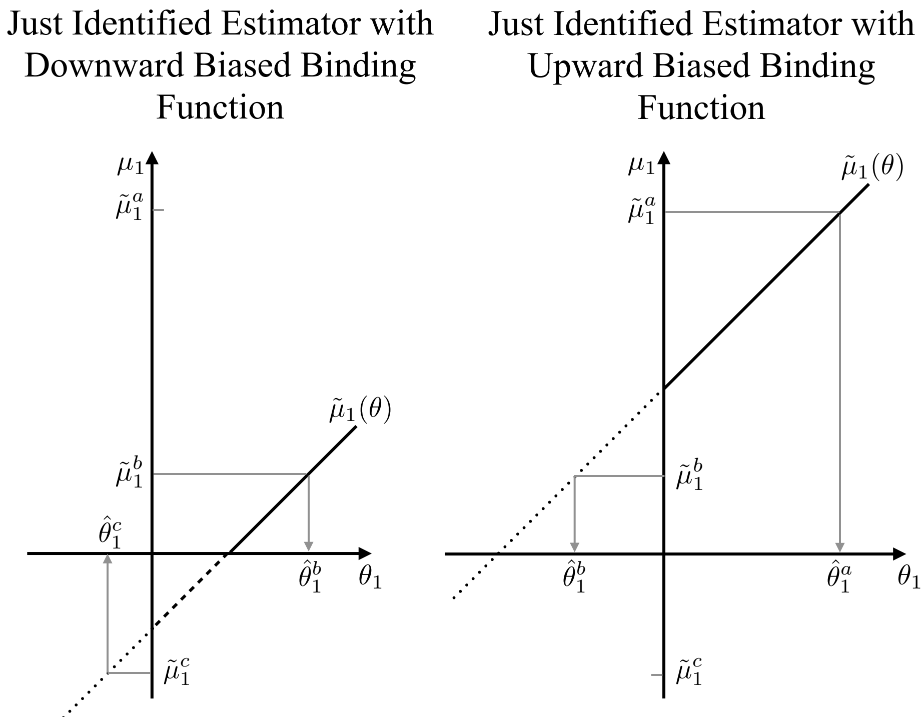


Figure 1: Mechanics of finding the just identified indirect estimate for (a) a downward biased binding function, (b) an upward biased binding function.

the OU process in (35): the expected value of the score depends on the following second moments

$$E_{F_\theta}[y_t^2] = \frac{\theta_2^2}{2\theta_1} + \frac{\theta_0^2}{\theta_1^2}\,. \qquad E_{F_\theta}[y_t y_{t-\Delta}] = \frac{\theta_2^2}{2\theta_1}e^{-\theta_1\Delta} + \frac{\theta_0^2}{\theta_1^2} \tag{48}$$

Note, that the presence of θ_1 in the denominator causes the second moment to quickly diverge to infinity as $\theta_1 \to 0$. Therefore, for $S \to \infty$ Tauchen (1998) suggested that constraining only the auxiliary estimate of an AR(1) model is sufficient to avoid explosive structural parameter estimates for EMM and indicated a similar result for II. To see the impact of constraints on the indirect estimators, note the following: [3]

(1) EL1 is based on $\tilde{g}_{Sn}(y_t(\theta); x_{t-\Delta}(\theta), \tilde{\mu})$ and IL is based on $\tilde{\mu} - \tilde{\mu}_{Sn}(\theta)$. As shown by Tauchen (1998), constraining $\tilde{\mu}$ will ensure nonexplosive solutions for $S \to \infty$ because $\hat{\theta}^{EL1} \to \tilde{\mu}$ and $\hat{\theta}^{IL} \to \tilde{\mu}$ if the auxiliary model nests the structural model. However, for finite S, the distribution of $\hat{\theta}^{EL1}$ and $\hat{\theta}^{IL}$ will include explosive values of θ, especially if $\tilde{\mu}$ is close to the boundary. To see this note that for finite S the simulated binding function is a random function. An unbiased auxiliary estimator implies that the binding function will fluctuate around a 45° line crossing the origin. For

[3] assume that $\theta > 0$ and $\mu > 0$ represents the stable region of the structural and auxiliary parameter space respectively

a downward biased auxiliary estimator there is a high probability of the simulated binding function falling below this 45° line (Figure 1a), but there will also be some realizations of the random binding function above this line (Figure 1b). In the former case, if the auxiliary estimate is not constrained, evaluating the inverted binding function with an explosive auxiliary estimate, such as $\tilde{\mu}_3$, will result in an explosive structural estimate $\hat{\theta}_3$. In the latter case, evaluating the inverted binding function with an auxiliary estimate close to the boundary, such as $\tilde{\mu}_2$, will result in an explosive structural estimate $\hat{\theta}_2$. If explosive structural parameter values are attained, the machine arithmetic may run into trouble during the calibration process. Because constraining the auxiliary estimate only may not be sufficient to prevent these numerical problems, constraining the structural parameter space might be necessary (whenever it is feasible).

(2) IM is based on $\tilde{\mu} - \tilde{\mu}_n^S(\theta)$. If the auxiliary estimate $\tilde{\mu}$ is on the boundary, the IM estimator's bias correcting property implies that the unconstrained IM estimate $\hat{\theta}^{IM}$ will be in the nonstationary region with very high probability. To see this note that an upward biased auxiliary estimator[4] implies that the mean simulated binding function is not a 45° line crossing the origin, but it is shifted up (Figure 1b). Therefore evaluating the inverted binding function with an auxiliary estimate close to the boundary, such as $\tilde{\mu}_2$, will result in an explosive $\hat{\theta}_2$ estimate. In this case, the need to constrain $\tilde{\mu}$ will guarantee explosive $\hat{\theta}^{IM}$ even for large S because $S \to \infty \Rightarrow \hat{\theta}^{IM} \to \theta < \tilde{\mu}$ if the auxiliary estimator is upward biased. Thus, constraining the auxiliary parameter is not enough to avoid simulations with explosive values; θ has to be constrained too.

(3) The E2 estimator can be considered a hybrid estimator consisting of two steps. In the first step the simulation based binding function $\tilde{\mu}_S(\theta)$ is calculated just as in II. In II this simulated binding function is directly compared to the auxiliary estimate $\tilde{\mu}$. In the E2 estimator the mean score evaluated with $\tilde{\mu}_S(\theta)$ 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 E2 objective function can be attributed to the simulated binding function $\tilde{\mu}_S(\theta)$, just like in the case for the II type objective function. Therefore the E2 and II type estimates will be close and their objective functions will look similar in finite samples.

EM2 is based on $\tilde{g}_n^S(y_t; x_{t-\Delta}, \tilde{\mu}(\theta))$. If the observed data y is such that $\tilde{\mu}$ is explosive, $\tilde{\mu}(\theta)$ will tend be explosive too because $\tilde{\mu}(\theta)$ and $\tilde{\mu}$ contain the same bias. Constraining $\tilde{\mu}$ does not help because the objective function does not depend on $\tilde{\mu}$. It depends on the explosive sequence that generated $\tilde{\mu}$. Again, the distribution of $\hat{\theta}^{EM2}$ will include explosive values of θ, especially if $\tilde{\mu}$ is close to the boundary. This exacerbates the problem of spurious results at false minima of the objective function, illustrated in Section 4.2.1. As in the case of II, to avoid explosive EN2 parameter estimates, the structural parameter space has to be constrained. Gouriéroux and Monfort (1996) showed that in just identified models EN1, EN2 and IN should return identical results, but if the auxiliary estimator is constrained, the EN2 estimates will differ because the auxiliary constraint affects the EN1 and IN estimators, but it has no impact on the EN2 estimator.

(4) So far in this section, the binding function was assumed to be unconstrained. Calzolari et al.

[4] For example, the mean reversion parameter estimate of an OU process is upward biased in finite samples.

(2004) proposed a method using multipliers to constrain the auxiliary parameter space including the binding function. They analyzed indirect estimators with inequality constraints on the auxiliary model, and include the Lagrange multipliers in the definition of the binding function. This approach ensures that the derivative of the binding function (including the multipliers) with respect to the structural parameter has full column rank. If the auxiliary parameter space were constrained by re-parameterization, no multipliers would be available to augment the binding function, and a flat area could be introduced in the objective function. That is, changing θ would not change $\tilde{\mu}(\theta)$, and the simulated binding function would not be invertible in a just identified setting. This can be seen in Figure 1a: if the auxiliary estimate $\tilde{\mu}$ lied on the boundary, a parametrically constrained binding function would not be invertible. In addition, Figure 1b illustrates that constraining the binding function cannot prevent explosive structural parameters. Thus constraining the binding function does not benefit the stability of the model or calculations. Instead, constraining both the auxiliary estimate $\tilde{\mu}$, and the structural θ parameter space seem to be sufficient to avoid inadmissible results.

Those favoring constraints on the auxiliary model note that they may be easier to implement than the constraints on the structural model. However, observations from diffusion models can be generated by a fine Euler discretization, for which the boundary of the stable parameter space is, in general, the same as for the auxiliary model. For simple discrete models of the interest rate the constraints are known: Broze et al. (1995) study the ergodic properties of common discrete time interest rate models and show conditions for their stationarity. Therefore, the constraints can be directly imposed on the simulations.

3.7 Issues Related to Over-Identified Indirect Estimators of Continuous Time Processes

Over-identification takes place if the number of auxiliary parameters exceeds the number of structural ones. In this case the estimator is not independent of the weighting matrix. Over-identification enables the researcher to test wether the model is correctly specified by evaluating the objective function with the structural parameter estimates (Gallant and Tauchen, 1996). The imbalance between the number of auxiliary and structural parameters occurs either because a large number of auxiliary parameters is used to approximate the structural model, or because some of the structural parameters are assumed to be known. For example, on the one hand Gallant and Tauchen (1996) approximate the structural model with a highly parameterized auxiliary score generator; on the other hand, Duffee and Stanton (2008) achieve over-identification by estimating only a subset of the structural parameters. In the latter case some structural parameters are assumed to be fixed at their true values, and at the same time all auxiliary parameters are estimated.

Over-identification arising from model based restrictions has been successfully used by Chan et al. (1992) and Broze et al. (1995) to compare specific interest rate models nested within a general model. For example Chan et al. (1992) used GMM to estimate a wide variety of short term interest rate models. Their approach exploits the fact that many term structure models can be nested within the following stochastic differential equation

$$dy = (\theta_0 - \theta_1 y)dt + \theta_2 y^{\theta_3} dW \ , \quad dW \sim \text{iid N}(0, dt) \ . \tag{49}$$

The various models are obtained by placing appropriate restrictions on the four θ parameters. For example for the OU model $\theta_3 = 0$, and for the CIR model $\theta_3 = 0.5$. Chan et al. (1992) use four moment conditions to estimate the discrete time approximation to (49)

$$y_t = \theta_0 + (1 - \theta_1)y_{t-1} + \theta_2 y_{t-1}^{\theta_3} \epsilon_t \ , \quad \epsilon_t \sim \text{iid N}(0, 1) \tag{50}$$

by GMM. In addition to (50) they also estimate the discrete time versions of the nested models. The GMM estimator is just identified for model (50), because the number of moment conditions and the number of parameters are equal, but it becomes over-identified if any restrictions are imposed on the parameter space, such as in the case of the discrete time OU and CIR models.

The logic is similar in the case of indirect estimators: if (50) represents both the structural and the auxiliary models, the indirect estimator is just identified, but if (50) only represents the auxiliary model, and the structural model is a discrete time OU or CIR model, the indirect estimator becomes over-identified. In the latter case there are restrictions placed on the structural parameter space. While the value of the just-identified objective function is equal to zero at the parameter estimates, the value of the over-identified one is an asymptotically χ^2 distributed random variable. If one or more of the structural parameters are being held fixed at their true values and the model is correctly specified, asymptotic theory indicates that the estimates of the remaining parameters will also converge to their true values. However, we show in Section XX that in finite samples over-identification introduces a bias caused by the nonlinearity of the auxiliary statistics.

4 EMM and II Estimation of an OU Process

Because it has an exact analytical solution, the OU process represents the closest connection between a discrete time AR(1) model and a diffusion, and we use it to address some of DS's criticism. In this section, we apply the EMM and II estimators to an OU model and evaluate their finite sample performance in a Monte Carlo study.

4.1 Model Setup

The analytically tractable OU process gives us the opportunity to directly compare the performance of indirect estimators to the benchmark MLE. The true data generating process is an OU process of the form given in (34). Weekly observations ($\Delta = 1/50$) of annualized interest rates are generated from its exact solution in (35) for two sets of true parameters. The first one, $\theta = (0.01, 0.1, 0.1)$, also used by Phillips and Yu (2007), is based on the following empirical findings (see for example Chan et al. (1992)): (1) the mean annualized interest rate θ_0/θ_1 on short term government bonds has not exceeded 10% in the last 20 years, (2) interest rates are highly persistent and the annual mean reversion rate θ_1 is around 10%, (3) while the volatility θ_2 of interest rates is proportional to their level, the annualized volatility is below 10%. The value $\theta_1 = 0.1$ implies that the half life of a shock is approximately seven years.

The second set of parameters is derived from the AR(1) parameterization of DS $\theta^{DS} = (0, 0.9868, 1)$ to achieve comparability with their results. DS assumed weekly observations, and used $\Delta = 1$ to represent the observation interval. Therefore their parameter values have weekly as opposed to annual interpretation in the following AR(1) process

$$y_t = \theta_0^{DS} + \theta_1^{DS} y_{t-\Delta} + \theta_2^{DS} \epsilon_t, \quad \epsilon_t \sim \text{iid N}(0,1) \;, \tag{51}$$

The AR(1) parameters can be mapped to the OU parameters by inverting the following correspondence between (35) and (51)

$$\theta_0^{DS} = \frac{\theta_0}{\theta_1}(1 - e^{-\theta_1 \Delta}), \quad \theta_1^{DS} = e^{-\theta_1 \Delta}, \quad \theta_2^{DS} = \theta_2 \sqrt{\frac{1 - e^{-2\theta_1 \Delta}}{2\theta_1}} \;. \tag{52}$$

Thus, we obtain the second set of parameters from the transformation: $\theta_1 = -\Delta^{-1} \log \theta_1^{DS} = -50 \log 0.9868 = 0.66$, $\theta_0 = \theta_0^{DS} \theta_1 / (1 - \theta_1^{DS}) = 0 \cdot 0.66/(1 - 0.9868) = 0$, $\theta_2 = \theta_2^{DS} \sqrt{2\theta_1/(1 - (\theta_1^{DS})^2)} = 1\sqrt{2 \cdot 0.66/(1 - 0.9868^2)} = 7.071$. Here, θ_1 can be interpreted as the annualized mean reversion toward the long run mean of zero, and θ_2 as the annualized volatility of the OU process. The value $\theta_1 = 0.66$ implies that the half life of a shock is approximately one year. We consider a time horizon of 20 years, which corresponds to 1000 observations.

The auxiliary model is a crude Euler discretization of the OU process (36), and the auxiliary score and estimator are given in Section 3.2. We use their analytical forms (38) and (39) for calculating the binding function in the calibration process. Following the recommendation of Tauchen (1998), we constrain the maximum likelihood estimate of the auxiliary parameter $\tilde{\mu}$ from the "observed data". That is, we impose $\tilde{\mu}_1 > 0$ and $\tilde{\mu}_2 > 0$ in the optimizer in the maximum likelihood estimator. The structural model can be constrained or unconstrained, and we analyze the impact of constraints by comparing the two results.

We consider two just-identified models: model (3×3), where all three parameters are estimated in both the structural and auxiliary models; and model (1×1), where only the mean-reversion parameter is estimated in both the structural and auxiliary models; the constant and the volatility parameter are held fixed at their true values. In addition we consider an over-identified model (1×3), where only the mean reversion parameter is estimated in the structural model, but all three parameters are estimated in the auxiliary model.

4.2 Objective Functions and Confidence Intervals

In this section we illustrate the objective functions and confidence intervals of the indirect estimators. The results and diagrams are based on a representative sample with $\theta = (0.01, 0.1, 0.1), T = 20$ and simulations from the exact solution of the OU process (35) with $S = 20, \Delta = 1/50$, and

	$QMLE$	$EN1$	$EL1$	$EA1$	
OU parameter estimates					
θ_0	0.0159	0.0159	0.0153	0.0135	
θ_1	0.2645	0.2652	0.2344	0.2105	
θ_2	0.1033	0.1035	0.1035	0.1034	
		$EN2$	$EL2$	$EA2$	$EM2$
θ_0		0.0159	0.0153	0.0136	-0.0102
θ_1		0.2649	0.2341	0.2103	-0.0685
θ_2		0.1035	0.1035	0.1034	0.1035
		IN	IL	IA	IM
θ_0		0.0159	0.0153	0.0135	-0.0102
θ_1		0.2652	0.2344	0.2106	-0.0684
θ_2		0.1035	0.1035	0.1034	0.1035

Table 1: Unconstrained parameter estimates of true $\theta = (0.01, 0.1, 0.1)$, $T = 20$ in linear just-identified model of OU process. Simulations are from exact solution with $S = 20$, and $\Delta = 1/50$.

$n = 1000$.

4.2.1 Shapes of the Objective Functions for 3×3 Just-Identified Models

Figure 3 illustrate the shapes of objective functions of all considered estimators for just identified models in an unconstrained parameter space. For all diagrams, θ_2 is held fixed at its estimate. The corresponding parameter estimates are displayed in Table 1, and denoted with a green dot in the diagrams (QMLE represents the naive auxiliary estimates). Looking at Table 1, I observe what Phillips and Yu (2005b) and Phillips and Yu (2005a) also indicate: θ_2 can be estimated quite precisely, but the estimates of θ_1 suffer from strong finite sample bias in highly persistent processes. Also, the M-type estimates of θ_1 have smaller finite sample bias than the other estimates.

For the AR(1) model, DS and Tauchen (1998) pointed out that some of the moments in the EN1 estimator are scaled by the population variance, whereas the corresponding moments in the IN and EN2 estimators are scaled by the sample variance. The population variance grows rapidly as the process becomes more persistent, and it causes the objective function for the EN1 estimator to peak sharply for θ near the stability boundary. Because the sample variance is being held fixed during calibration, the EN2 and IN objective functions remain roughly symmetric around their minimum. While the point estimates are approximately equal across the E1, E2 and II estimators, the different objective function shapes imply that the performance of the E1 estimator in tests and the associated confidence intervals will be different from those of the E2 and II estimators.

In general, the A and L-type criterion functions most closely match the corresponding N-type criterion functions which suggests that the A and L-type estimators have similar finite sample properties as the N-type estimators. However, the M-type criterion functions are shifted toward the non-stationary region which suggests that the M-type estimators have different finite sample properties than the other estimators. If the structural parameter space is unconstrained, the optimizer may evaluate the E2 and II type objective function with the parameter in the unstable region. While the unconditional mean of explosive processes is not defined, the starting value of simulation is still set to its implied value θ_0/θ_1 which is infinity for $\theta_1 = 0$. As the diagrams show, the II type objective function remains roughly symmetric around its minimum, but the E2 type will have false minima in the non-stable region.

The false minima of E2 are caused by the computer running out of significant digits, and as a result giving imprecise estimates of the simulated binding function. For an explosive value of the slope parameter, the generated observations quickly diverge towards infinity. The least squares estimator of the mean reversion parameter $\mu_1(\theta)$ in (39), is a *ratio* where both the numerator and the denominator are of similar magnitude. Therefore this estimate, while biased, is relatively close to θ_1, or 1. However, the estimator of the intercept $\mu_0(\theta)$ is the *difference* between two terms of the same magnitude, and the result can be very large in absolute terms. This makes the estimate of the intercept extremely imprecise, and the error is propagated to the OLS estimate of $\mu_2(\theta)$.

The wild fluctuation of the E2 objective function is caused by the occasional large value of $\mu_2(\theta)$ in the denominator of the score (38): if the $\mu_2(\theta)$ estimate is large, all terms in the score will be close to zero. While the II objective function is also affected by the estimation error of the binding function, it doesn't give false minima because none of the terms of the binding function appear in the denominator. In fact, all estimators are subject to a large error in explosive series, but only E2 will fluctuate violently as a result of $\mu_2(\theta)$ in the denominator. However, as indicated in Section 3.6, the structural parameter space has to be constrained in order to avoid explosive E1, E2 and II estimates in finite samples, and this constraint also eliminates the potential for the erratic behavior of the E2 estimator.

4.2.2 Confidence Intervals for the Mean-Reversion Parameter for 3×3 Just-Identified Models

Figure 4 illustrates the LR-type statistics (31), (32) and their simulation based counterparts for testing $H_0 : \theta_1 = \theta_1^0$ as functions of θ_1^0 for the just identified estimators. The point estimates for θ_1 and 95% confidence intervals obtained by inverting the LR statistics are displayed in the title of each diagram. The 95% confidence intervals contain values of θ_1^0 such that the value of the LR statistic lies below the 95% quantile of the chi-square distribution with 1 degree of freedom.

While the estimates are very similar across estimators using the same amount of information, the LR statistics for the score based (E1) and binding function based (E2, II) estimators can have very different shapes. For the highly persistent cases, LR^{E1} is highly asymmetric due to the scaling of some moments by the population variance. It is extremely flat for θ_1^0 values above $\hat{\theta}_1^{\mathrm{E1}}$, and peaks sharply as θ_1^0 approaches zero. In contrast, the LR functions for the E2 and II estimators are almost identical and are roughly symmetric in θ_1^0, because they are scaled by the variance of the observed sample, which is constant for any θ_1^0.

The E1 confidence interval covers a wide range of θ above the point estimate, but only little of the range below the point estimate. While the shape of the criterion function puts a high penalty for θ close to the boundary of stationarity, it causes point estimates above the true value to be rejected with high probability. Because the E1 estimator delivers very limited finite sample bias correction, most of the point estimates will fall above the true value, thus causing a high rejection rate. Also, the optimizer of the E1 estimator often converges slower than the optimizer of the II

Average estimation time										
$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution										
True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Unconstrained estimation.										
0.03	5.15	5.92	0.09	4.16	5.79	11.40	0.01	4.03	5.65	11.06

Table 2: Average estimation time in seconds.

estimator if the starting value falls far above the true value and the search is conducted along the flat side of the E1 objective function. This is apparent from Table 2, which shows estimation times with the Nelder-Mead optimizer. From the analytic estimators, IN is the fastest, followed by EN1, and EN2. The last two positions flip in the ranking of the L and A type estimators, and here the E1 estimator is the slowest. In addition, gradient and Hessian based optimizers might converge slowly for E1 because the shape of the E1 objective function is non-quadratic. The M-type estimators are the slowest overall, but only less then three times as slow as the L-type estimators.

4.2.3 Objective Functions and Confidence Intervals for 1×3 Over-Identified Models

In order to illustrate the impact of over-identification on the objective functions and confidence intervals of the indirect estimators, the θ_0 and θ_2 parameters are restricted to their true values, and only θ_1 is being estimated in the structural model.

Figure 5 illustrates the LR-type statistics (31), (32) and their simulation based counterparts for testing $H_0 : \theta_1 = \theta_1^0$ as functions of θ_1^0 for the over-identified estimators. The corresponding parameter estimates are displayed in Table 3, and denoted with a green dot in the diagrams (QMLE represents the naive auxiliary estimates). Here the point estimates differ across estimators using the same amount of information, and similar to the results reported in DS, LR^{E1} has its minimum the farthest away from the true value $\theta_1 = 0.1$. Based on the diagrams, the EN1 estimator has the most negatives going against it: the largest bias, widest asymmetric confidence interval which does not contain the true parameter value. On the other hand, EM2 has the smallest bias, narrowest confidence interval which contains the true parameter value.

OU parameter estimates of θ_1				
$QMLE$	$EN1$	$EL1$	$EA1$	
0.2645	0.2490	0.2241	0.1912	
	$EN2$	$EL2$	$EA2$	$EM2$
	0.2325	0.2087	0.1892	0.0905
	IN	IL	IA	IM
	0.2336	0.2095	0.1900	0.0904

Table 3: Constrained θ_1 parameter estimates for true $\theta = (0.01, 0.1, 0.1)$, $T = 20$ in linear over-identified model of OU process. Simulations are from exact solution with $S = 20$, and $\Delta = 1/50$.

The EN objective function of the OU process depends on the population moments (48) and therefore its shape is asymmetric: it quickly increases as $\theta_1 \to 0$ but is relatively flat for $\theta_1 >> 0$. Figure 6 shows the shape of the just identified EN objective function in two cases: (1.) when θ_2 is being held fixed at $\hat{\theta}_2^{EN}$ (the just identified estimate), and (2.) when θ_2 is being held fixed at $\theta_2 = 0.1$, the true value. It is clear that the large deviation of the over-identified EN estimate from the just identified one is caused by holding θ_2 fixed at a value that is not optimal for the current set of observations.

Figure 7 shows the third term of the expected value of the score vector (derivative of likelihood with respect to μ_2). The green dot at level $E_\theta[\tilde{g}(y_t(\theta); y_{t-\Delta}(\theta), \tilde{\mu})] = 0$ represents the just identified estimate providing the best fit given the observed data. In the over-identified model θ_0 and θ_2 are being held fixed and θ_1 has to compensate for those restrictions when minimizing the objective function. The second and third terms of $E_\theta[\tilde{g}(y_t(\theta); y_{t-\Delta}(\theta), \tilde{\mu})]$ depend on the population moments (48) which explode to infinity as $\theta_1 \to 0$. Because of the large penalty of moving towards 0, θ_1 will tend to adjust by moving away from 0 as illustrated by the gray dot in Figure 7. This tendency of θ_1 to adjust by moving away from 0 will cause the over-identified EN1 estimator to have a larger, upward bias than the just identified estimator.

4.3 Monte Carlo Results

In this section, we conduct a Monte Carlo study of the bias of the indirect estimators. Unless otherwise stated, the results and diagrams are based on representative samples with $\theta = (0.01, 0.1, 0.1), T = 20$ and simulations from the exact solution of the OU process (35) with $S = 20, \Delta = 1/50$, and $n = 1000$.

4.3.1 Bias and RMSE in 3×3 Just-Identified Models

Figures 8 illustrate the distribution of estimates based on 1000 Monte Carlo simulations for 3×3 just-identified models in an unconstrained parameter space. The results are shown for the constrained auxiliary QMLE, EXAct (constrained) structural MLE, unconstrained auxiliary OLS, and the indirect estimators. The indirect estimates of θ_0 and θ_1 are distributed approximately equally to the auxiliary estimates, except for the M-type estimators, which show significant bias correction, and have lower RMSE compared to the other ones. The estimates of θ_2 are essentially unbiased for all indirect estimators.

The upper portion of Table 4 displays the mean, mean bias and root mean squared error (RMSE) for the estimators. As Gouriéroux and Monfort (1996) indicate, the E1, E2, and II estimators of the same type (that is N, L, A, and M type if available) give nearly identical results in this just identified setting. Because we use the exact solution for data generation, there is no simulation bias in the estimates. The discretization bias of the QMLE is apparent from its downward deviation from the EXAct solution (-0.5%), but the finite sample bias (+270%) clearly dominates at this parameterization. The analytic (N-type) estimators perfectly correct the discretization bias, but they have no effect on the finite sample bias. The L and A type estimators slightly correct the finite sample bias, but remain very close to the analytic estimators. The M-type estimator delivers the largest finite sample bias reduction and the smallest RMSE. However, it does not completely eliminate the bias. Gouriéroux et al. (2006) find that setting $S = 250$ in their dynamic panel models, does essentially eliminate the bias, but the diagrams of Gouriéroux et al. (2000) still show some bias at $S = 15000$ (see the impact of S on the distribution of the L and M-type estimators in Figure 15). It is important to remember the M-type estimator is "b_T mean unbiased", but not necessarily mean unbiased.

The non-mean unbiasedness is caused by the non-linearity of the binding function for $\theta < 0$. Figures 16 and 17 illustrate this point. The N-type binding function is essentially unbiased around the true parameter values because the asymptotic discretization bias is negligible. Thus the distributions of the N, L and A-type estimators are similar to the distribution of the auxiliary estimator, as indicated in the top right panel of Figure 17. The bottom left panel of Figure 17 shows that if the M-type binding function contained a uniform bias, $\hat{\theta}_1^{IM}$ and $\hat{\theta}_1^{EM2}$ would be mean unbiased. However, Figure 16 indicates that the super-consistency of M-type binding function for $\theta \leq 0$ causes the M-type binding function to be non-linear in this range of θ, and therefore the M-type estimators remain biased, as shown in the bottom right panel of Figure 17.

In the case of unconstrained estimation, the EL1 and EA1 estimators deviate (interestingly, upward) from their same simulation length E2 and II counterparts, but constraining the structural parameter space essentially eliminates the difference, indicating that the deviation of the E1 estimator is caused by the parameter entering the non-stationary region. For the alternative parameterization $\theta = (0, 0.66, 7.071)$ in the lower portion of Table 4 the ranking of the estimators based on bias and RMSE remains the same as above.

4.3.2 Bias and RMSE in 1×1 Just-Identified Models

Gouriéroux et al. (2000) point out, that the median unbiased estimator of Andrews (1993) is an application of the IM estimator to a just-identified model with one parameter, when the mean is replaced by the median in the simulated binding function (MEM2 and MIM). In this section I investigate the relationship between the M-type estimator and these median unbiased estimators. Figure 18 illustrates the distribution of estimates based on 1000 Monte Carlo simulations for 1×1 just-identified models in an unconstrained parameter space. The results are shown for the constrained auxiliary QMLE, EXAct (constrained) structural MLE, unconstrained auxiliary OLS, and the indirect estimators, including two median unbiased estimators (MEM2 and MIM).

Because the distribution of the mean reversion parameter μ_1 in the auxiliary model has a positive skew, the M-type simulated binding function is upward biased compared to a typical (median) auxiliary estimate. Therefore the distance minimization between the auxiliary estimate $\tilde{\mu}$ and the simulated binding function $\tilde{\mu}_S^M(\theta)$ results in a larger shift of the distribution of the structural estimates $\hat{\theta}$ than would be necessary to get a median unbiased estimator. The diagram shows that the MEM2 and MIM estimators are essentially *median* unbiased, while the M-type estimators are not quite *mean* unbiased, implying that the median based estimators are more robust against the non-linearities of the binding function than the M-type estimator. The impact of the nonlinearity of the binding function for highly persistent processes becomes even more apparent by comparing the diagrams in Figure 18. The lower one displays the results for $\theta = (0, 0.66, 7.071)$, and the mean bias of the M-type estimators at this parameterization almost disappears.

4.3.3 Bias and RMSE in 1×3 Over-Identified Models

Figure 9 illustrates the distribution of estimates based on 1000 Monte Carlo simulations for 1×3 over-identified models in a constrained structural parameter space. The diagrams are based on representative samples with $\theta = (0, 0.66, 7.071), T = 20$ and simulations from the exact solution of the OU process (35) with $S = 20, \Delta = 1/50$. Throughout, θ_0 and θ_2 are being held fixed at their true values. In addition to the diagram, the results are also given in Table 8 for the constrained auxiliary QMLE, EXAct (constrained) structural MLE, unconstrained structural OLS, and the indirect estimators.

As apparent from the diagram, the distribution of the over-identified E1 type estimators clearly differs from the just-identified ones. As indicated in Section 4.2.3, this behavior can be attributed to the sensitivity of the E1 objective function to θ_2 when it is being held fixed at a value that is suboptimal for the current set of observations.

The over-identification restriction, $\theta_0 = 0$, implies that $\mu_0(\theta) = 0$ in (40). Correspondingly, $\mu_1(\theta)$ and therefore $\hat{\theta}_1^{IN}$ will converge towards the OLS estimate of θ_1 when the structural model does not contain a constant term as opposed to the QMLE. This is apparent in the results in Table 8 and can be verified by substituting $\mu_0(\theta) = 0$ in the II moment function

$$\widetilde{H}_n(\tilde{\mu} - \mu(\theta)) = \frac{\Delta}{\tilde{\mu}_2^2} \begin{pmatrix} -[\overline{y_t} - (1 - \mu_1(\theta))\overline{y_{t-\Delta}} - \mu_0(\theta)] \\ \left[(\overline{y_t y_{t-\Delta}} - \mu_0(\theta)\overline{y_{t-\Delta}})/\overline{y_{t-\Delta}^2} - (1 - \mu_1(\theta))\right] \overline{y_{t-\Delta}^2} \\ -\frac{2}{\Delta}(\tilde{\mu}_2 - \mu_2(\theta)) \end{pmatrix}. \tag{53}$$

It follows from (40) that the asymptotic discretization bias of θ_1^{OLS} is about 0.005 at this parameterization. Because the constraint on the EXAct structural MLE is never binding at this parameterization, the difference between θ_1^{EXA} and θ_1^{OLS} gives the discretization bias measured in the Monte Carlo study. While the IN estimator slightly over-corrects the discretization bias, the EN2 estimator does not fully correct it. However, the discretization bias is negligible compared to the finite sample bias, which leaves all N, L and A-type estimators biased. In general, the N, L and A-type II and E2 estimators outperform the E1 estimators, but more importantly, the M-type estimators seem to overcorrect the finite sample bias.

4.3.4 Impact of Over-Identification on the Bias of the M-type Estimators

As observed in the previous section, the over-identified M-type estimator over corrects the bias for the indicated parameterization. This is surprising given the finite sample bias correcting property of the M-type estimator in just identified models: if the binding function is linear, the M-type estimator is mean unbiased, otherwise it is b_T-mean unbiased (Gouriéroux et al., 2000). As described in Chapter 3, in the just-identified setting the M-type estimator can be viewed as an inverted binding function, and the weighting matrix does not play a role. However, in over-identified models the weighting matrix is essential in forming linear combinations of the indirect moment functions, and it is these linear combinations that are solved by the estimator.

In this section, we explore the effect of the weighting matrix and the nonlinearity of the moments on the M-type estimates if over-identification restrictions are imposed on the structural parameters. Column F-II in Table 2 of Duffee and Stanton (2008) shows that the IM estimates tend to exceed the true AR(1) parameter value in all cases they considered. I confirmed a similar result for the OU process: note that in the lower part of Table 8 the IM type estimator seems to overcorrect for the $\theta = (0, 0.66, 7.071)$ parameterization. That is, the IM bias in the over-identified model is in the opposite direction (down) compared to the finite sample bias (up). I contend that the finite sample bias correcting property of the IM estimator does not carry over from just identified models to over-identified ones.

To see if the observed bias is caused by insufficient simulation size, I first looked at the impact of S on the bias of the IL and IM estimator of θ_1 in both the just-identified and the over-identified setting. Figure 10 shows the distribution of the auxiliary, the IL, and the IM estimators in a just-identified setting, and Figure 11 shows those same estimators in an over-identified setting. First look at the just identified case. For $S = 1$, the IL and IM estimators are identical, but for S increasing the IL estimator's distribution approaches the distribution of the auxiliary estimator, and the IM estimator's distribution shifts so that it's mean approaches the true parameter value. Given the analysis in Chapter 3, and the fact that the discretization bias is negligible in comparison with the finite sample bias, this is the expected outcome:

(1) For $S \to \infty$ the IL estimator corrects the negligible discretization bias, but not the finite sample bias of the auxiliary estimator, and therefore the distributions of the IL and auxiliary estimators are very similar for large S.

(2) For $S \to \infty$ the IM estimator corrects both biases of the auxiliary estimator, and therefore the mean of the IM estimator is shifted by approximately the amount of the total bias of the auxiliary estimator (exactly if the binding function is linear (Gouriéroux et al., 2000)).

The diagrams for the over-identified case in Figure 11 do not follow this logic: for increasing S, the means of the IL and the IM estimators seem to converge to values below the corresponding values in the just identified setting; that is, both seem to be downward biased compared to their just-identified counterparts. In the following, I will focus on the IM estimator, but later I show that the arguments are valid for the IL estimator too, albeit the impact on the IL estimator is somewhat milder.

The bias of the IM estimator in the over-identified model is caused by two sources: (1.) the interaction between the terms in the weighting matrix and the simulated binding function, and (2.) the relationship between the structural parameter θ_1 in $\theta = (\theta_0, \theta_1, \theta_2)$ and the auxiliary estimates $\tilde{\mu}(\theta)$ in the simulated binding function during calibration. During calibration, $\tilde{\mu}^{\mathrm{M}}_{S0}(\theta)$ and $\tilde{\mu}^{\mathrm{M}}_{S1}(\theta)$

are jointly estimated by OLS, and their bias influences the bias of the structural estimates in over-identified models. In the Appendix, we show that the approximate biases of the auxiliary estimators are

$$E[\tilde{\mu}_1(\theta)] - \theta_1 \approx \frac{3e^{-\theta_1 \Delta} + 1}{T} + \frac{1}{\Delta}(1 - e^{-\theta_1 \Delta}) - \theta_1 \ , \tag{54}$$

and

$$E[\tilde{\mu}_0(\theta)] - \theta_0 \approx \theta_0 \left(\frac{3e^{-\theta_1 \Delta} + 1}{\theta_1 T} + \frac{1 - e^{-\theta_1 \Delta}}{\theta_1 \Delta} \right) - \theta_0 \ . \tag{55}$$

For small positive values of θ_1 the constant term in the simulated auxiliary estimate, $\tilde{\mu}_0(\theta)$, will be biased away from zero; that is, the estimates of positive constants will be biased up and those of negative constants will be biased down in finite samples.

While the true value of the constant in the DS parameterization is fixed at $\theta_0 = 0$, the auxiliary estimate $\tilde{\mu}_0$ can be positive or negative. In the just identified model both θ_0 and θ_1 can be optimally chosen to set the distance between the vector of auxiliary estimates $\tilde{\mu}$ and the binding function $\tilde{\mu}(\theta)$ equal to zero. In the over-identified model, only θ_1 will adjust to minimize the objective function, and it will contribute to the reduction of $d_0 = \tilde{\mu}_0 - \tilde{\mu}_0(\theta)$ by deviating from the just identified θ_1 estimate. The amount of deviation will be determined by the weighting matrix and the sensitivity of $\tilde{\mu}(\theta)$ to θ_1.

The dependence of $\tilde{\mu}_0(\theta)$ on θ_0 and θ_1 for true $\theta = (0.01, 0.1, 0.1)$, and $n = 1000$ data points generated with $\Delta = 1/50$ from a particular seed is shown in Figure 12. The explosive behavior of $\tilde{\mu}_0(\theta)$ as $\theta_1 \to 0$ can also be deduced from (60). As the contour-plot in Figure 12 shows, there is a strengthening relationship between $\tilde{\mu}_0(\theta)$ and θ_1 when the latter is approaching zero. Depending on the random draw, the distance d_0 might be reduced by moving from the just identified θ_1 estimate towards zero or away from zero. The larger sensitivity of $\tilde{\mu}_0(\theta)$ to θ_1 in the neighborhood of zero implies a larger benefit of a deviation of θ_1 towards zero than a deviation of θ_1 away from zero. Therefore on balance the deviations towards zero will dominate, which explains the downward bias of the over-identified IM estimator.

The nonlinear behavior of the binding function for $\theta_1 \to 0$ will cause the objective function to become steeper and correspondingly the confidence intervals to be shorter. Because asymptotically the finite sample bias term in (61) goes to zero, and the discretization bias is declining for $\theta_1 \to 0$, the IL binding function and estimator will be less affected by these nonlinearities (see Figures 13 and 19).

A comparison of Tables 8 and 4 reveals that the over-identified E2 and II type estimates of θ_1 have lower bias and RMSE for the $\theta = (0.01, 0.1, 0.1)$ parameterization than the just-identified ones. This can be explained by the bias of over-identified estimators toward zero described for the M-type estimator in the previous section. Because the mean reversion parameter is closer to the boundary of stationarity, the nonlinearities present in the binding function are stronger, and the b_T-mean unbiased IM estimator is farther from being mean unbiased than in the $\theta = (0, 0.66, 7.071)$ case. Then the "over-correction" present in the over-identified estimator pushes the estimates closer to the true value.

4.3.5 Test Statistics

Table 6 shows the joint LR-type tests, and Table 7 shows the LR-type test of the mean reversion parameter θ_1 for the just-identified models. The parameterization and simulation type is displayed in the table headings.

In the joint LR tests, the E1-type estimators have the highest rejection frequencies, followed by the M-type estimators. Interestingly, the IM estimators have the lowest rejection frequencies in the LR tests for θ_1. This can be explained by looking at Figure 19. The level curves of the objective function have a shape elongated in the $45°$ direction. An increase in θ_1 implies a decrease in the "slope" of the process, and assuming that the unconditional mean lies in the positive quadrant, to minimize squared residuals, the "intercept" or θ_0 has to increase. The M-type objective function is steeper that the L-type one as θ moves off the estimate. This is primarily caused by the nonlinearity of the binding function. Because the objective function is re-optimized while θ_1 is held fixed in the simple test, its value will be considerably smaller than when all θ parameters are being held fixed at their true values. Because the L-type estimates have a larger bias, the M-type estimator actually performs better than the L-type one in the simple test. A comparison of the upper and the lower portions of the Tables 6 and 7 shows that the rejection rate of LR-type tests increases as the process becomes more persistent.

Mean unbiasedness implies that the estimates from the positively skewed IM estimator will fall below the true parameter value with high probability (most of the density lies below the truth). In addition, the confidence interval decreases as θ_1 approaches zero. The combination of these two effects implies that the empirical rejection frequency of the LR-type test will be larger for the M-type estimator than for a median unbiased estimator.

Table 10 shows the over-identification statistic, or J-test, for the over-identified models, and Table 11 shows the joint LR-type tests. The parameterization and simulation type is displayed in the table headings.

In all cases, the E1-type estimators have the highest rejection frequencies. The rejection rate of LR-type and over-identification tests increases as the process becomes more persistent. The size of the LR-type tests is closest to the nominal size for the N, L, and A-type E2 and II estimators. The lower LR-type test rejection frequencies in Table 11 compared to Table 7 can be explained by the lower upward bias of the over-identified estimators discussed in the previous section.

5 Conclusion

[5] In this paper, we study indirect estimation methods with a special emphasis on issues related to continuous time models of the interest rate.

The EMM-2 (E2) estimator is based on the auxiliary score evaluated with the binding function and the observed data. It is asymptotically equivalent to the EMM (E1) estimator, but in finite samples it behaves differently. Along with the E2 estimator we subject the E1 and II estimators

[5] should we include any CIR or bond pricing results?

to Monte Carlo experiments to analyze their finite sample properties, and I discuss the issues pertinent to indirect estimation of continuous time models. Among other things, we demonstrate the dominance of the finite sample bias over the discretization bias in highly persistent time series, and this result led us to focus on the finite sample behavior of different estimator types (N, L, A, M). In our Monte Carlo study, we find that for the naturally occurring just identified estimators of continuous time models, the performance in point estimation mainly depends on these types as opposed to whether the estimator is score based or binding function based. The N, L, A type estimators deliver none, or very limited amount of, finite sample bias correction, and because the finite sample bias dominates at realistic parameter values, the M-type estimator could be the preferred estimator if its slightly slower speed is not an issue.

We noticed that the nonlinearity of the auxiliary statistics (auxiliary score and estimator) combined with model based restrictions may lead to additional bias in comparison to the bias of just identified estimators. In particular, we showed that the excessive bias of EMM criticized by Duffee and Stanton (2008) is caused by over-identification restrictions with parameter values that are sub-optimal for a given set of observations. To satisfy the first order conditions, the estimated parameters have to pick up the slack caused by these over-identifying restrictions, and as a result they will be biased. In fact, the situation is similar for all simulation based estimators, and we illustrated the distortion of the IM estimator caused by the interaction between the nonlinear terms of the binding function.

When the auxiliary model is a crude Euler discretization of the underlying diffusion, the same number of parameters will be present in both. If the same restrictions are placed on both the structural and auxiliary models, the estimator remains just identified, and the additional bias caused by over-identification can be avoided. Thus, in light of our findings, we recommend to keep the indirect estimators of interest rate diffusions just identified by either estimating all model parameters, or imposing the same restrictions on both the structural and auxiliary models.

We also find that the performance of E2 is comparable to II in most aspects. This result counters the criticism of the EMM estimator by Duffee and Stanton (2008) who compared E1 and II in an over-identifed setting: not only does EMM give the same point estimates in a just identified setting as II, but if it is based on the binding function, it also performs similarly to II in tests. [6]

The direct comparison of E1, E2 and II reveals that EMM based on the binding function, E2, is approximately equivalent to II. The added benefit of the binding function based E2 estimator is the availability of finite sample bias correction of the M-type estimator, which is unavailable for the E1 estimator. Given the dominance of the finite sample bias in discrete time approximations of continuous time interest rate models, the finite sample bias correction property of the M-type estimators is highly desirable, and my analysis shows that EM2 (and IM) deliver the most accurate

[6] Further, we find that the beneficial finite sample properties of the M-type estimators carry over to bond prices for some models, like the CIR model, but this result can not be generalized to all interest rate models. The reason is that the bond price is a nonlinear function of the underlying parameters, and a nonlinear transformation of a mean unbiased estimator does not imply mean unbiasedness of the transformed value. However, as long as the transformation is only mildly nonlinear, the illustrated plug-in method based on M-type estimates of the underlying parameter values becomes a feasible bias corrected bond price estimator.

point estimates. If the stability constraints on the structural simulator are known, as in the case of some single factor models of the interest rate, E1 should be avoided, and instead EM2 (or IM) can be used to achieve accurate finite sample results. Again, EMM is not necessarily inferior to II, and in fact it has the same benefits as II if it is based on the binding function, as I illustrated in this study.

References

Andersen, T. and Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77(2):343–377.

Andrews, D. (1993). Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models. *Econometrica*, 61(1):139–165.

Ball, C. and Torous, W. (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance*, 3(2):215–238.

Broze, L., Scaillet, O., and Zakoaan, J. (1998). Quasi-Indirect Inference for Diffusion Processes. *Econometric Theory*, 14(2):161–86.

Broze, L., Scaillet, O., and Zakoïan, J. (1995). Testing for continuous-time models of the short-term interest rate. *Journal of Empirical Finance*, 2(3):199–223.

Calzolari, G., Fiorentini, G., and Sentana, E. (2004). Constrained Indirect Estimation. *Review of Economic Studies*, 71(4):945–973.

Chan, K., Karolyi, G., Longstaff, F., and Sanders, A. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47(3):1209–1227.

Cox, J., Ingersoll Jr, J., and Ross, S. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407.

Czellar, V. and Ronchetti, E. (2008). Accurate and robust indirect inference for diffusion models. *Working paper.*

Czellar, V. and Zivot, E. (2008). Improved Small Sample Inference for Efficient Method of Moments and Indirect Inference Estimators. *Working paper.*

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.

Fuleky, P. (2009). A Note on the Choice of a Unit Period in Time Series Models. *Dissertation Chapter.*

Fuleky, P. and Zivot, E. (2010). Efficient Method of Moments Based on the Binding Function. *Working paper.*

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

Gallant, A. and Tauchen, G. (1997). Estimation of Continuous-Time Models for Stock Returns and Interest Rates. *Macroeconomic Dynamics*, 1(1):135–68.

Gallant, A. and Tauchen, G. (2004). EMM: A Program for Efficient Method of Moments Estimation. *Manuscript*.

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. (2006). Indirect Inference for Dynamic Panel Models. *Cowles Foundation Discussion Paper No. 1550*.

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*, pages 365–383.

Jiang, W. and Turnbull, B. (2004). The Indirect Method: Inference Based on Intermediate Statistics—A Synthesis and Examples. *Stat Sci*, 19(2):239–263.

Lo, A. (1988). Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data. *Econometric Theory*, 4(2):231–247.

Lord, R., Koekkoek, R., and van Dijk, D. (2006). A Comparison of Biased Simulation Schemes for Stochastic Volatility Models. *Tinbergen Institute Discussion Paper*.

Marriott, F. and Pope, J. (1954). Bias in the Estimation of Autocorrelations. *Biometrika*, 41(3/4):390–402.

Phillips, P. and Yu, J. (2005a). A Two-Stage Realized Volatility Approach to the Estimation for Diffusion Processes from Discrete Observations. *Cowles Foundation Discussion Paper No. 1523*.

Phillips, P. and Yu, J. (2005b). Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics. *Statistical Science*, 20(4):338–343.

Phillips, P. and Yu, J. (2005c). Jackknifing Bond Option Prices. *The Review of Financial Studies*, 18(2).

Phillips, P. and Yu, J. (2007). Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Cowles Foundation Discussion Paper No. 1597*.

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

Tang, C. and Chen, S. (2007). Parameter Estimation and Bias Correction for Diffusion Processes. *Working paper*.

Tauchen, G. (1998). The Objective Function of Simulation Estimators Near the Boundary of the Unstable Region of the Parameter Space. *Review of Economics and Statistics*, 80(3):389–398.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5(2):177–188.

Yu, J. (2008). Bias in the Estimation of Mean Reversion Parameter in Continuous Time Models. Technical report, Working Paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University.

Zhou, H. (2001). Finite Sample Properties of EMM, GMM, QMLE, and MLE for a Square-Root Interest Rate Diffusion Model. *Journal of Computational Finance*, 5:89–122.

6 Appendix: Bias of the Auxiliary Estimator of an OU Model

During calibration, $\tilde{\mu}^{\mathrm{M}}_{S0}(\theta)$ and $\tilde{\mu}^{\mathrm{M}}_{S1}(\theta)$ are jointly estimated by OLS. Let $\tilde{\mu}_i(\theta)$ represent $\tilde{\mu}^{\mathrm{M}}_{1i}(\theta)$, and $E[\tilde{\mu}_i(\theta)]$ represent $\tilde{\mu}^{\mathrm{M}}_{\infty i}(\theta)$. To find the bias of $\tilde{\mu}_1(\theta)$, I start out with (35) rewritten in the AR(1) form used by DS in (51). Marriott and Pope (1954) showed that up to order $O(n^{-2})$ the finite sample bias of the AR(1) coefficient, θ_1^{DS}, when the estimating equation includes a constant, θ_0^{DS}, is [7]

$$E[\hat{\theta}_1^{DS}] - \theta_1^{DS} = -\frac{3\theta_1^{DS} + 1}{n} + O(n^{-2}) = -\frac{3e^{-\theta_1\Delta} + 1}{n} + O(n^{-2}) \ . \tag{56}$$

This translates to the following finite sample bias of auxiliary estimate of the mean reversion parameter, $\tilde{\mu}_1(\theta)$,

$$E[\tilde{\mu}_1(\theta)] - \tilde{\mu}_1(\theta) = \frac{1}{\Delta}\left[(1 - E[\hat{\theta}_1^{DS}]) - (1 - \theta_1^{DS})\right] = \frac{1}{\Delta}\frac{3e^{-\theta_1\Delta} + 1}{n} + O(n^{-2}) \ , \tag{57}$$

as can be verified by examining the relationship between (36) and (51). However, given the continuous data generating process, the discretized auxiliary model is misspecified (Lo, 1988), and the auxiliary parameter estimates contain a discretization bias in addition to the finite sample bias

$$E[\tilde{\mu}_1(\theta)] - \theta_1 = E[\tilde{\mu}_1(\theta)] - \tilde{\mu}_1(\theta) + \tilde{\mu}_1(\theta) - \theta_1 \approx \frac{1}{\Delta}\frac{3e^{-\theta_1\Delta} + 1}{n} + \frac{1}{\Delta}(1 - e^{-\theta_1\Delta}) - \theta_1 \ , \tag{58}$$

where $\frac{1}{\Delta}(1 - e^{-\theta_1\Delta}) - \theta_1$ is the asymptotic discretization bias. Its value is negative and is dominated by the positive finite sample bias at small $\theta_1\Delta$ values.

To get an approximation to the bias of the auxiliary constant $\tilde{\mu}_0(\theta)$ evaluate its estimator with the biased $\tilde{\mu}_1(\theta)$

$$\tilde{\mu}_0(\theta) = \overline{y_t} - (1 - E[\tilde{\mu}_1(\theta)])\overline{y_{t-\Delta}} \ . \tag{59}$$

Replacing $E[\overline{y_{t-\Delta}}]$ with $E[y_{t-\Delta}] = \theta_0/\theta_1$ one gets

$$E[\tilde{\mu}_0(\theta)] = \frac{\theta_0}{\theta_1}E[\tilde{\mu}_1(\theta)] = \theta_0\frac{E[\tilde{\mu}_1(\theta)]}{\theta_1} = \theta_0\left(1 + \frac{E[\tilde{\mu}_1(\theta)] - \theta_1}{\theta_1}\right) \tag{60}$$

$$\approx \theta_0\left(\frac{3e^{-\theta_1\Delta} + 1}{\theta_1 T} + \frac{1 - e^{-\theta_1\Delta}}{\theta_1\Delta}\right) \ , \tag{61}$$

where $T = n\Delta$. The second term in (61) is the relative discretization bias of $\tilde{\mu}_1(\theta)$, which is close to 1 for small values of $\theta_1\Delta$. The last expression in (60) indicates that the bias of $\tilde{\mu}_0(\theta)$ depends on the relative bias of $\tilde{\mu}_1(\theta)$, which quickly increases as $\theta_1 \to 0$. Thus, for small positive values of θ_1 the constant term in the simulated auxiliary estimate, $\tilde{\mu}_0(\theta)$, will be biased away from zero; that is, the estimates of positive constants will be biased up and those of negative constants will be biased down in finite samples.

[7] Tang and Chen (2007) and Yu (2008) derive the finite sample bias of a directly estimated mean reversion parameter, θ_1, up to second order. Theirs is a refinement of the Marriott and Pope (1954) result for direct estimation of the continuous time process.

For $\theta_0 = 0.01$, $\theta_1 = 0.1$, $\delta = 1/50$, and $n = 1000$, the analytical expressions above give $bias(\tilde{\mu}_0(\theta)) \approx 0.02$ and $bias(\tilde{\mu}_1(\theta)) \approx 0.2$.

OU RESULTS

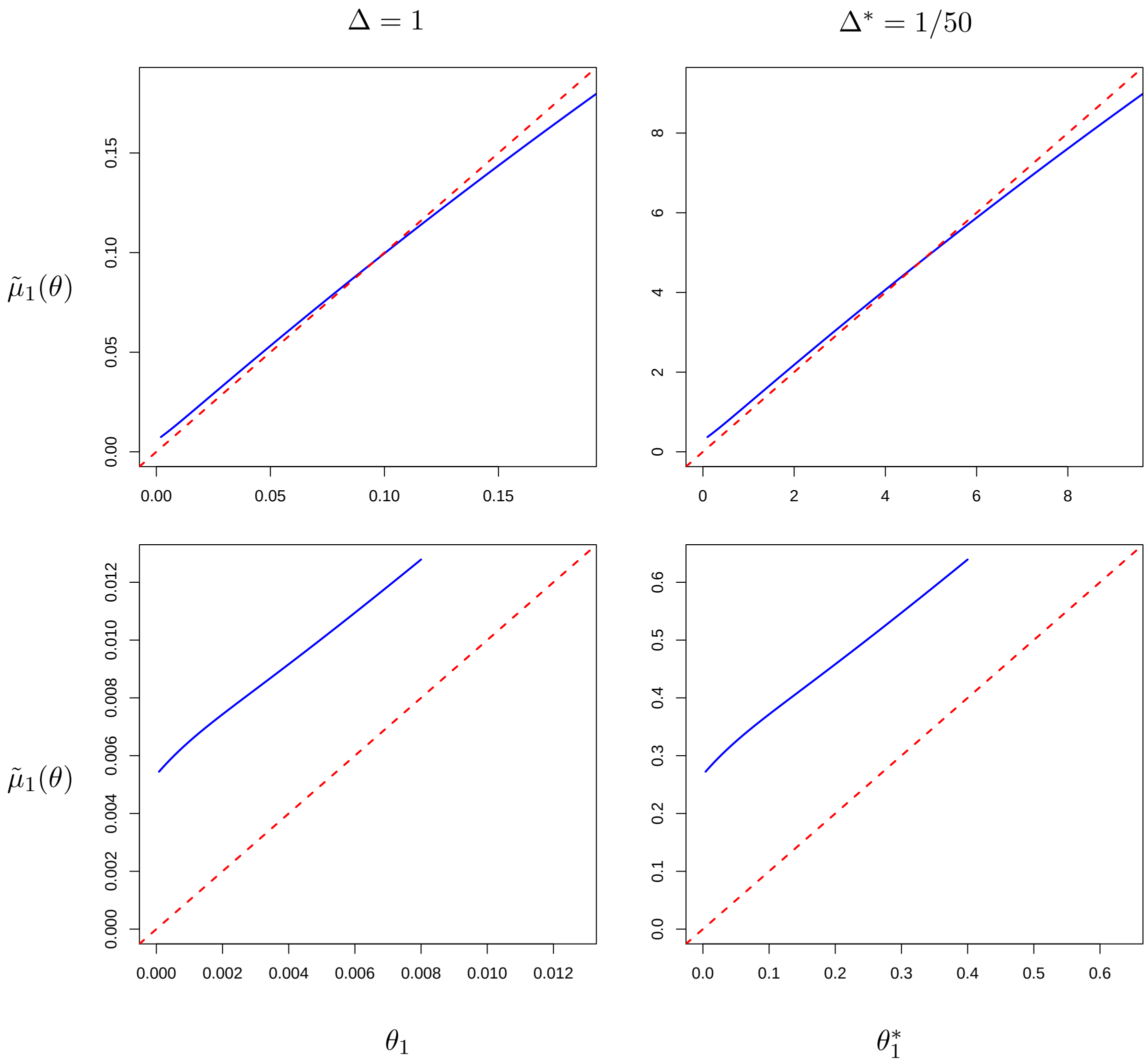


Figure 2: Red dashed 45° line: theoretical unbiased estimator. Blue line: mean of 1000 estimates of the mean reversion parameter μ_1 as a function of θ_1. The estimates are based on weekly observations. The difference between the left and right diagrams is the choice of the unit period - left: $\Delta = 1$ (parameter values in weekly terms), right: $\Delta = 1/50$ (parameter values in annual terms). The generated data are the same both, left and right. The range of θ_1 and θ_1^* is shown on the horizontal axis, and the remaining parameters are being held constant at $\theta_0 = 0.1,\ \theta_2 = 0.1/\sqrt{50}$, and $\theta_0^* = 0.1,\ \theta_2^* = 0.1$. The bottom diagrams show a magnified portion of the top diagrams close to 0. The choice of the unit period has a scaling effect, but the relative bias is the same for both unit period choices. The top diagrams illustrate that the finite sample bias dominates close to 0, but at $\theta_1 = 0.1$ or $\theta_1^* = 5$ the discretization bias takes over.

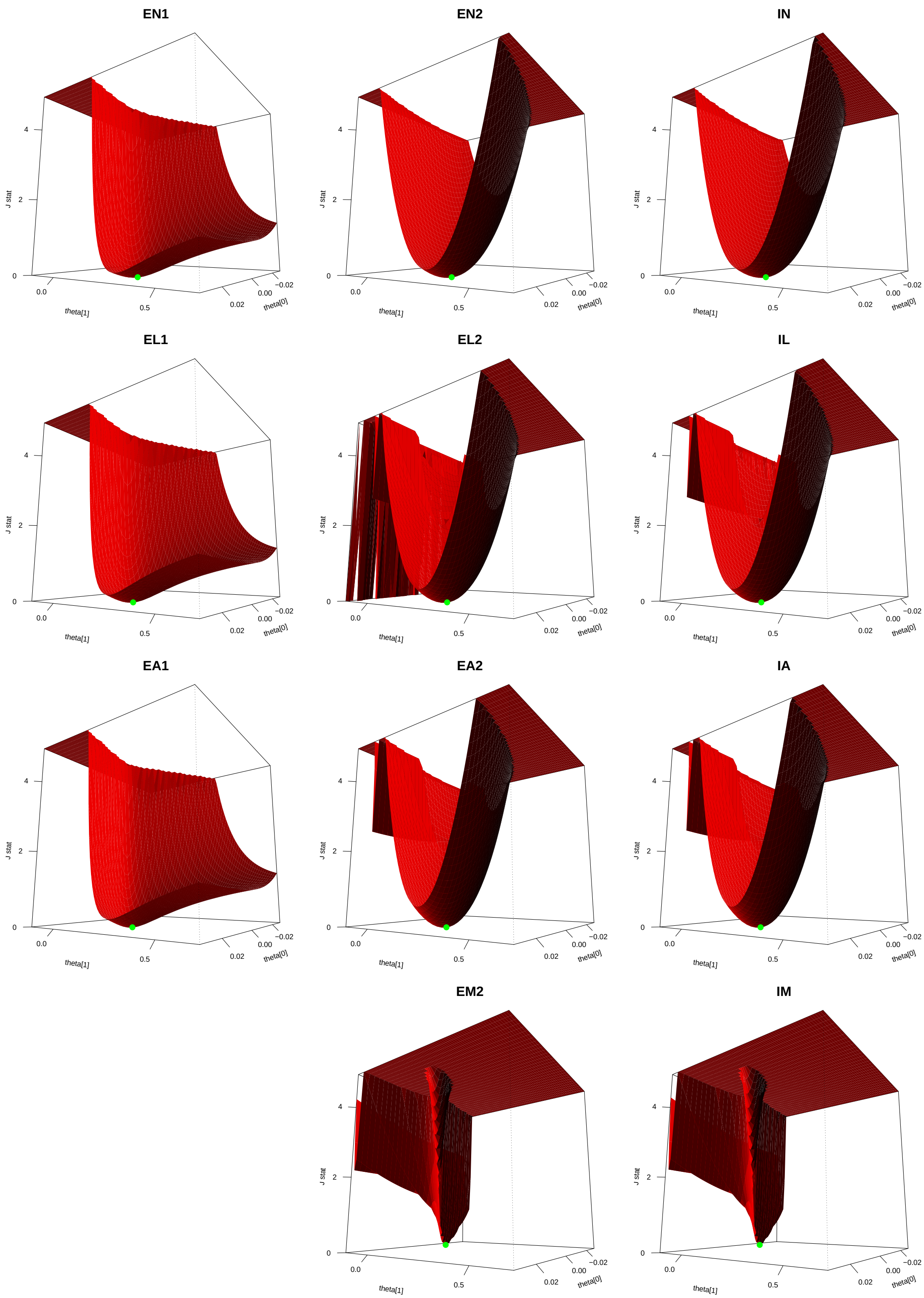


Figure 3: Objective functions of score based EMM, binding function based EMM-2, and II estimators. Table 1 displays the estimates (green dot).

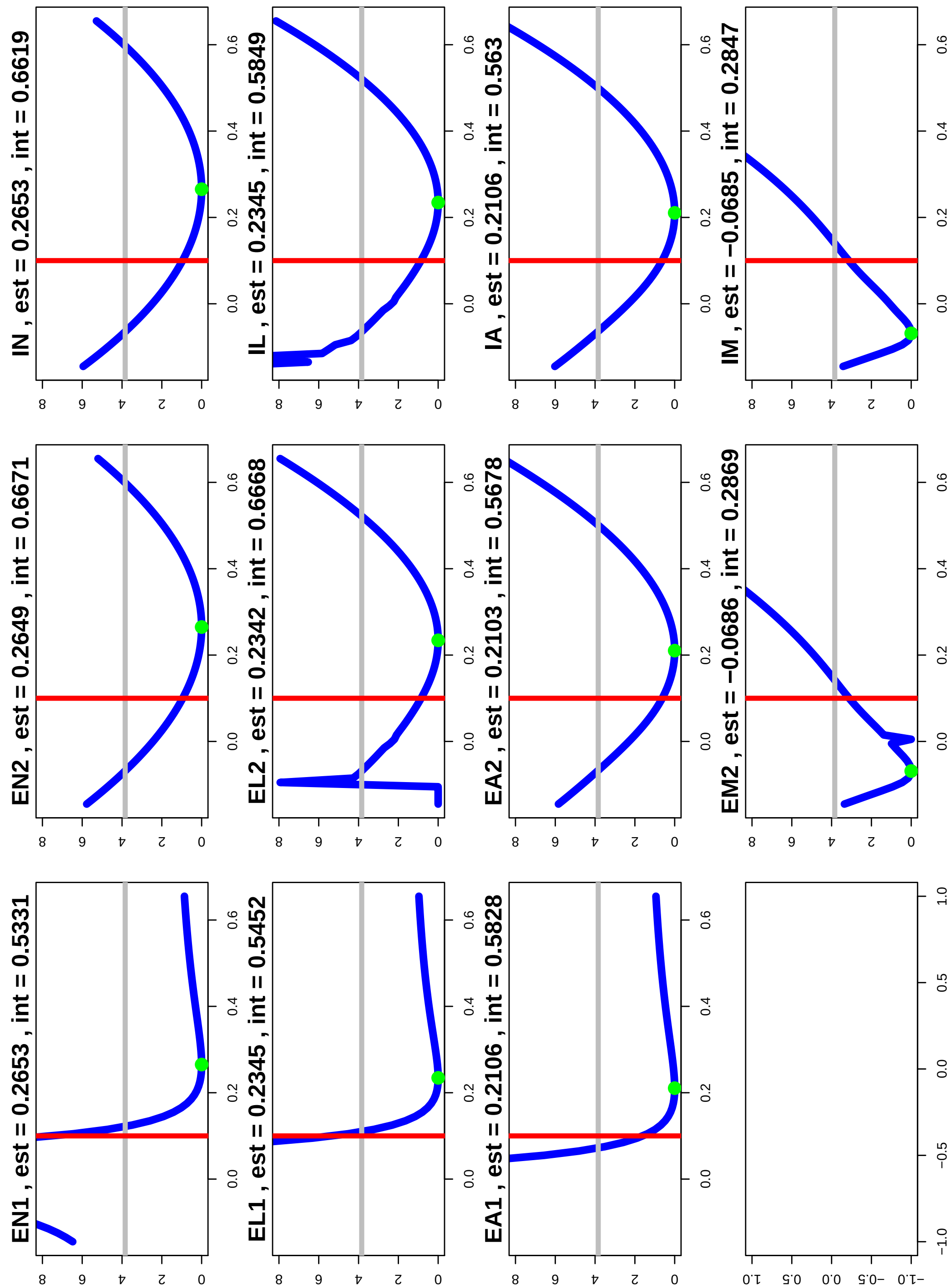


Figure 4: LR-type statistics for testing $H_0 : \theta_1 = \theta_1^0$ as functions of θ_1^0 in the just identified model. "est" represents the estimate (green dot) and "int" represents the length of the confidence interval. The auxiliary estimate is $\tilde{\mu}_1 = 0.2645$, and Table 1 displays the structural estimates (green dot). Red line=true value of $\theta_1 = 0.1$.

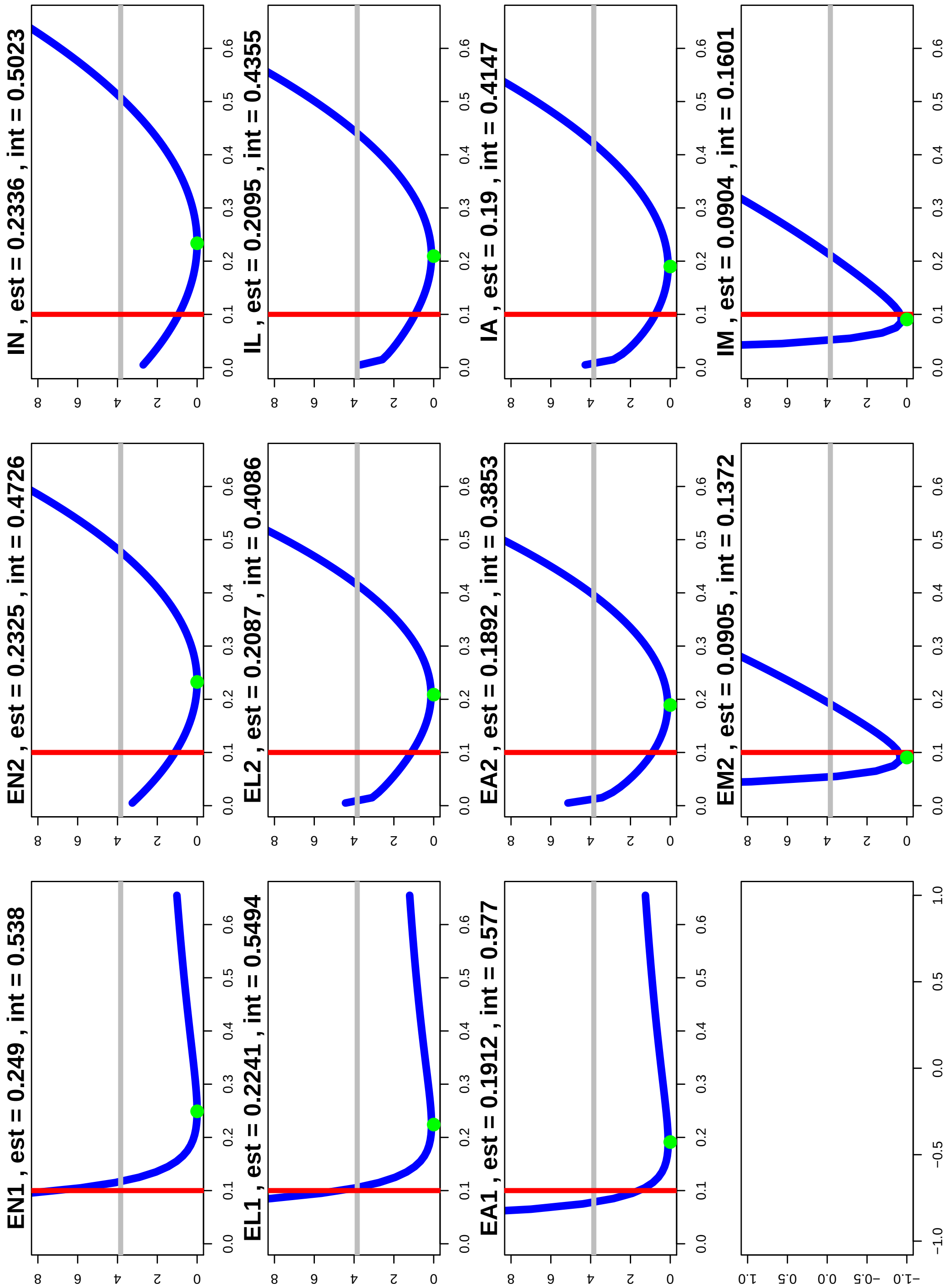
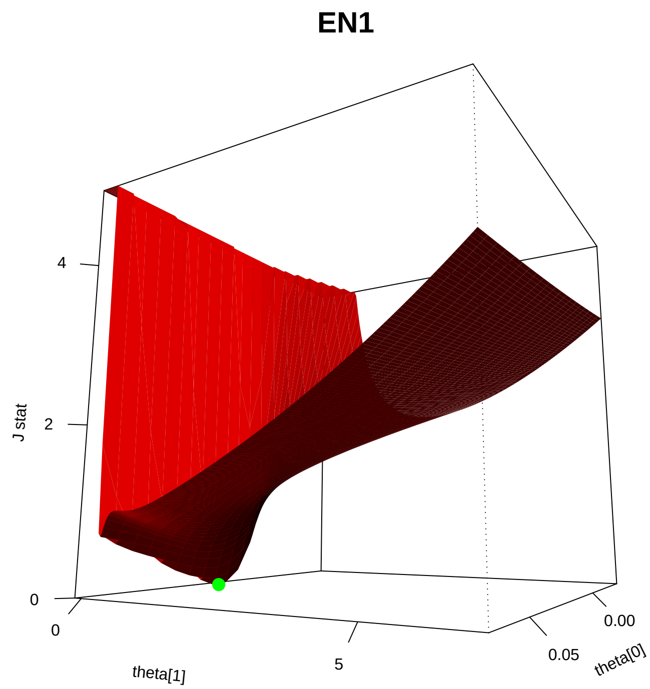


Figure 5: LR-type statistics for testing $H_0 : \theta_1 = \theta_1^0$ as functions of θ_1^0 in the over-identified model. "est" represents the structural estimate (green dot) and "int" represents the length of the confidence interval. The auxiliary estimate is $\tilde{\mu}$=0.2645. Red line=true value of $\theta_1 = 0.1$



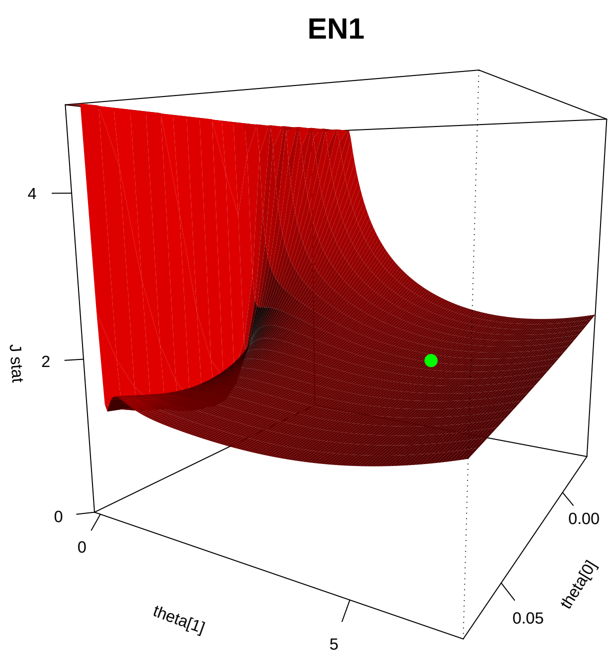


Figure 6: EN1 objective function evaluated with the just identified estimate of $\theta_2 = 0.098$ (top), and the true value of $\theta_2 = 0.1$ (bottom). The green dot represents the just identified estimate $\hat{\theta}^{\mathrm{EN1}} = (0.0287, 0.1967, 0.0980)$ on the top, and the over-identified estimate $\hat{\theta}_1^{\mathrm{EN1}} = 4.2748$, with $\theta_0 = 0.01$ and $\theta_2 = 0.1$, on the bottom.

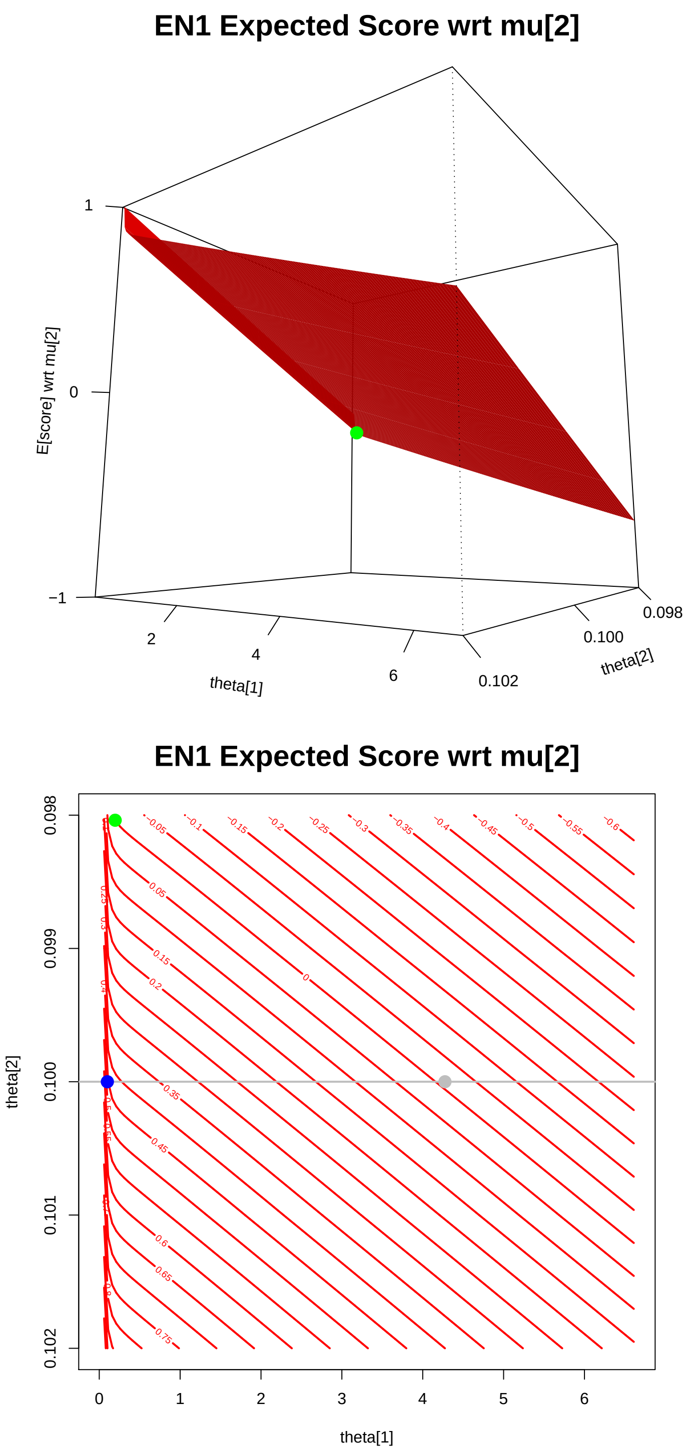


Figure 7: Third term in the expected score (derivative of the likelihood with respect to μ_2) as a function of θ_1 and θ_2. θ_0 is being held fixed at its just identified estimate. Color notation: blue = true value: $\theta = (0.01, 0.1, 0.1)$; green = just identified estimate: $\hat{\theta}^{\text{EN1}} = (0.0287, 0.1967, 0.0980)$; grey = over-identified estimate: $\hat{\theta}_1^{\text{EN1}} = 4.2748$.

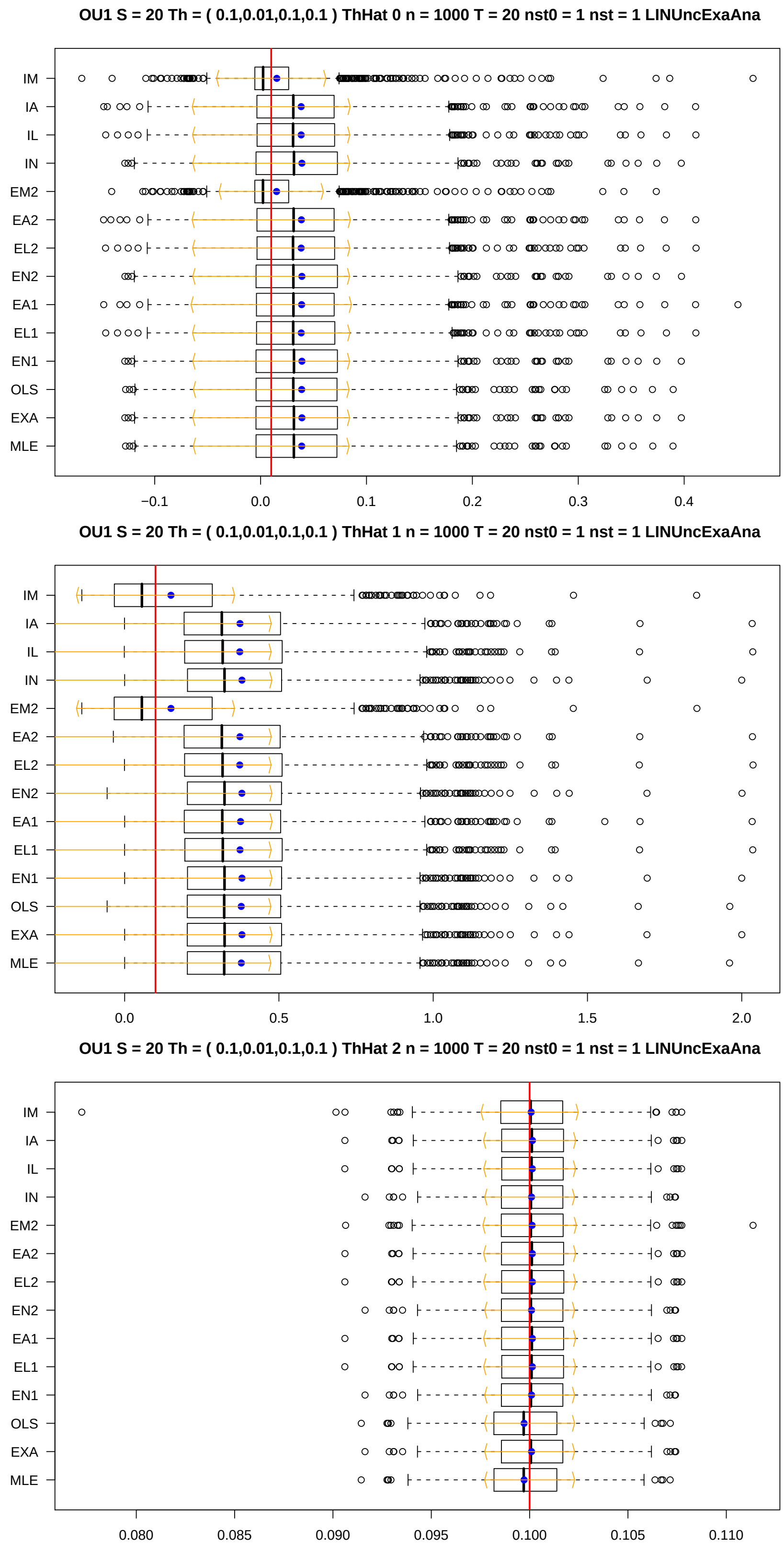


Figure 8: Distribution of θ_0, θ_1, and θ_2 estimates in a 3×3 just-identified model with $\theta = (0.01, 0.1, 0.1)$, and $T = 20$ based on simulations from exact discretization with $S = 20$, $\Delta = 1/50$. The red line represents the true value, blue dot the mean of the estimates, and the yellow arrow the RMSE of the estimates.

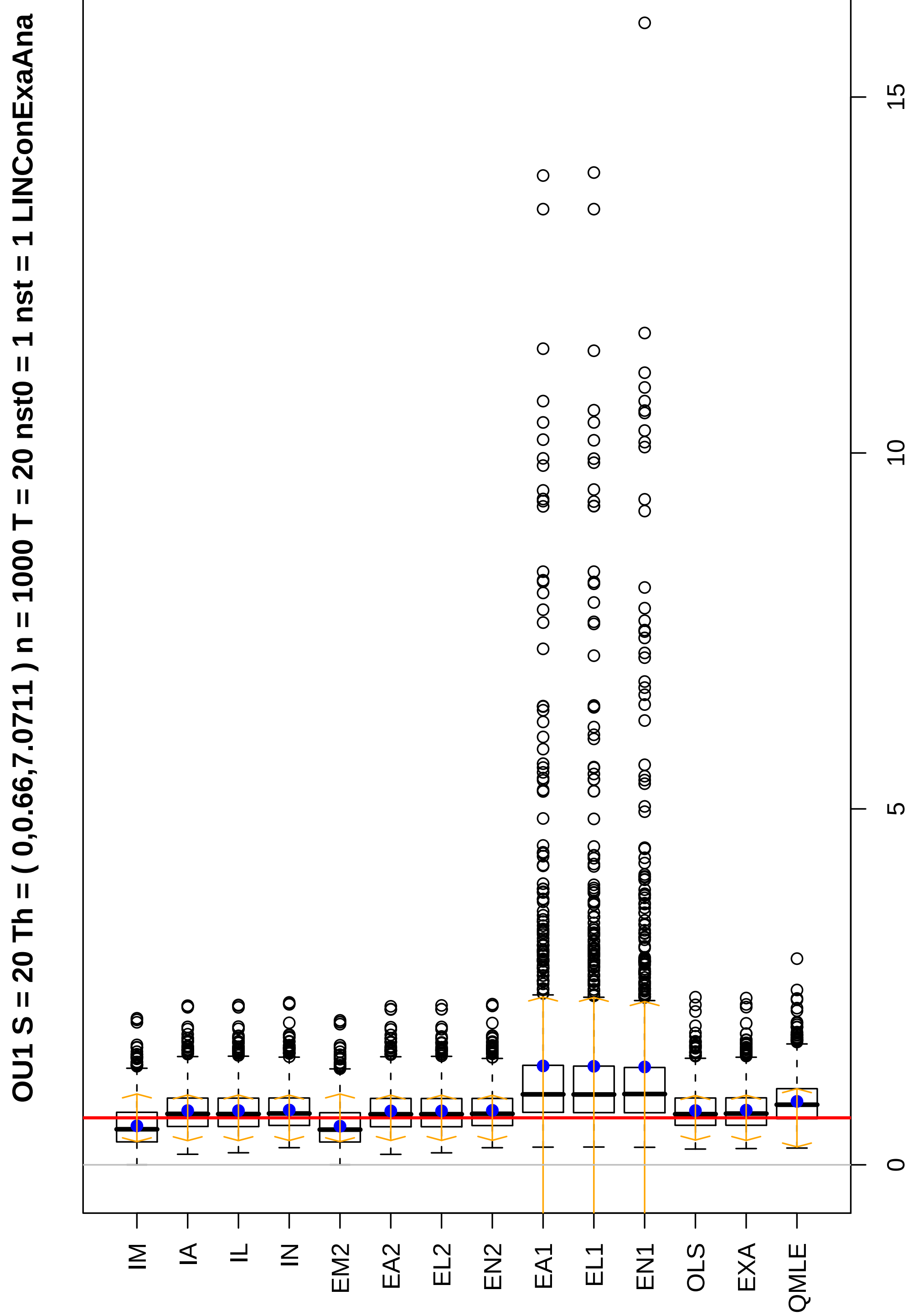


Figure 9: Distribution of θ_1 estimates in a 1×3 over-identified model with $\theta = (0, 0.66, 7.071)$, and $T = 20$ based on simulations from exact discretization with $S = 20$, $\Delta = 1/50$. The red line represents the true value, blue dot the mean of the estimates, and the yellow arrow the RMSE of the estimates.

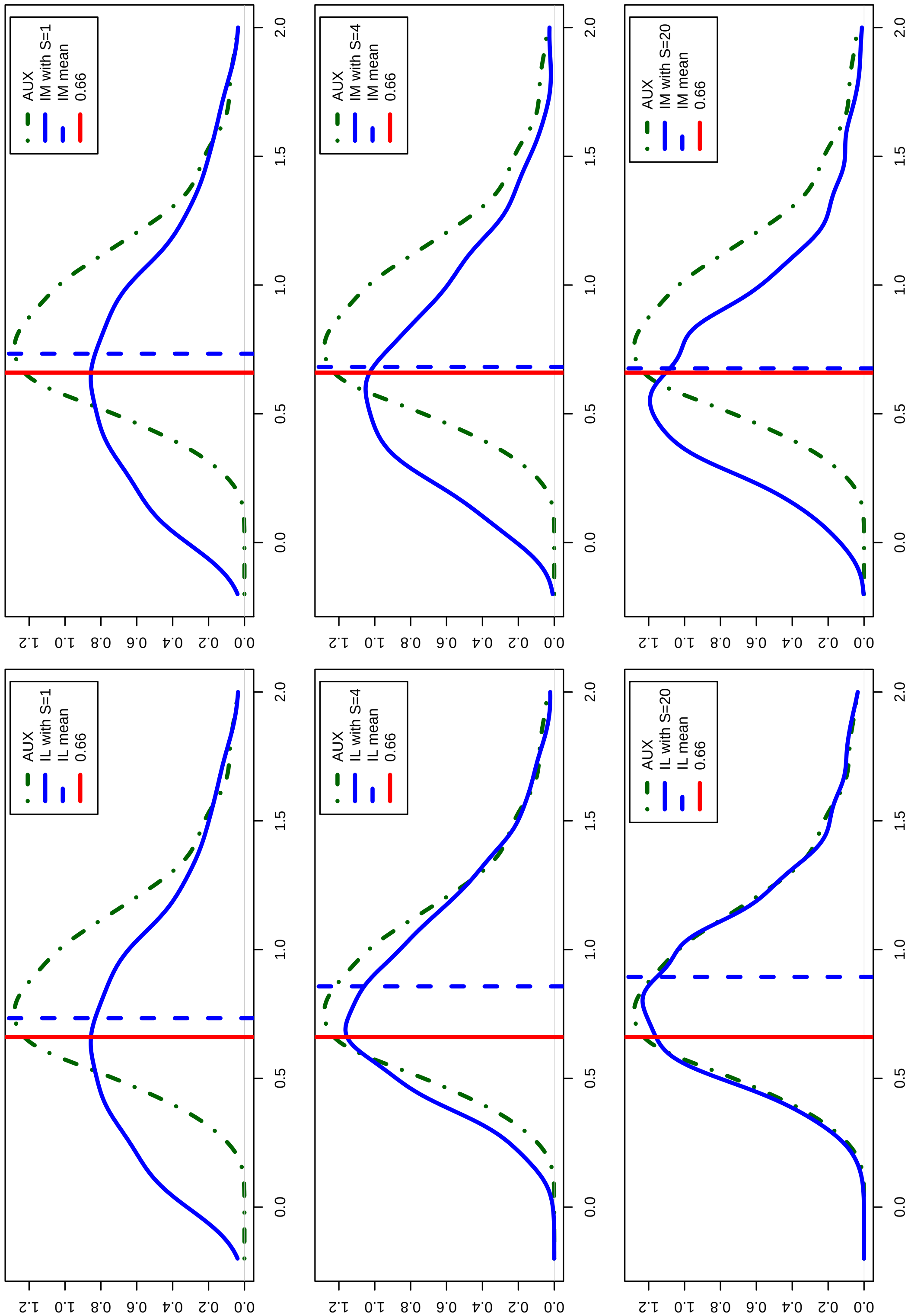


Figure 10: Impact of simulation size on the just-identified IL and IM estimators of θ_1 for $\theta = (0, 0.66, 7.071)$, $T = 20$, $\Delta = 1/50$. The red line represents the true $\theta_1 = 0.66$, the green line is the density of the auxiliary estimator, and the blue lines are the density of the indirect estimator (left IL, right IM), and its mean.

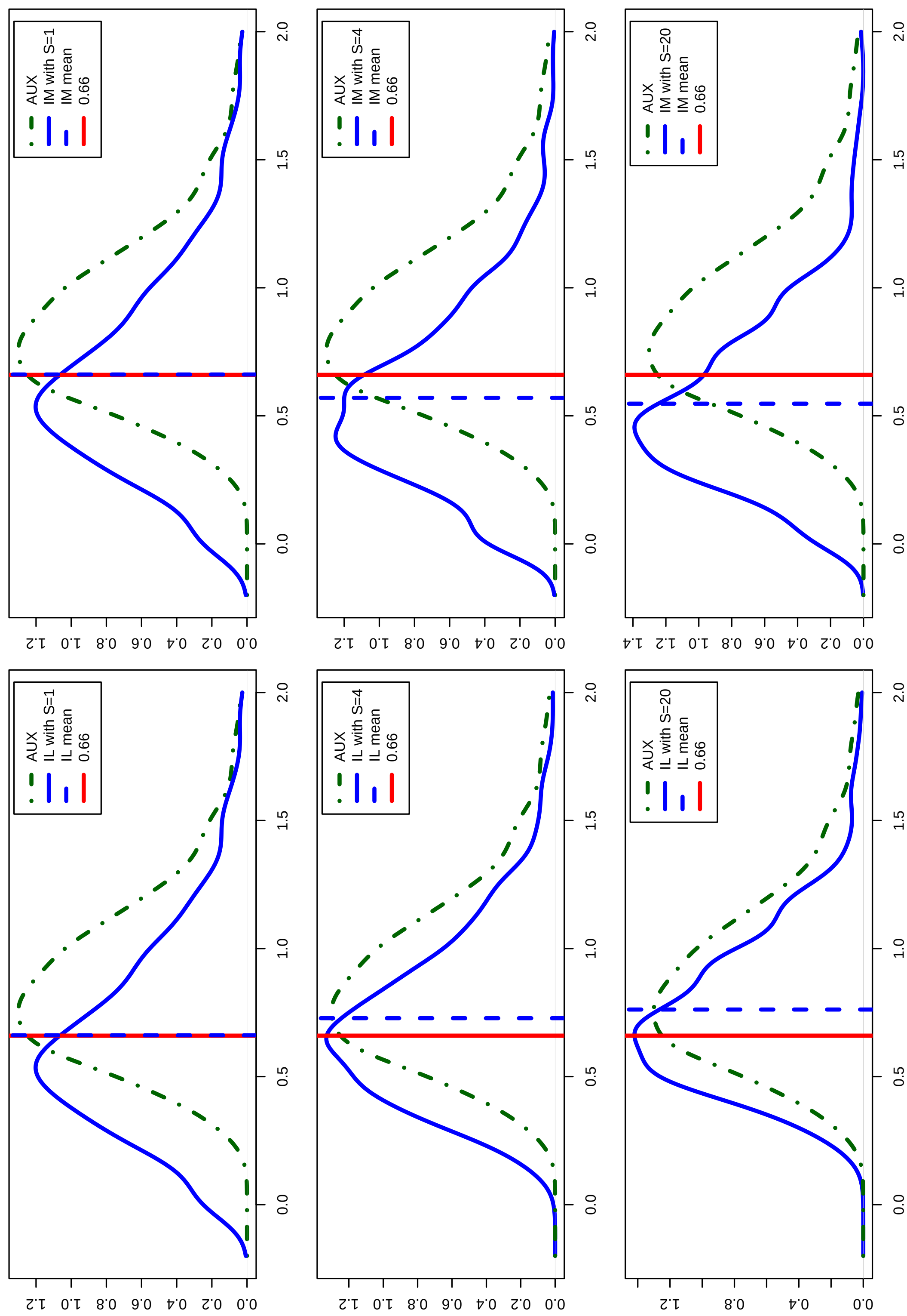
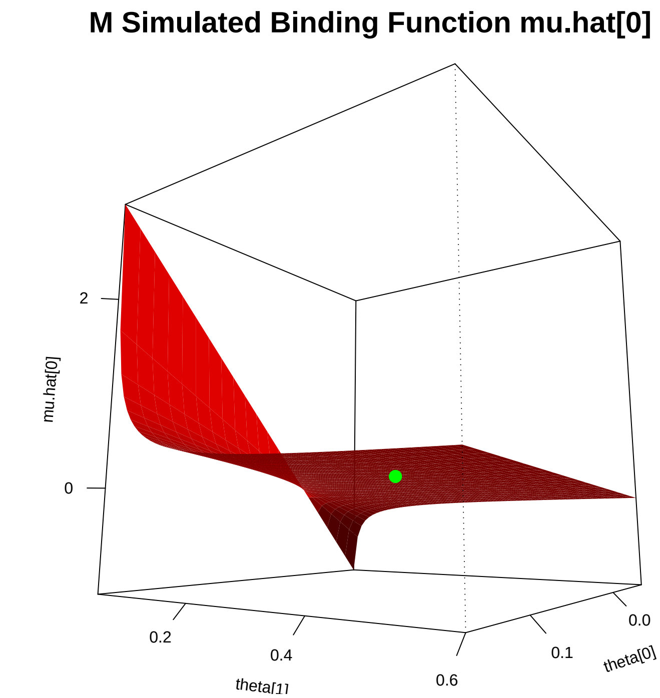


Figure 11: Impact of simulation size on the over-identified IL and IM estimators of θ_1 for $\theta = (0, 0.66, 7.071)$, $T = 20$, $\Delta = 1/50$. The red line represents the true $\theta_1 = 0.66$, the green line is the density of the auxiliary estimator, and the blue lines are the density of the indirect estimator (left IL, right IM), and its mean.



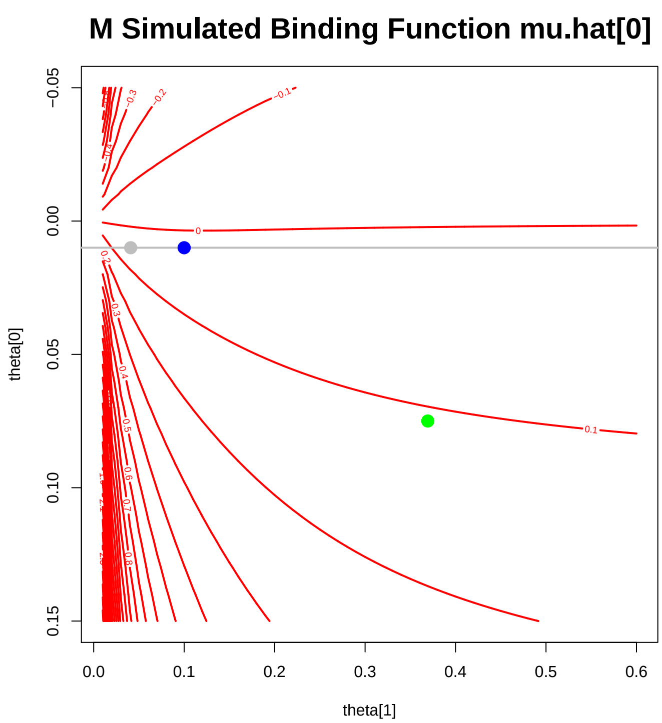


Figure 12: Simulated binding function of M type estimator, $\tilde{\mu}_0^{\mathrm{IM}}(\theta)$, as a function of θ_0 and θ_1. θ_2 is being held fixed at its just identified estimate. Color notation: blue = true value: $\theta = (0.01, 0.1, 0.1)$; green = just identified estimate: $\hat{\theta}^{\mathrm{IM}} = (0.0749, 0.3693, 0.1001)$; grey = over-identified estimate: $\hat{\theta}_1^{\mathrm{IM}} = 0.0409$.

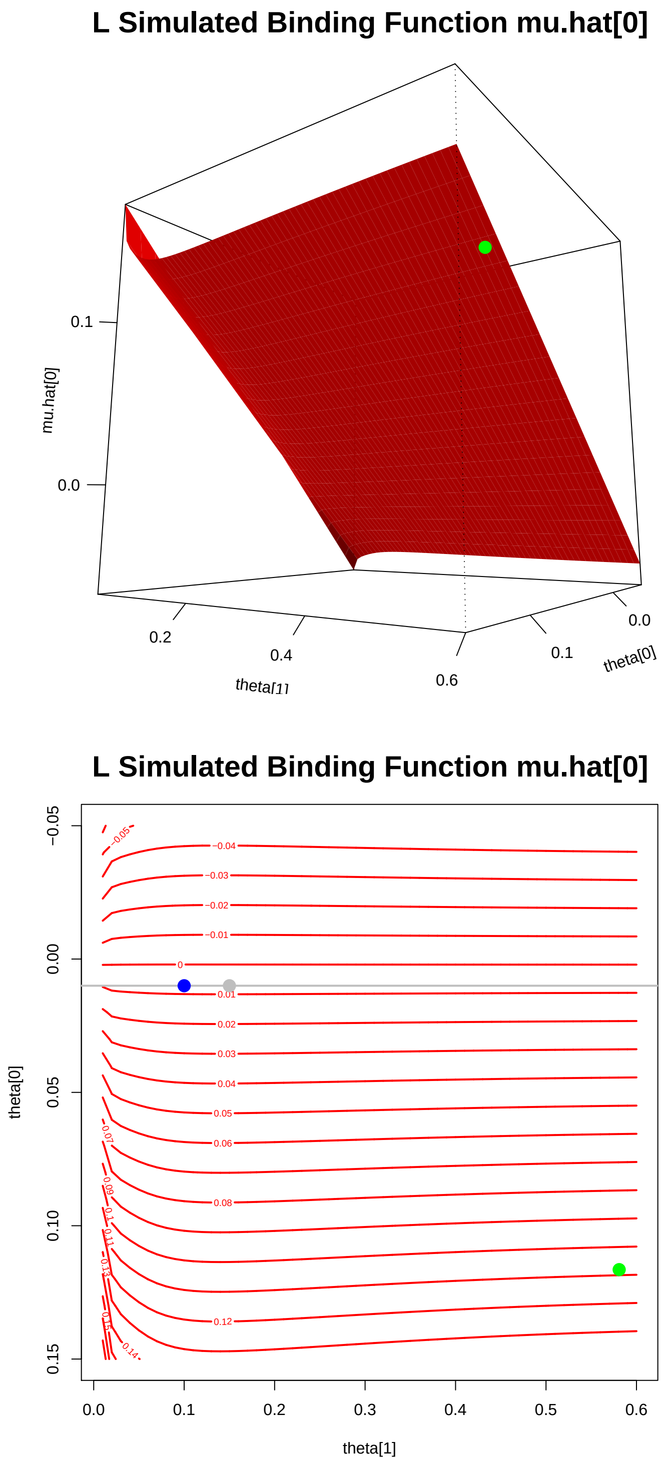
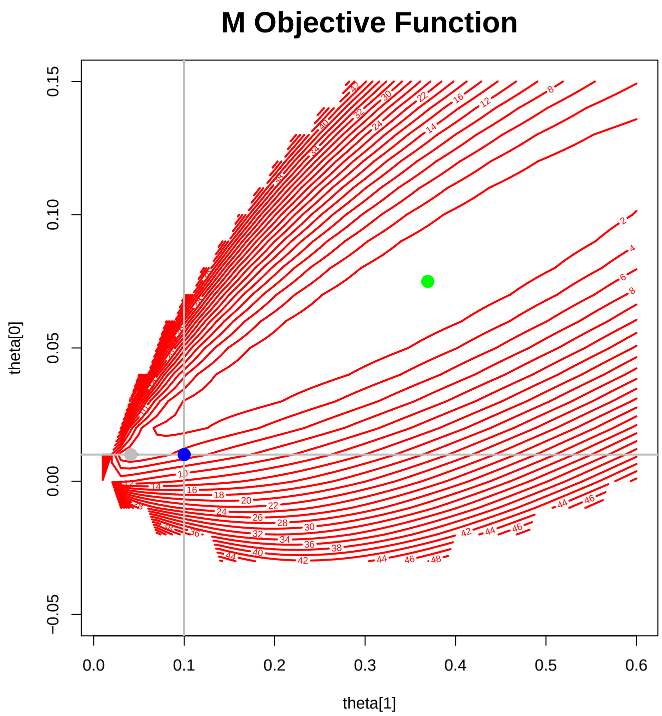


Figure 13: Simulated binding function of L type estimator, $\tilde{\mu}_0^{\mathrm{IL}}(\theta)$, as a function of θ_0 and θ_1. θ_2 is being held fixed at its just identified estimate. Color notation: blue = true value: $\theta = (0.01, 0.1, 0.1)$; green = just identified estimate: $\hat{\theta}^{\mathrm{IL}} = (0.1164, 0.5808, 0.1002)$; grey = over-identified estimate: $\hat{\theta}_1^{\mathrm{IL}} = 0.1500$.



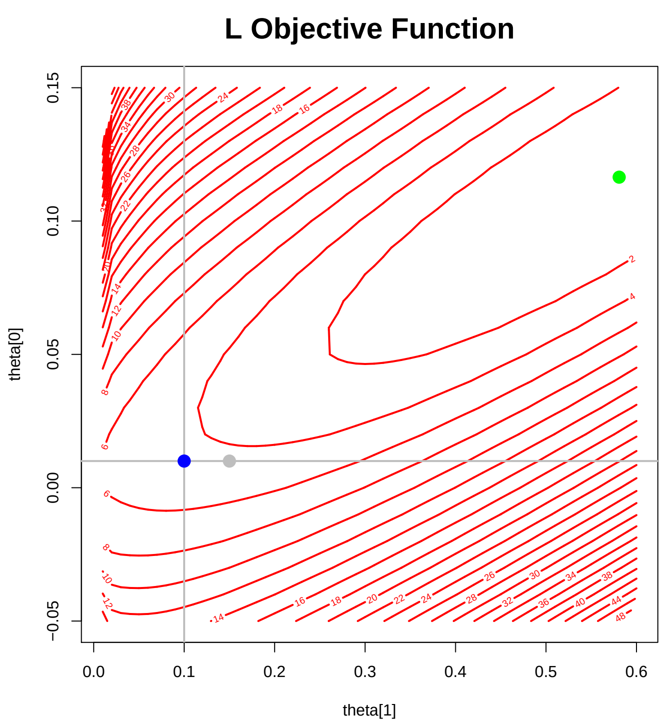


Figure 14: TOP: Contours of M-type objective function, as a function of θ_0 and θ_1. θ_2 is being held fixed at its just identified estimate. Color notation: blue = true value: $\theta = (0.01, 0.1, 0.1)$; green = just identified estimate: $\hat{\theta}^{\text{IM}} = (0.0749, 0.3693, 0.1001)$; grey = over-identified estimate: $\hat{\theta}_1^{\text{IM}} = 0.0409$. BOTTOM: Contours of L-type objective function, as a function of θ_0 and θ_1. θ_2 is being held fixed at its just identified estimate. Color notation: blue = true value: $\theta = (0.01, 0.1, 0.1)$; green = just identified estimate: $\hat{\theta}^{\text{IL}} = (0.1164, 0.5808, 0.1002)$; grey = over-identified estimate: $\hat{\theta}_1^{\text{IL}} = 0.1500$.

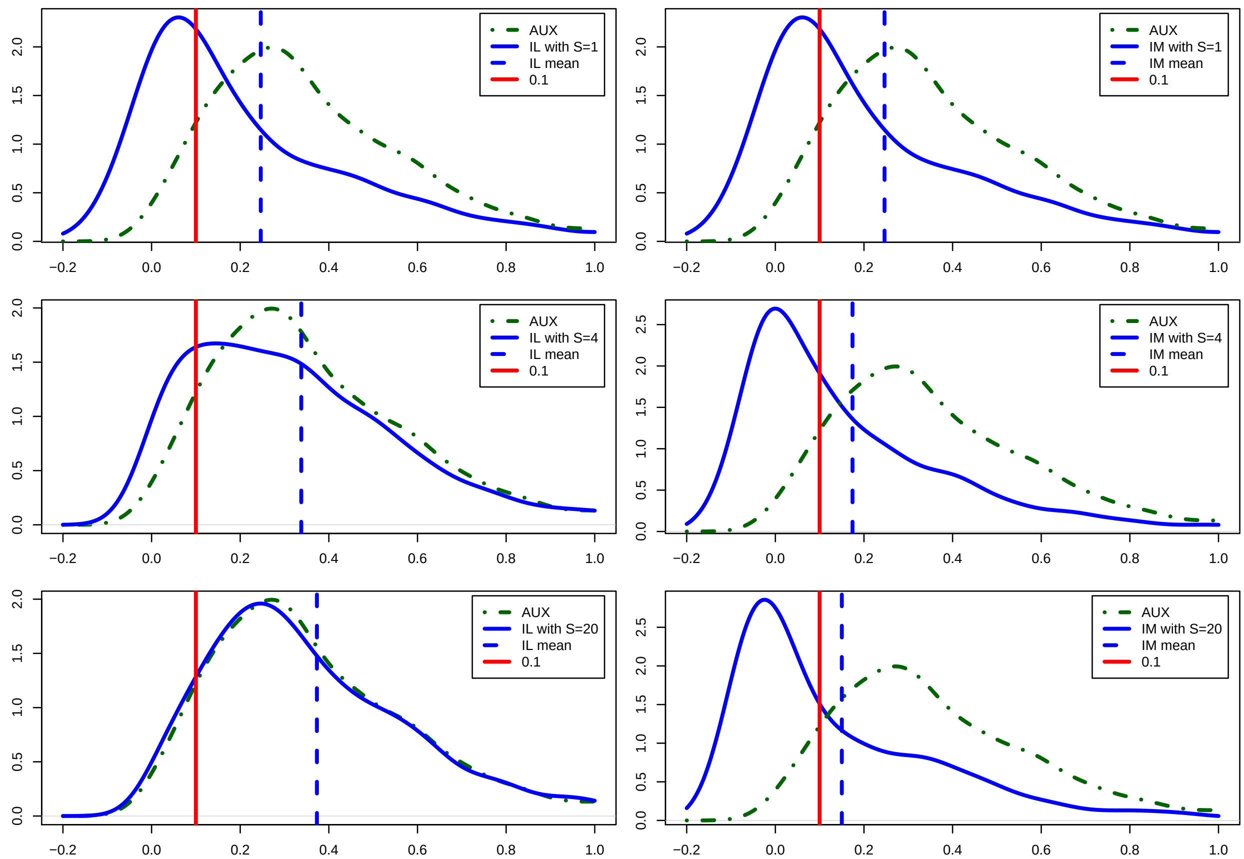


Figure 15: Distribution of $\tilde{\mu}_1$ (green), $\hat{\theta}_1^{IL}$ (blue, left) and $\hat{\theta}_1^{IM}$ (blue, right) for $\theta = (0.01, 0.1, 0.1)$, n=1000, $\Delta = 1/50$. From top to bottom, S takes on the values 1, 4, and 20.

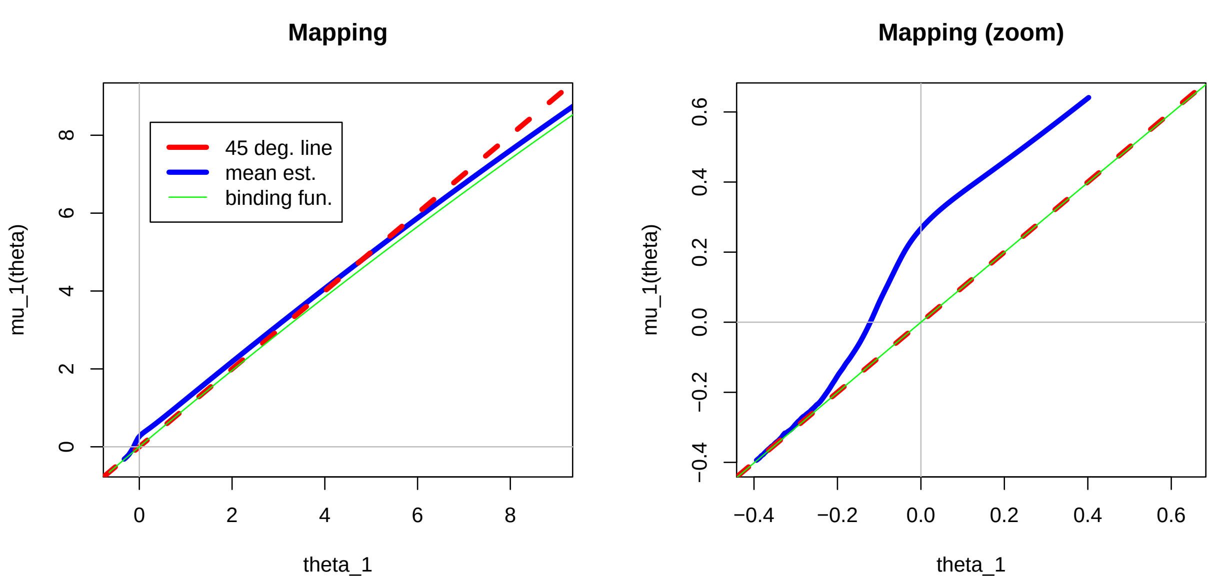


Figure 16: Mapping from θ to μ_1 for $\theta = (0.01, \theta_1, 0.1)$, n=1000, $\Delta = 1/50$. Green line: analytic binding function, $\mu_1(\theta)$. Blue line: mean estimate, $\tilde{\mu}_1^{IM}(\theta)$, based on $S = 1000$. The nonlinearity of $\tilde{\mu}_1^{IM}(\theta)$ is caused by the super-consistency of the estimator for $\theta_1 \leq 0$.

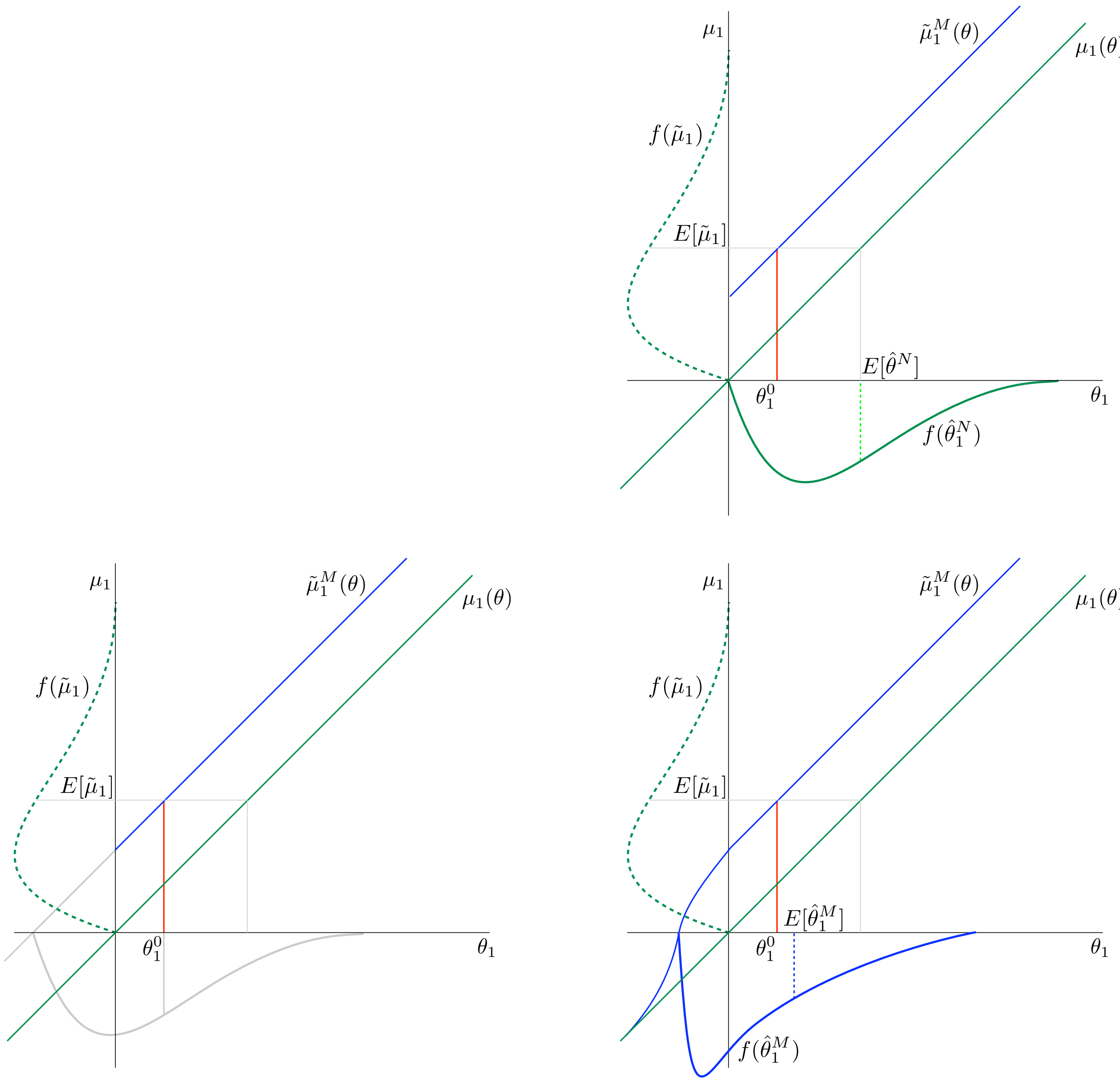


Figure 17: Finding the distribution of the θ_1 estimates by mapping the distribution of the $\tilde{\mu}_1$ estimates onto θ_1. The bias of the θ_1 estimator depends on the bias present in the binding function. An unbiased binding function $\mu_1(\theta)$ results in the same bias in $\hat{\theta}_1$ as there is in $\tilde{\mu}_1$ (top right). A linear binding function shifted by the amount of bias would eliminate the bias present in $\tilde{\mu}_1$, that is, it would result in a mean unbiased θ_1 estimator (lower left). If the binding function is non-linear, full bias correction of $\hat{\theta}_1$ will not be achieved (lower right).

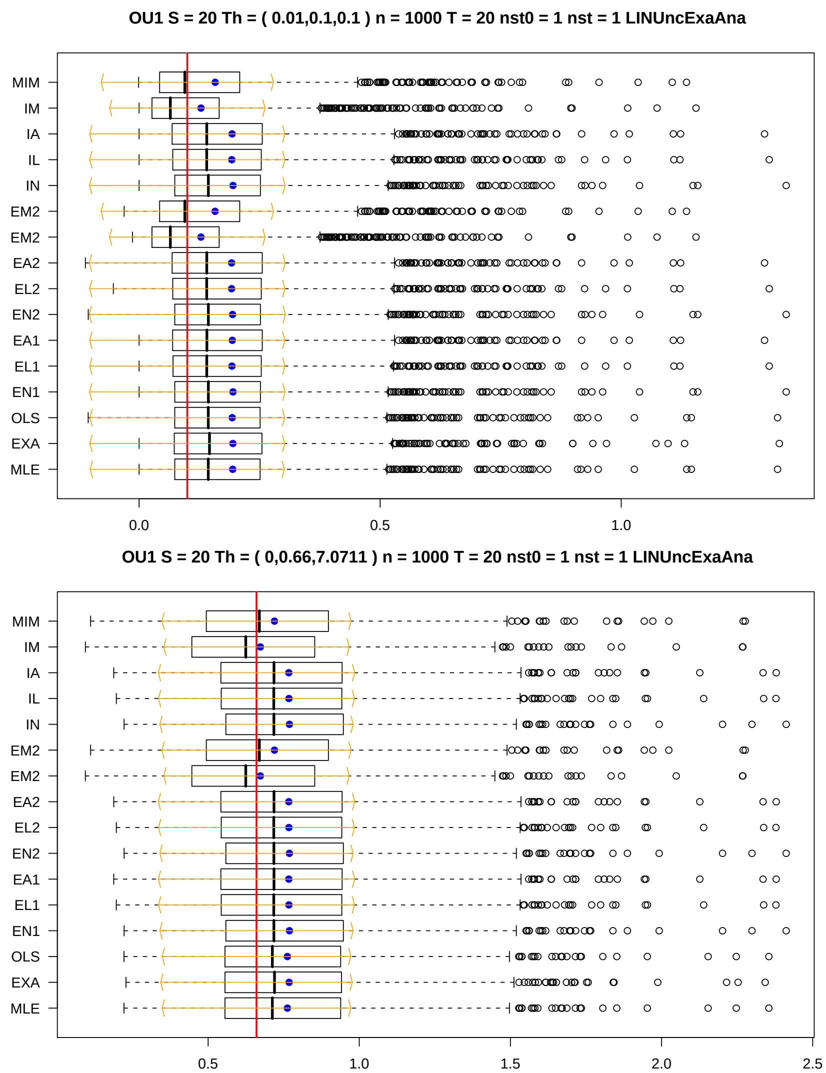
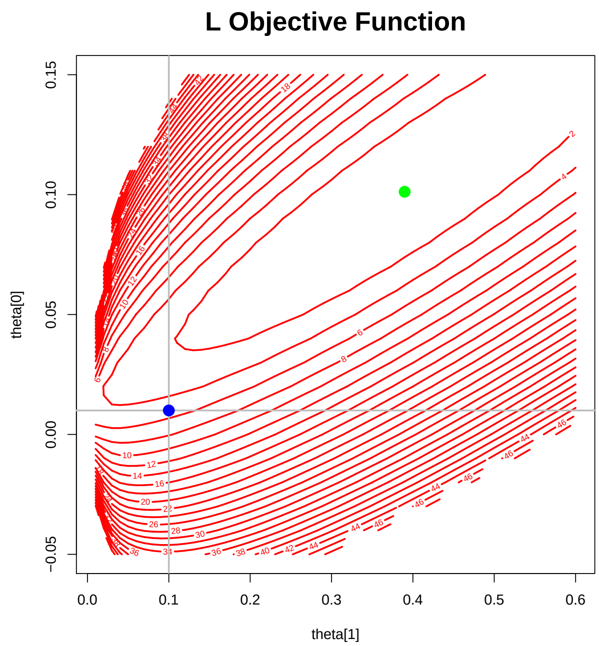


Figure 18: Distribution of θ_1 estimates in a 1×1 just-identified model with $\theta = (0.01, 0.1, 0.1)$, and $\theta = (0, 0.66, 7.071)$, and $T = 20$ based on simulations from exact discretization with $S = 20$, $\Delta = 1/50$. The red line represents the true value, blue dot the mean of the estimates, and the yellow arrow the RMSE of the estimates.



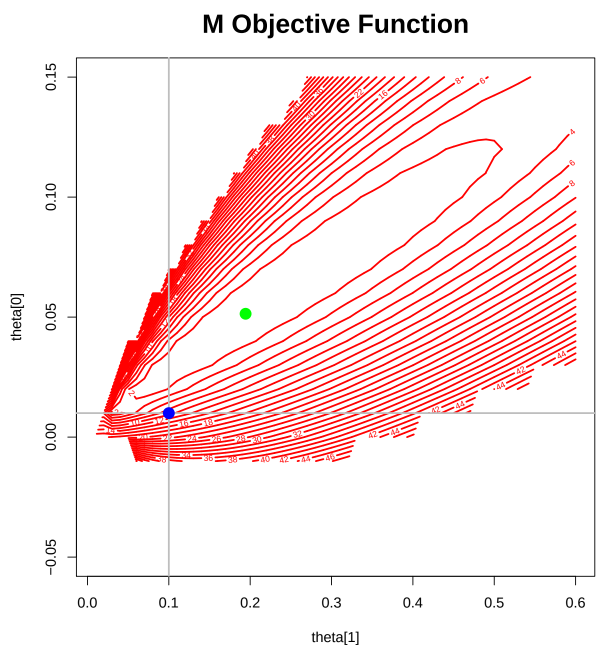


Figure 19: TOP: Contours of L-type objective function, as a function of θ_0 and θ_1. θ_2 is being held fixed at its just identified estimate. Color notation: blue = true value: $\theta = (0.01, 0.1, 0.1)$; green = just identified estimate: $\hat{\theta}^{\mathrm{IL}} = (0.1011, 0.3898, 0.1014)$. BOTTOM: Contours of M-type objective function, as a function of θ_0 and θ_1. θ_2 is being held fixed at its just identified estimate. Color notation: blue = true value: $\theta = (0.01, 0.1, 0.1)$; green = just identified estimate: $\hat{\theta}^{\mathrm{IM}} = (0.0513, 0.1943, 0.1013)$.

Parameter estimates in linear model														
	$QMLE$	EXA	OLS	$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution, true $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Unconstrained estimation.														
θ_0	0.0388	0.0390	0.0388	0.0390	0.0384	0.0387	0.0390	0.0383	0.0385	0.0151	0.0390	0.0383	0.0383	0.0153
	(0.0288)	(0.0290)	(0.0288)	(0.0290)	(0.0284)	(0.0287)	(0.0290)	(0.0283)	(0.0285)	(0.0051)	(0.0290)	(0.0283)	(0.0283)	(0.0053)
	[0.0736]	[0.0741]	[0.0735]	[0.0741]	[0.0743]	[0.0757]	[0.0740]	[0.0743]	[0.0747]	[0.0493]	[0.0741]	[0.0743]	[0.0746]	[0.0516]
θ_1	0.3782	0.3802	0.3779	0.3802	0.3739	0.3757	0.3800	0.3730	0.3737	0.1501	0.3802	0.3731	0.3738	0.1500
	(0.2782)	(0.2802)	(0.2779)	(0.2802)	(0.2739)	(0.2757)	(0.2800)	(0.2730)	(0.2737)	(0.0501)	(0.2802)	(0.2731)	(0.2738)	(0.0500)
	[0.3740]	[0.3774]	[0.3741]	[0.3774]	[0.3751]	[0.3781]	[0.3775]	[0.3748]	[0.3752]	[0.2563]	[0.3774]	[0.3749]	[0.3753]	[0.2564]
θ_2	0.0997	0.1001	0.0997	0.1001	0.1001	0.1001	0.1001	0.1001	0.1001	0.1001	0.1001	0.1001	0.1001	0.1001
	(-0.0003)	(0.0001)	(-0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0024]	[0.0023]	[0.0023]	[0.0023]	[0.0025]
Simulations from exact solution, true $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.														
θ_0	0.0388	0.0390	0.0388	0.0390	0.0383	0.0383	0.0390	0.0383	0.0408	0.0168	0.0390	0.0383	0.0409	0.0169
	(0.0288)	(0.0290)	(0.0288)	(0.0290)	(0.0283)	(0.0283)	(0.0290)	(0.0283)	(0.0308)	(0.0068)	(0.0290)	(0.0283)	(0.0309)	(0.0069)
	[0.0736]	[0.0741]	[0.0735]	[0.0741]	[0.0743]	[0.0744]	[0.0740]	[0.0742]	[0.0885]	[0.0446]	[0.0741]	[0.0742]	[0.0924]	[0.0446]
θ_1	0.3782	0.3802	0.3779	0.3802	0.3732	0.3738	0.3802	0.3729	0.3736	0.1693	0.3802	0.3730	0.3736	0.1694
	(0.2782)	(0.2802)	(0.2779)	(0.2802)	(0.2732)	(0.2738)	(0.2802)	(0.2729)	(0.2736)	(0.0693)	(0.2802)	(0.2730)	(0.2736)	(0.0694)
	[0.3740]	[0.3774]	[0.3741]	[0.3774]	[0.3749]	[0.3753]	[0.3774]	[0.3749]	[0.3753]	[0.2458]	[0.3774]	[0.3749]	[0.3753]	[0.2458]
θ_2	0.0997	0.1001	0.0997	0.1001	0.1001	0.1001	0.1001	0.1002	0.1002	0.1003	0.1001	0.1001	0.1001	0.1001
	(-0.0003)	(0.0001)	(-0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0023]	[0.0024]	[0.0024]	[0.0029]	[0.0023]	[0.0023]	[0.0024]	[0.0024]
Simulations from exact solution, true $\theta = (0, 0.66, 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.														
θ_0	0.0708	0.0719	0.0701	0.0716	0.0704	0.0695	0.0708	0.0696	0.0688	0.0519	0.0716	0.0704	0.0695	0.0527
	(0.0708)	(0.0719)	(0.0701)	(0.0716)	(0.0704)	(0.0695)	(0.0708)	(0.0696)	(0.0688)	(0.0519)	(0.0716)	(0.0704)	(0.0695)	(0.0527)
	[2.2726]	[2.2989]	[2.2799]	[2.2996]	[2.3268]	[2.3259]	[2.3073]	[2.3338]	[2.3330]	[1.8751]	[2.2996]	[2.3268]	[2.3259]	[1.8675]
θ_1	0.8910	0.9002	0.8913	0.9002	0.8937	0.8935	0.9005	0.8939	0.8938	0.6768	0.9002	0.8937	0.8935	0.6766
	(0.2310)	(0.2402)	(0.2313)	(0.2402)	(0.2337)	(0.2335)	(0.2405)	(0.2339)	(0.2338)	(0.0168)	(0.2402)	(0.2337)	(0.2335)	(0.0166)
	[0.4085]	[0.4198]	[0.4092]	[0.4199]	[0.4227]	[0.4231]	[0.4206]	[0.4234]	[0.4237]	[0.3657]	[0.4199]	[0.4227]	[0.4231]	[0.3651]
θ_2	7.0152	7.0785	7.0152	7.0785	7.0814	7.0814	7.0785	7.0814	7.0815	7.0759	7.0785	7.0814	7.0814	7.0759
	(-0.0558)	(0.0074)	(-0.0558)	(0.0074)	(0.0104)	(0.0104)	(0.0074)	(0.0104)	(0.0104)	(0.0048)	(0.0074)	(0.0104)	(0.0104)	(0.0048)
	[0.1692]	[0.1625]	[0.1692]	[0.1625]	[0.1662]	[0.1662]	[0.1625]	[0.1662]	[0.1662]	[0.1658]	[0.1625]	[0.1662]	[0.1662]	[0.1658]

Table 4: Mean, empirical bias ( ) and RMSE [ ] of estimates for 1000 Monte Carlo simulations.

Average estimation time										
$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution. True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Unconstrained estimation.										
0.03	5.15	5.92	0.09	4.16	5.79	11.40	0.01	4.03	5.65	11.06
Simulations from exact solution. True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.03	2.99	4.66	0.08	2.50	4.27	11.64	0.01	2.40	4.23	11.57
Simulations from exact solution. True $\theta = (0, 0.66, 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.03	3.09	4.72	0.08	2.69	4.60	5.75	0.01	2.61	4.61	5.55

Table 5: Average estimation time in seconds.

Rejection frequencies of likelihood ratio type tests										
$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution. True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Unconstrained estimation.										
0.767	0.760	0.728	0.129	0.114	0.116	0.419	0.121	0.112	0.113	0.415
Simulations from exact solution. True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.767	0.760	0.728	0.129	0.114	0.116	0.419	0.121	0.112	0.113	0.415
Simulations from exact solution. True $\theta = (0, 0.66, 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.336	0.330	0.323	0.077	0.079	0.080	0.137	0.073	0.071	0.073	0.130

Table 6: Empirical rejection frequencies of likelihood ratio type tests at 5% nominal level for 1000 Monte Carlo simulations.

Rejection frequencies of likelihood ratio type tests of θ_1										
$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution. True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Unconstrained estimation.										
0.698	0.640	0.593	0.197	0.169	0.175	0.130	0.210	0.176	0.183	0.149
Simulations from exact solution. True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.698	0.640	0.593	0.200	0.172	0.176	0.132	0.210	0.176	0.183	0.149
Simulations from exact solution. True $\theta = (0, 0.66, 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.263	0.251	0.244	0.085	0.076	0.076	0.079	0.086	0.078	0.078	0.083

Table 7: Empirical rejection frequencies of likelihood ratio type tests of θ_1 at 5% nominal level for 1000 Monte Carlo simulations.

Parameter estimates in linear overID model														
	$QMLE$	EXA	OLS	$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution, true $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.														
θ_1	0.3782	0.1946	0.1932	1.4201	1.5582	1.5342	0.1938	0.1885	0.1899	0.1271	0.1952	0.1901	0.1914	0.1283
	(0.2782)	(0.0946)	(0.0932)	(1.3201)	(1.4582)	(1.4342)	(0.0938)	(0.0885)	(0.0899)	(0.0271)	(0.0952)	(0.0901)	(0.0914)	(0.0283)
	[0.3740]	[0.2021]	[0.2019]	[2.5847]	[2.8086]	[2.8155]	[0.2010]	[0.1991]	[0.1991]	[0.1278]	[0.2024]	[0.2005]	[0.2004]	[0.1287]
Simulations from exact solution, true $\theta = (0, (1 - 0.9868) * 50 = 0.66, \sqrt{50} = 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.														
θ_1	0.8910	0.7679	0.7621	1.3746	1.3842	1.3893	0.7653	0.7573	0.7571	0.5429	0.7696	0.7619	0.7617	0.5474
	(0.2310)	(0.1079)	(0.1021)	(0.7146)	(0.7242)	(0.7293)	(0.1053)	(0.0973)	(0.0971)	(-0.1171)	(0.1096)	(0.1019)	(0.1017)	(-0.1126)
	[0.4085]	[0.3170]	[0.3122]	[1.6325]	[1.6913]	[1.6963]	[0.3158]	[0.3171]	[0.3177]	[0.3346]	[0.3185]	[0.3200]	[0.3206]	[0.3349]

Table 8: Mean, empirical bias ( ) and RMSE [ ] of estimates for 1000 Monte Carlo simulations.

Average estimation time										
$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.03	1.48	3.03	0.03	1.21	2.44	3.08	0.02	1.27	2.39	3.03
Simulations from exact solution True $\theta = (0, (1-0.9868)*50 = 0.66, \sqrt{50} = 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.02	0.63	0.88	0.03	0.52	0.87	1.12	0.01	0.51	0.85	1.11

Table 9: Average estimation time in seconds.

Rejection frequencies of over-identification tests										
$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.439	0.419	0.408	0.152	0.150	0.153	0.254	0.147	0.134	0.137	0.246
Simulations from exact solution True $\theta = (0, (1-0.9868)*50 = 0.66, \sqrt{50} = 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.131	0.121	0.121	0.079	0.074	0.073	0.079	0.082	0.070	0.070	0.076

Table 10: Empirical rejection frequencies of over-identification tests at 5% nominal level for 1000 Monte Carlo simulations.

Rejection frequencies of likelihood ratio type tests										
$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
Simulations from exact solution True $\theta = (0.01, 0.1, 0.1)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.719	0.724	0.694	0.068	0.077	0.076	0.302	0.075	0.081	0.083	0.304
Simulations from exact solution True $\theta = (0, (1-0.9868)*50 = 0.66, \sqrt{50} = 7.071)$, $T = 20$, $\Delta = 1/50$. Constrained estimation.										
0.379	0.364	0.355	0.051	0.053	0.052	0.129	0.049	0.056	0.054	0.123

Table 11: Empirical rejection frequencies of likelihood ratio type tests at 5% nominal level for 1000 Monte Carlo simulations.