Simulation-Based Estimation of Continuous Time Models in R

R/Finance 2010

Eric Zivot
University of Washington

Joint work with:

Peter Fuleky
University of Hawaii

Introduction

Goal: Estimate parameters of continuous time diffusion model from discretely sampled data

$$dy_t = \alpha(y_t,\theta)dt + \sigma(y_t,\theta)dW_t,\ dW_t \sim \text{iid } N(0,dt)$$

Examples

OU: $dy_t = (\theta_0 - \theta_1 y_t)dt + \theta_2 dW_t,\ \alpha(\mathrm{y}_t,\theta) = \theta_0 - \theta_1 y_t,\ \sigma(\mathrm{y}_t,\theta) = \theta_2$

CIR: $dy_t = (\theta_0 - \theta_1 y_t)dt + \theta_2\sqrt{y_t}\,dW_t,\ \alpha(\mathrm{y}_t,\theta) = \theta_0 - \theta_1 y_t,\ \sigma(\mathrm{y}_t,\theta) = \theta_2\sqrt{y_t}$

$$GCIR:\ dy_t = (\theta_0 - \theta_1 y_t)dt + \theta_2 y_t^{\gamma} dW_t\ ,$$

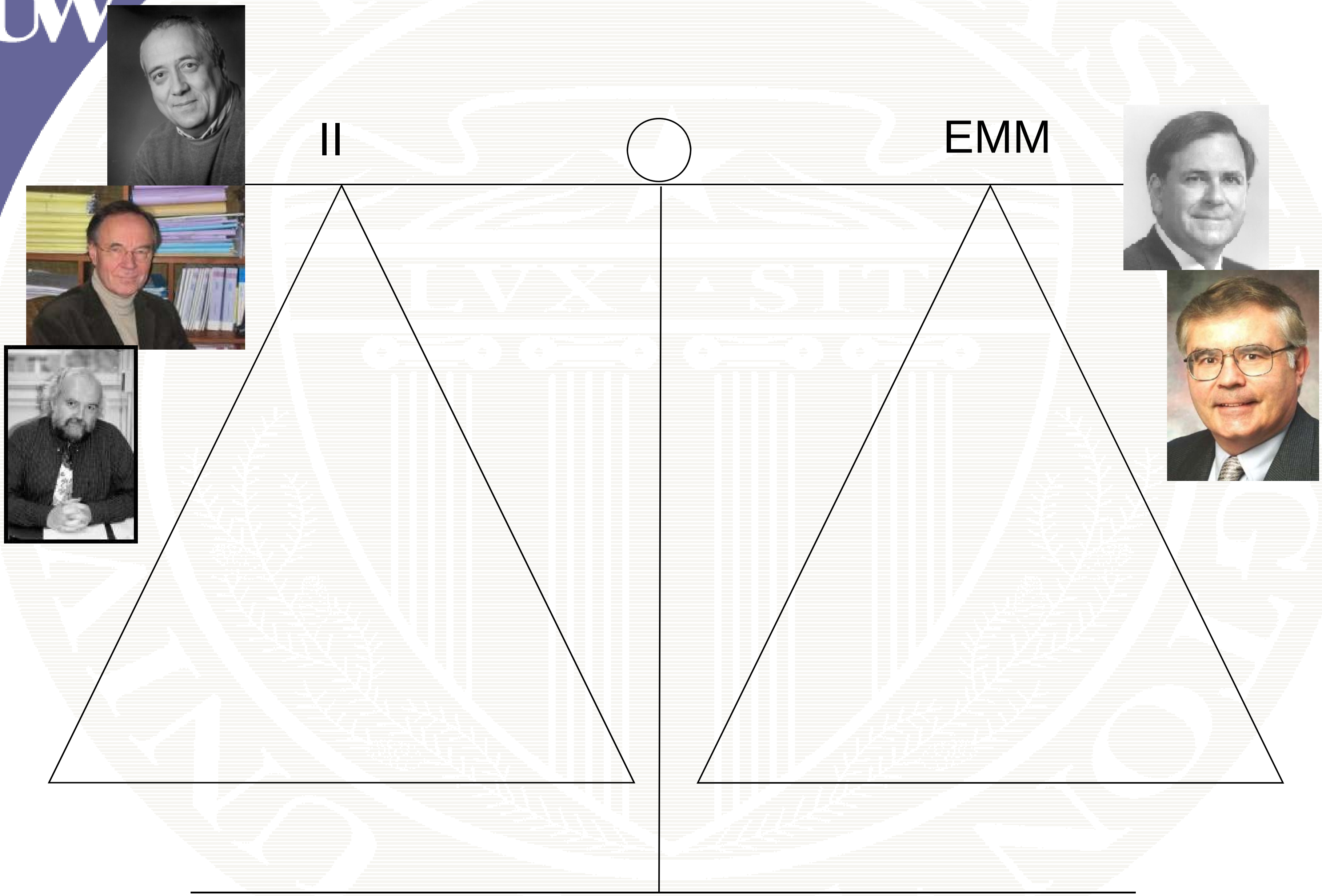
$$\alpha(\mathrm{y}_t,\theta) = \theta_0 - \theta_1 y_t,\ \sigma(\mathrm{y}_t,\theta) = \theta_2 y_t^{\gamma}$$

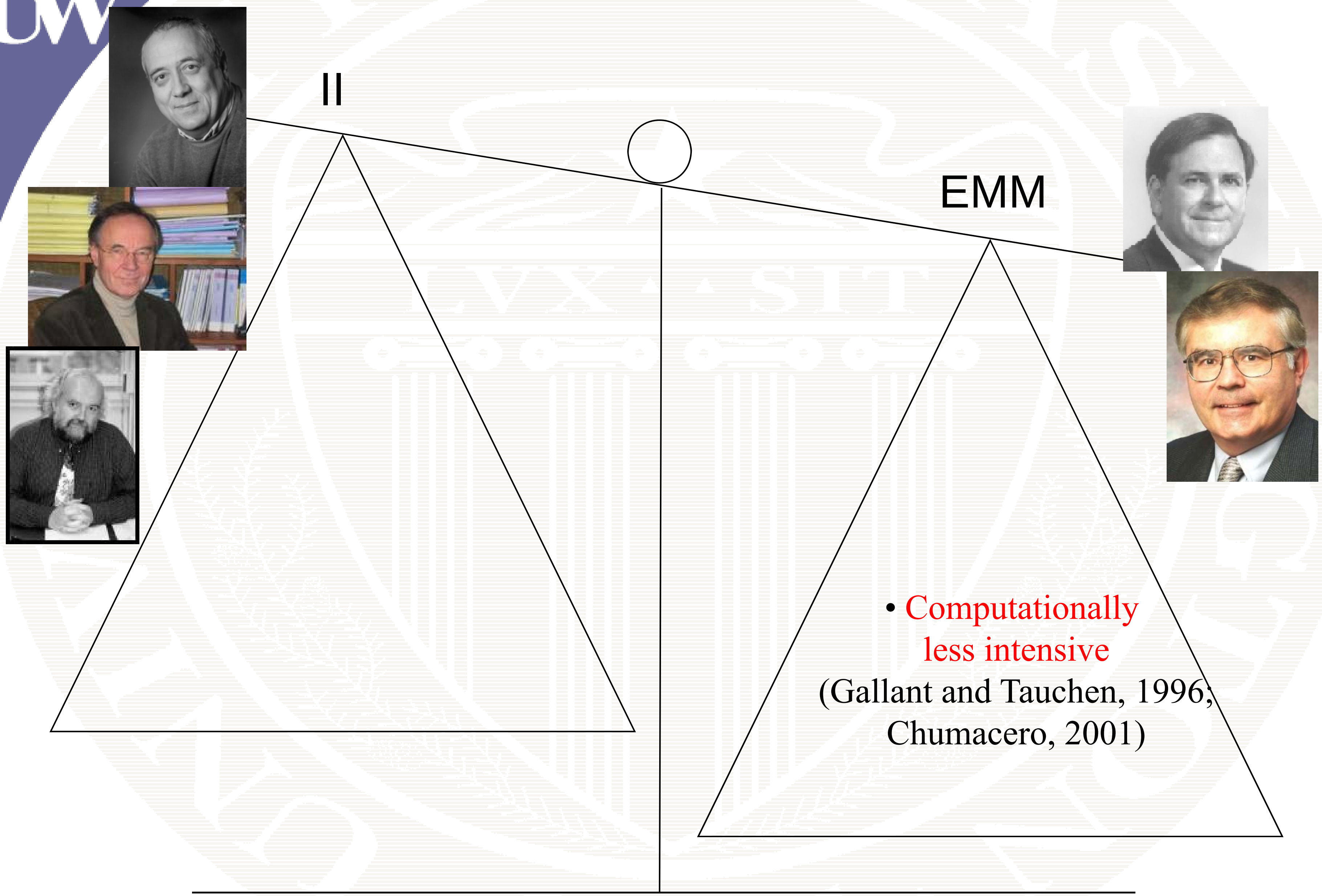
Estimation Methods

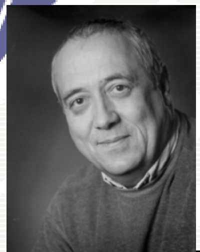
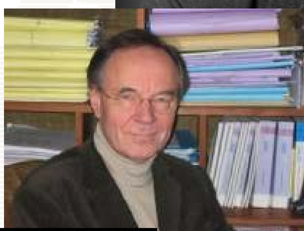
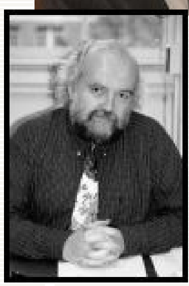
- MLE – often not feasible
- MLE of approximated model – difficult
- QMLE of discretized model – easy but biased
- GMM – inefficient and biased
- Bayesian MCMC Methods - promising
- Indirect Inference – Corrects bias in QMLE
  - focus of talk

Indirect Inference

- Distance-based methodology (aka II) developed by Smith (1993), Gourierioux, Monfort, and Renault (1993)
- Score-based methodology (aka EMM) developed by Gallant and Tauchen (1996)

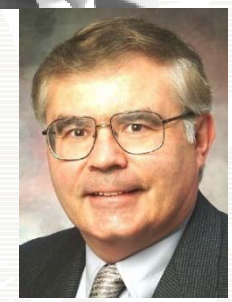

II
EMM


II
EMM
• Computationally less intensive
(Gallant and Tauchen, 1996; Chumacero, 2001)

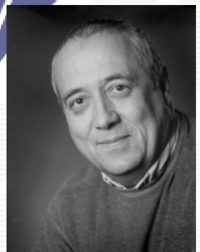
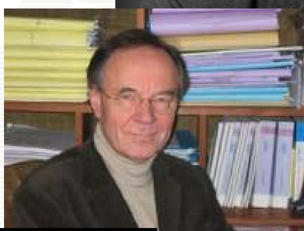
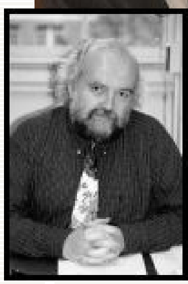
II

EMM

• Smaller bias and MSE in MA models (Ghysels, Khalaf and Vodounou, 2003)

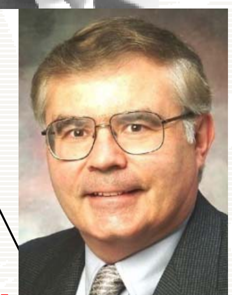
• Computationally less intensive (Gallant and Tauchen, 1996; Chumacero, 2001)

II

- Smaller bias and MSE in MA models (Ghysels, Khalaf and Vodounou, 2003)
- More accurate inference for AR models (Duffee and Stanton, 2008)

EMM

- Computationally less intensive (Gallant and Tauchen, 1996; Chumacero, 2001)

Research Agenda and R Contribution

- Implement indirect inference estimation techniques for some commonly used continuous time models (e.g., OU, CIR, etc.)
- Provide systematic comparison and evaluation of different estimators
- Create indirectInference R package
- Give practical advice on use of techniques

Indirect Inference Set-up

$\{y_t\}_{t=\Delta}^{n\Delta}$ observations with observation interval Δ

Structural model: $F_\theta,\ \theta \in \mathbb{R}^p$, stationary and ergodic

Auxiliary model: $\tilde{F}_\mu,\ \mu \in \mathbb{R}^r,\ r \geq p$

$$\tilde{\mu} = \arg\max_\mu \ Q_n\left(\{y_t\}_{t=\Delta}^{n\Delta}, \mu\right), \text{ where}$$

$$\tilde{Q}_n = \frac{1}{n-m} \sum_{t=(m+1)\Delta}^{n\Delta} \tilde{f}(y_t; x_{t-\Delta}, \mu),\ x_{t-\Delta} = \{y_i\}_{i=t-m\Delta}^{t-\Delta}$$

$\tilde{f}(y_t; x_{t-\Delta}, \mu) =$ conditional log density of y_t for the model $\tilde{F}_\mu$

$$\mu(\theta) = \arg\max_\mu \ E_{F_\theta}[\tilde{f}(y_t; x_{t-\Delta}, \mu)] = p\lim\ \tilde{\mu} \text{ under } F_\theta$$

$$= \text{binding function}$$

Example: OU Model

$$F_\theta : dy_t = (\theta_0 - \theta_1)dt + \theta_2 dW_t, \quad \theta_i > 0,\ p = 3,\ \Delta = 1/52$$

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

$$\tilde{F}_\mu : y_t = y_{t-\Delta} + (\mu_0 + \mu_1 y_{t-\Delta})\Delta + \mu_2\sqrt{\Delta}\varepsilon_{t-\Delta}$$

$$= \mu_0\Delta + (1 - \mu_2\Delta)y_{t-\Delta} + \mu_2\sqrt{\Delta}\varepsilon_{t-\Delta},\ \varepsilon_{t-\Delta} \sim iid\ N(0,1),\ r = 3$$

$$\mu(\theta) = p\lim \tilde{\mu} \neq \theta$$

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

Example: OU Model

- Estimating the “crude Euler” auxiliary model leads to biased estimates (Lo, 1988)
  - Asymptotic discretization bias = $\mu(\theta) - \theta$
  - $\mu(\theta) - \theta \rightarrow 0$ as $\Delta \rightarrow 0$
- $\mu(\theta)$ is invertible giving analytic II estimates

$$\hat{\theta}^{II} = \mu^{-1}(\tilde{\mu})$$

$$\hat{\theta}_0^{II} = \frac{\tilde{\mu}_0}{\tilde{\mu}_1 \Delta} \ln(1 - \tilde{\mu}_1 \Delta), \; \hat{\theta}_1^{II} = \frac{-1}{\Delta} \ln(1 - \tilde{\mu}_1 \Delta),$$

$$\hat{\theta}_2^{II} = \tilde{\mu}_2 \sqrt{\frac{2 \ln(1 - \tilde{\mu}_1 \Delta)}{1 - e^{\ln(1 - \tilde{\mu}_1 \Delta)}}}$$

Non-simulation based Estimation

- Assume $\mu(\theta)$ is known (very rare!)
- EMM is GMM with population moment

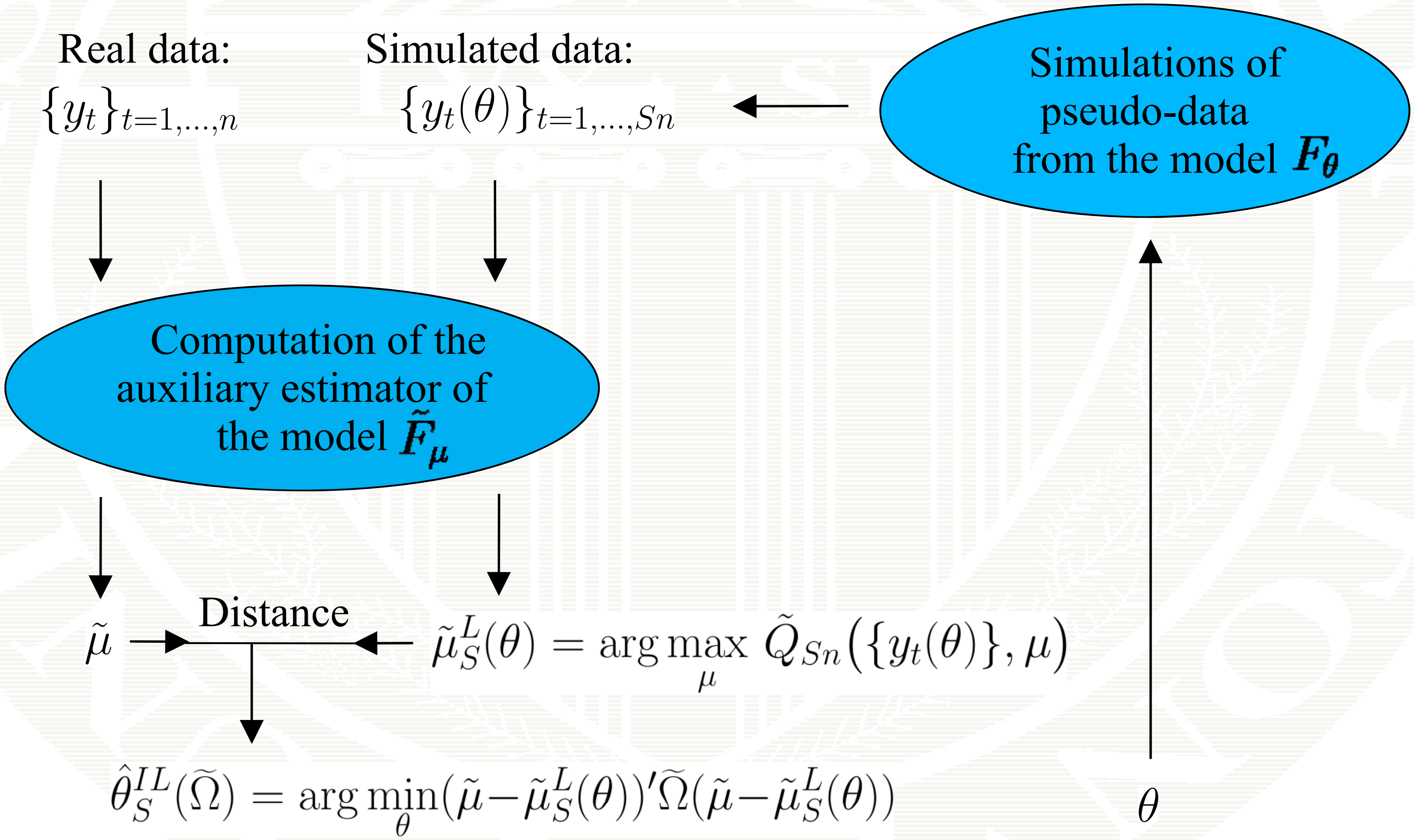
$$E_{F_\theta}\left[\frac{\partial \tilde{f}(y_t; x_{t-\Delta}, \mu)}{\partial \mu}\right]_{\mu=\mu(\theta)} = 0$$

- II minimizes distance between $\mu(\theta)$ and $\tilde{\mu}$
- Asymptotically equivalent to MLE when auxiliary model encompasses structural model

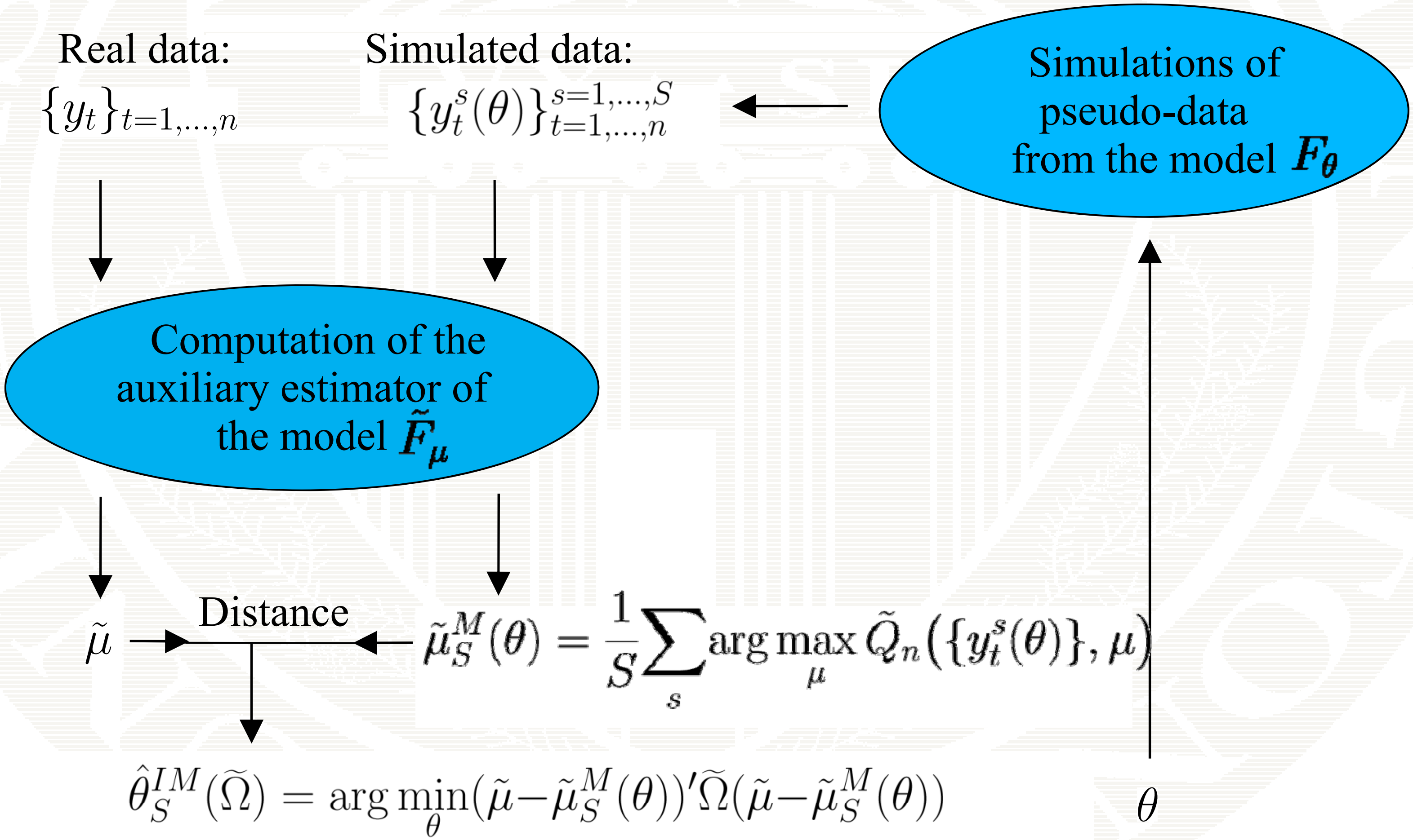
Simulation-based EMM and II

- $\mu(\theta)$ is unknown
- $\tilde{\mu}$ is used to estimate $\mu(\theta_{true})$
- With EMM, simulations for a given θ are used to approximate the expectation of sample score
- With II, simulations are used to approximate $\mu(\theta)$ for any θ
- Gouriéroux and Monfort (1996) describe 3 types of II estimators and 2 types of EMM estimators

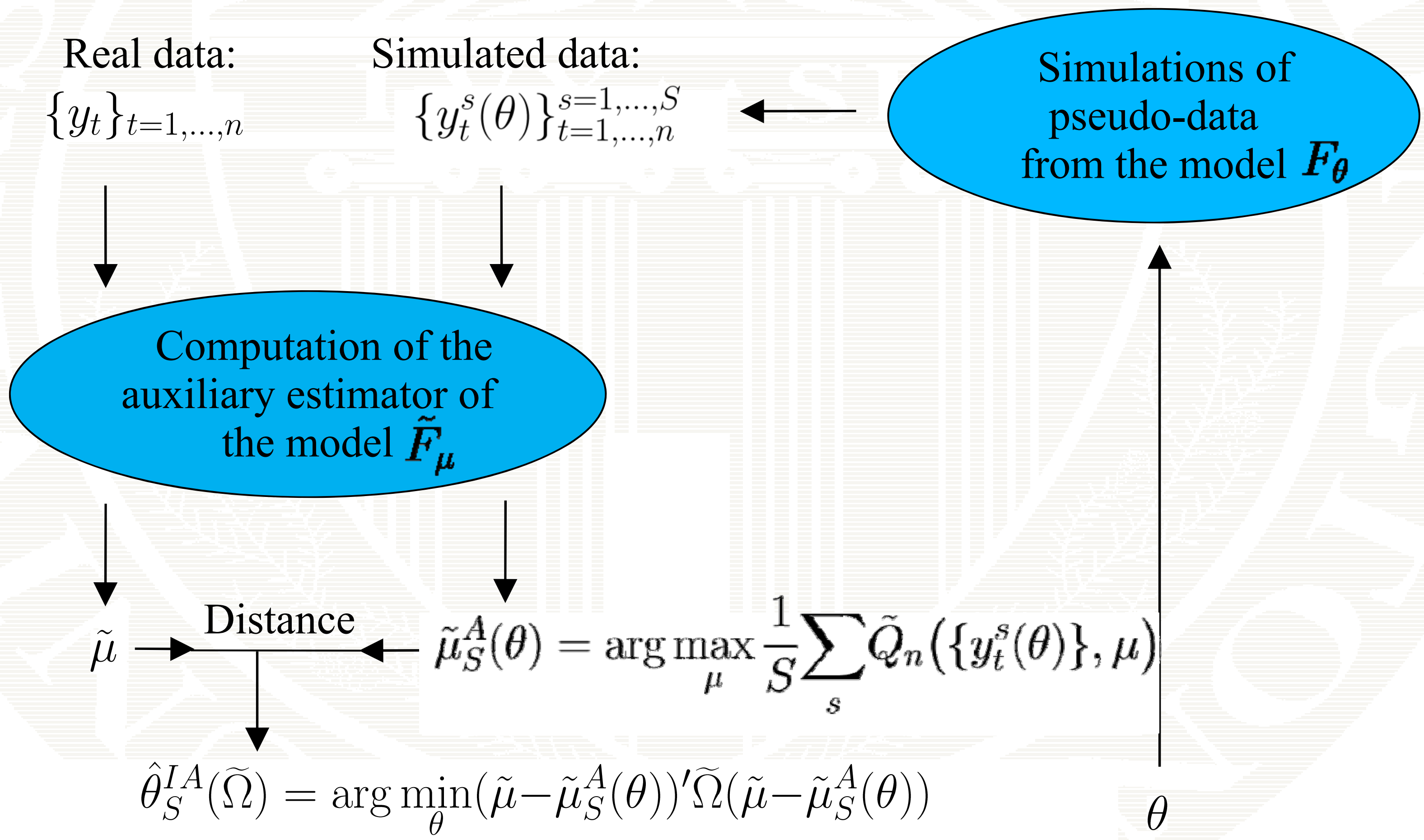
1st Type of Simulation-based II: IL



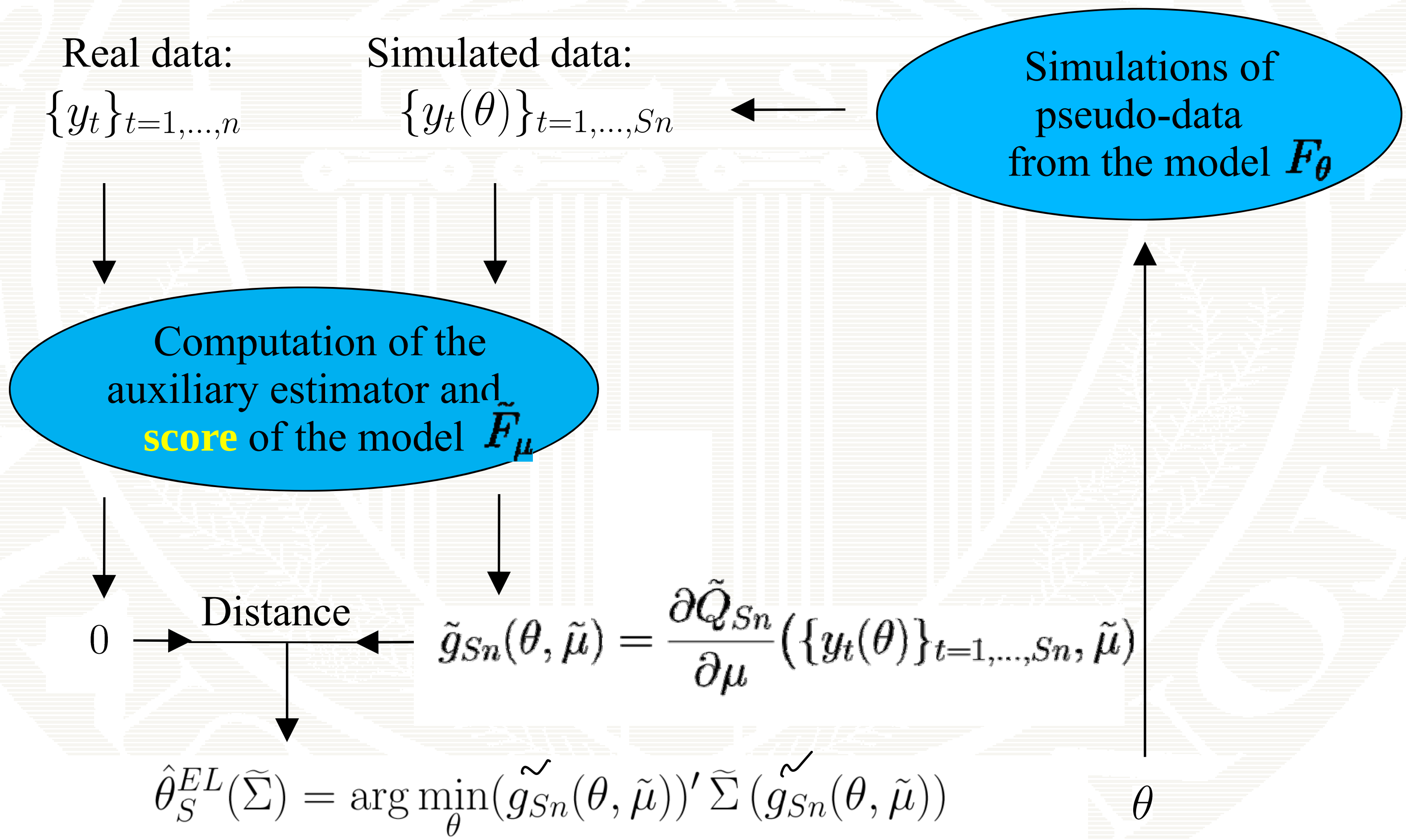
2^{nd} Type of II: IM



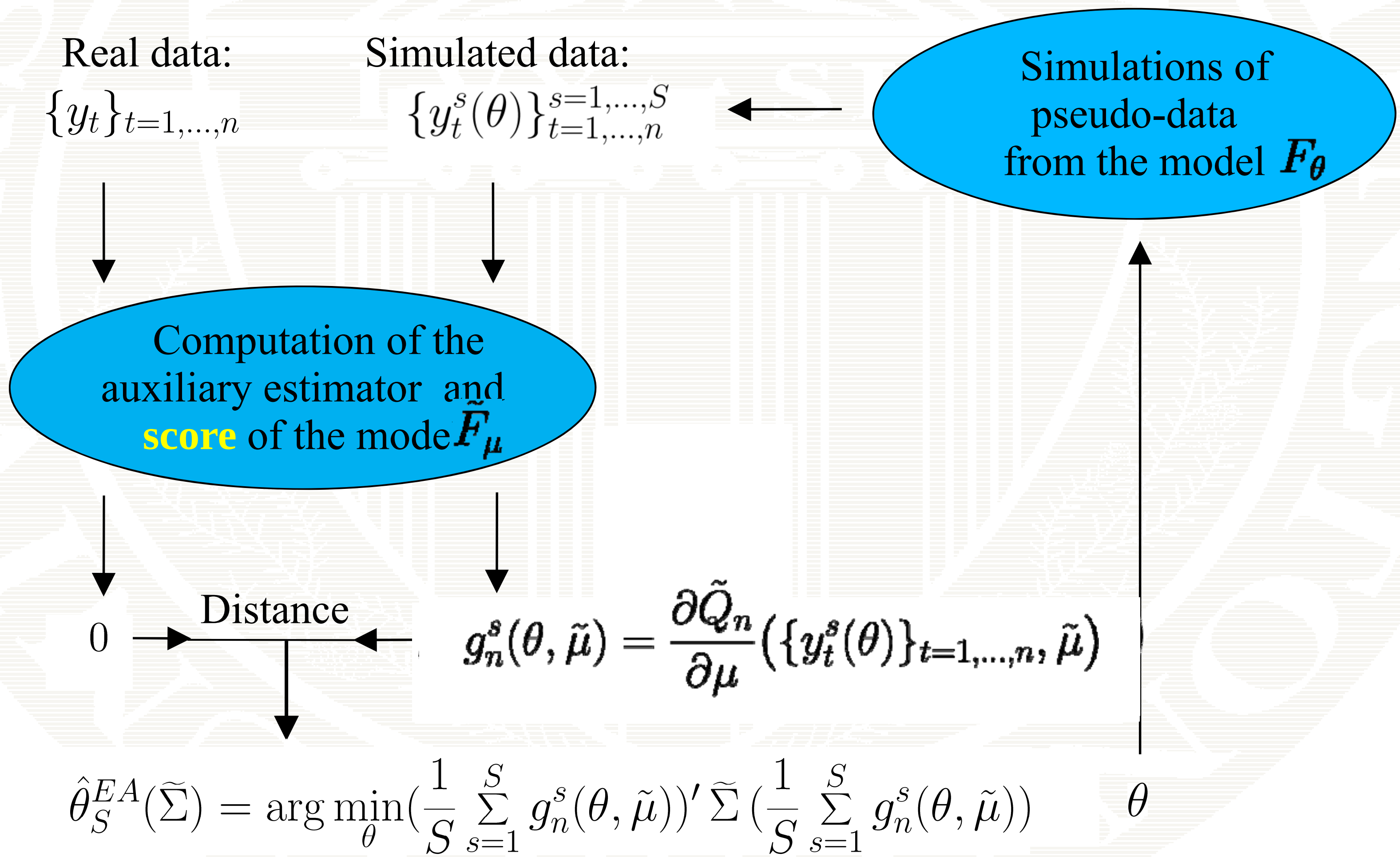
3rd Type of II: IA



1st Type of EMM: EL



2nd Type of EMM: EA



R Implementation of II

- Estimate Euler auxiliary model parameters μ from observed data $\{y_t\}$ by QMLE

$$y_{t+\Delta} - y_t = \alpha(y_t, \mu)\Delta + \sigma(y_t, \mu)\sqrt{\Delta} z_t, \; z_t \sim \text{iid } N(0,1)$$

$$\tilde{\mu} = \arg\max_{\mu} \; Q_n\left(\{y_t\}_{t=\Delta}^{n\Delta}, \mu\right), \text{ where}$$

$$\tilde{Q}_n = \frac{1}{n-m} \sum_{t=(m+1)\Delta}^{n\Delta} \tilde{f}(y_t; x_{t-\Delta}, \mu), \; x_{t-\Delta} = \{y_i\}_{i=t-m\Delta}^{t-\Delta}$$

  - Use function `EULERloglik()` from R package `sde`
  - Use R function `optim()`

R Implementation of II

- Simulate from F_θ
  - In general, cannot do exact simulations because transition density is not known
  - Simulate from very fine Euler discretization
  - Use function `sde.sim()` from R package `sde`
  - Use custom C code for fast simulation
  - Need to worry about "inadmissible" or "explosive" simulations from inappropriate θ – need to "bullet proof" the simulator

R Implementation of II

- For distance-based II, estimate binding function $\mu(\theta)$ from simulated data $\{y_t^s(\theta)\}$

$$\tilde{\mu}_S^L = \arg\max_{\mu} \; Q_{Sn}\left(\{y_t^s(\theta)\}, \mu\right), \text{ where}$$

$$\tilde{Q}_{Sn} = \frac{1}{n'} \sum_{t=(m+1)\Delta}^{Sn\Delta} \tilde{f}(y_t^s(\theta); x_{t-\Delta}^s(\theta), \mu),$$

  – Use same random number seed for all θ

R Implementation of II

- For score-based II, estimate auxiliary score from simulated data $\{y_t^s(\theta)\}$ and evaluate at auxiliary parameter estimate

$$g_{Sn}(\{y_t^s(\theta)\},\tilde{\mu}) = \frac{\partial Q_{Sn}(\{y_t^s(\theta)\},\tilde{\mu})}{\partial \mu}$$

  - User specified function to evaluate score function
  - Use same random number seed for all θ

R Implementation of II

- For distance-based II, estimate θ

$$\hat{\theta}_S^{II} = \arg\min_{\theta} \ (\tilde{\mu} - \tilde{\mu}_S^i(\theta))'\tilde{\Omega}(\tilde{\mu} - \tilde{\mu}_S^i(\theta)), \ i = L, M$$

- For score-based II, estimate

$$\hat{\theta}_S^{EMM} = \arg\min_{\theta} \ g_{Sn}(\theta, \tilde{\mu})'\tilde{\Sigma}\ g_{Sn}(\theta, \tilde{\mu})'$$

- If $p = r$ then use identity matrix for weight matrix
- For optimization, use R function `optim()` with Nelder-Meade simplex algorithm

Illustration

- OU Process calibrated to US interest rates used by Phillips and Yu (2009)

$$\theta_0 = 0.01,\ \theta_1 = 0.10,\ \theta_2 = 0.10$$

$$\frac{\theta_0}{\theta_1} = 0.10 = \text{ annualized avg rate,}$$

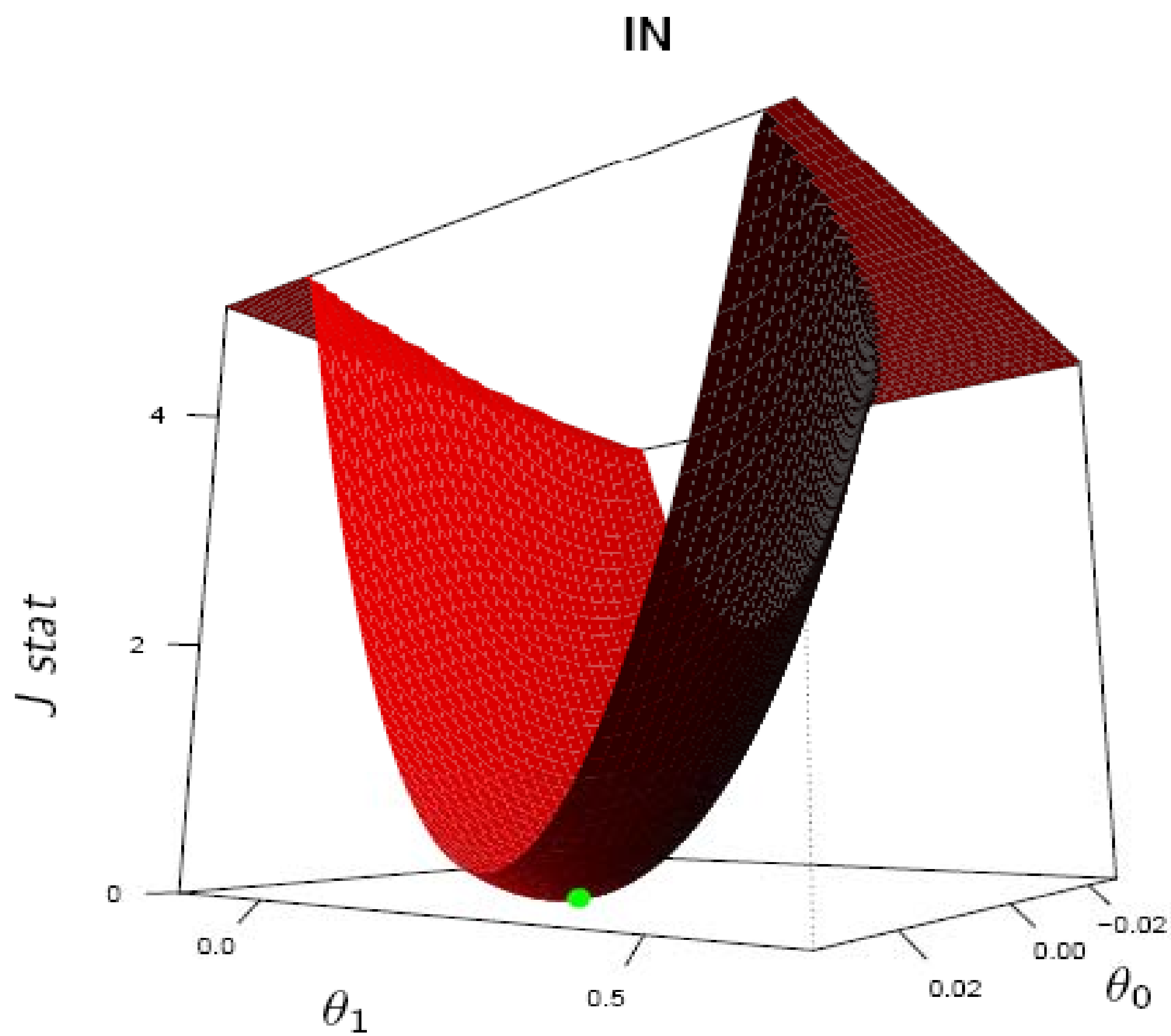
$$\theta_1 = 0.1 \Rightarrow \text{ 7 year half of rate shock}$$

$$\theta_2 = 0.10 = \text{ annualized rate volatility}$$

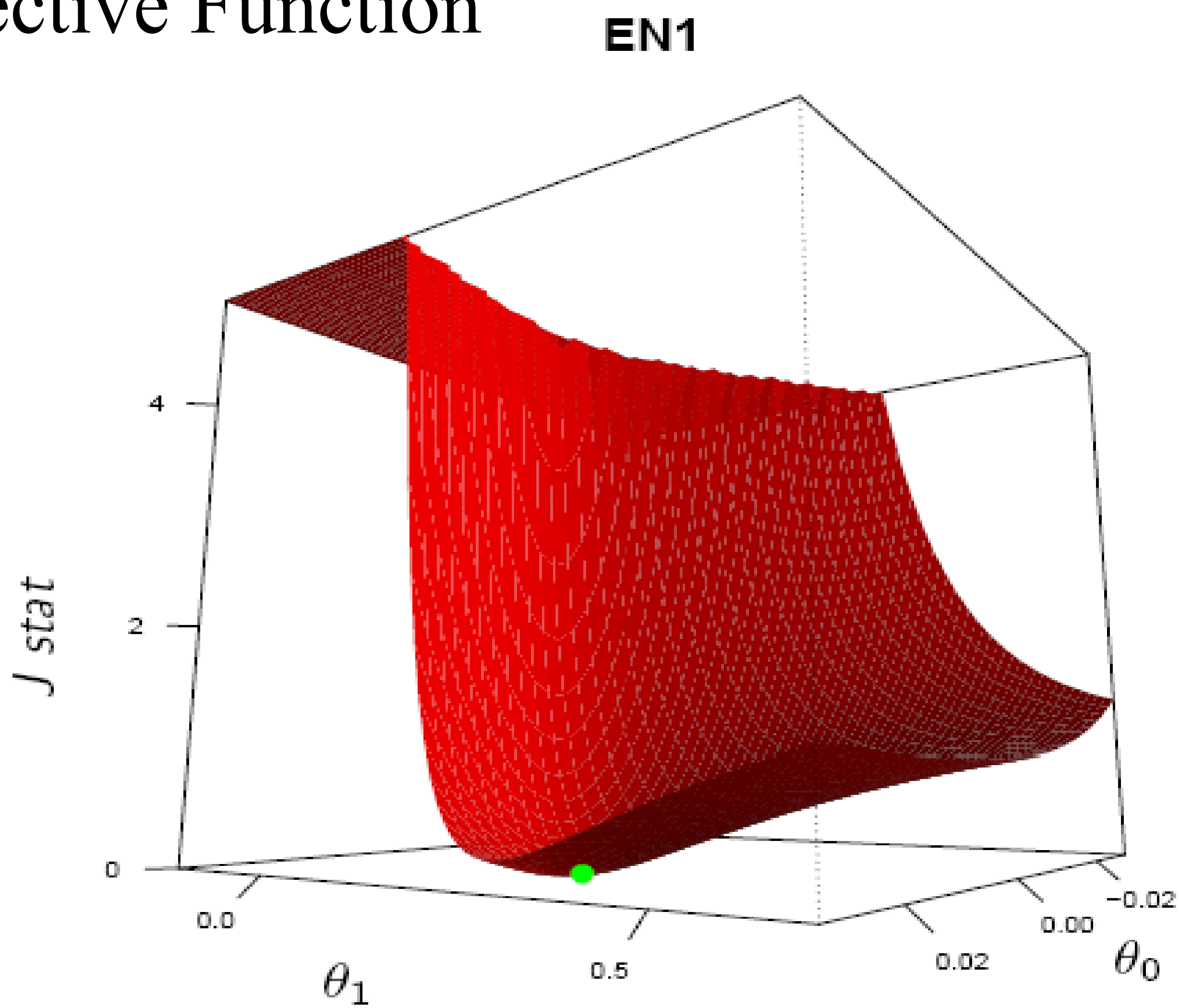
$$T = 19.23,\ \Delta = 1/52 \Rightarrow n = 1000$$

- θ_1 is the most difficult parameter to estimate

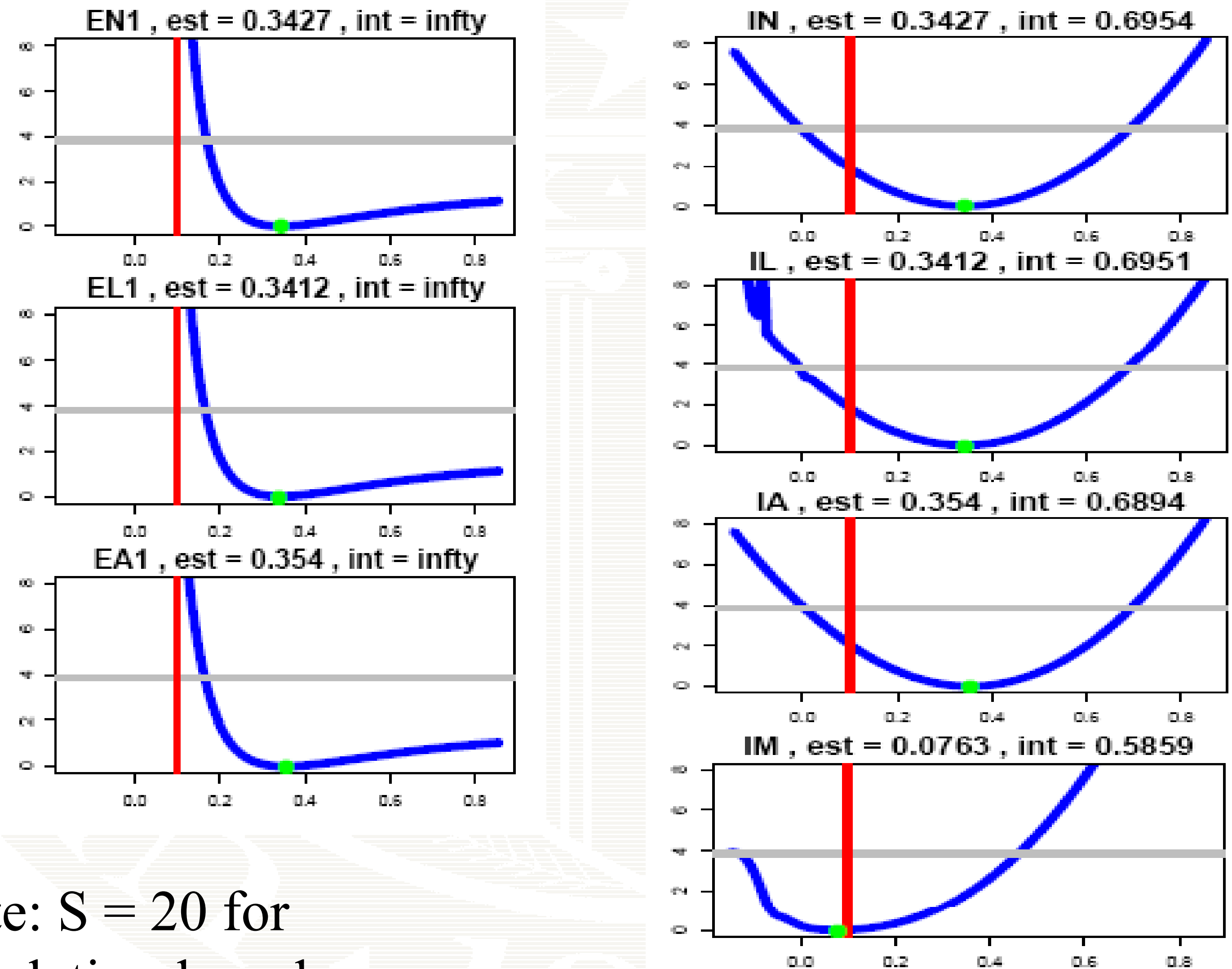
Shape of distance-based II Objective function



Shape of Gallant-Tauchen Score-based II Objective Function

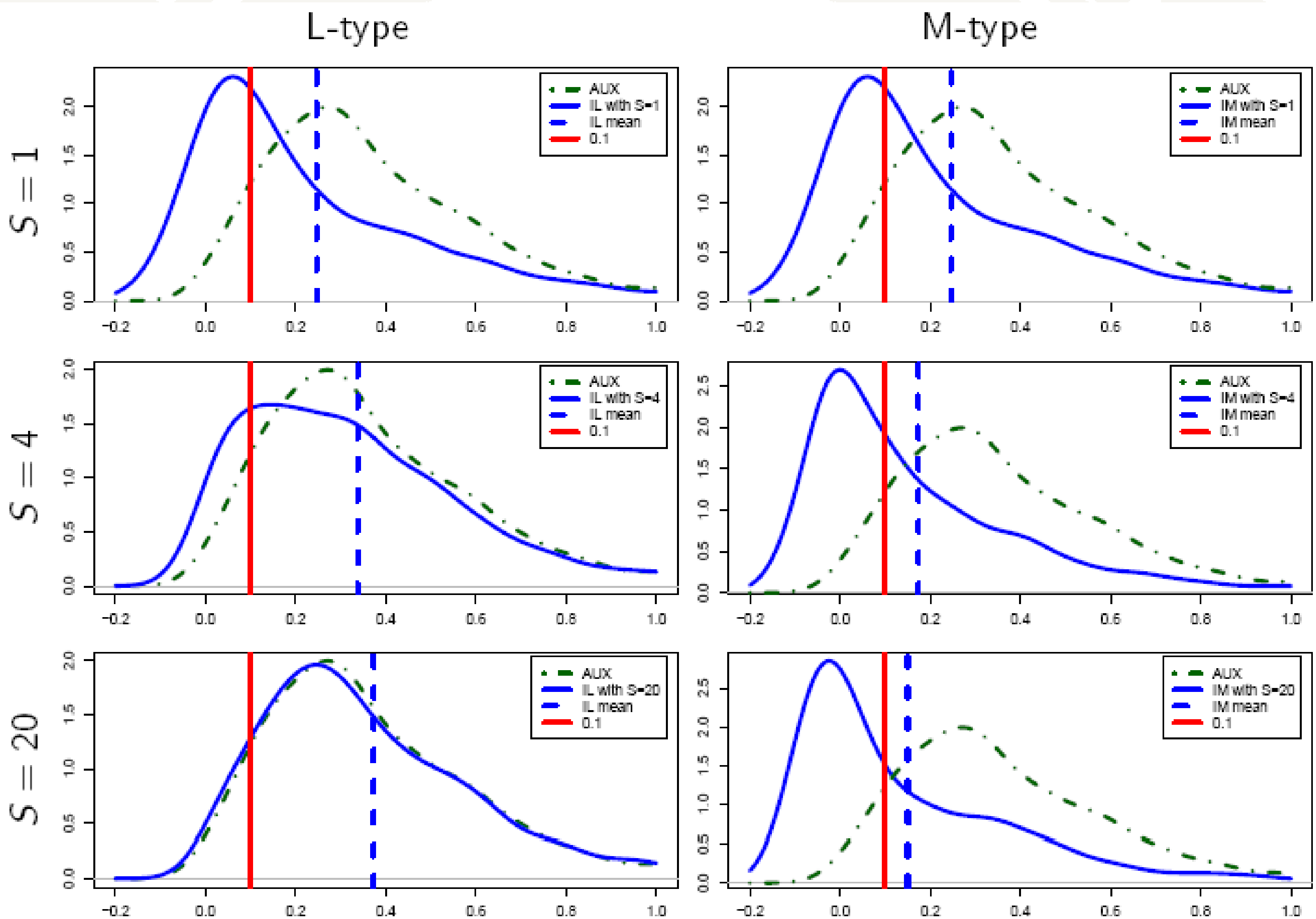


95% Confidence Intervals for θ_1

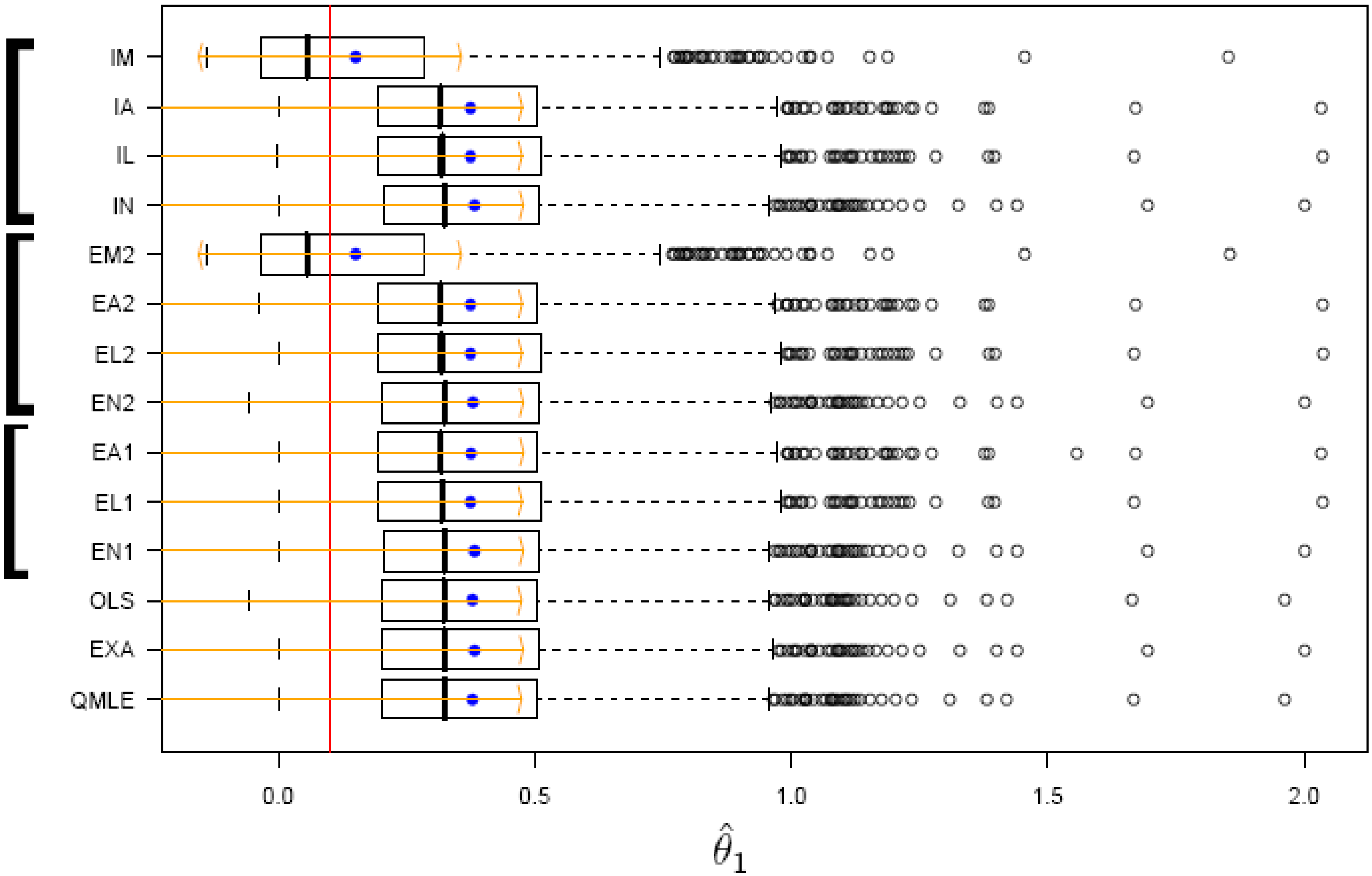


Note: S = 20 for simulation-based estimates

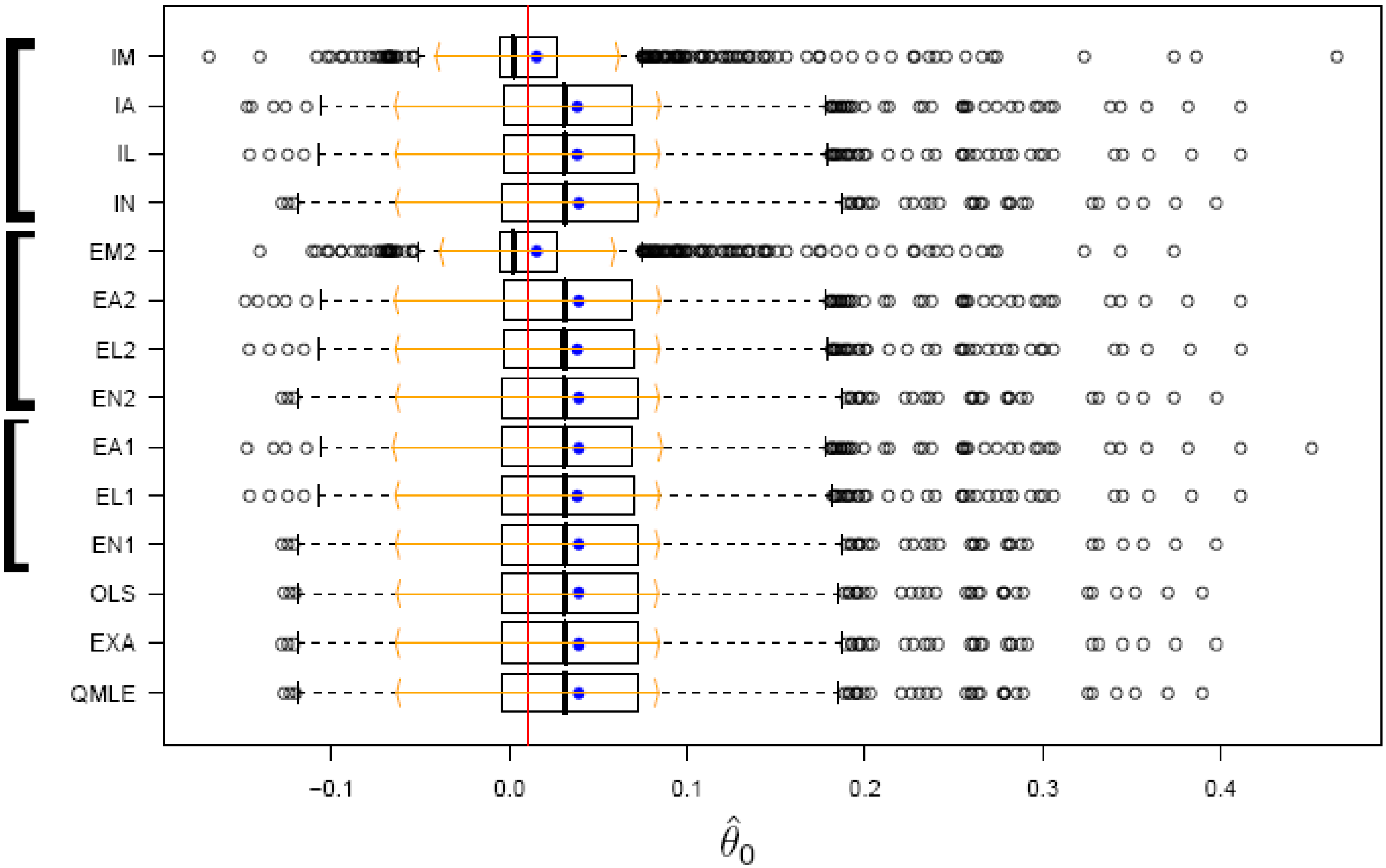
Impact of simulation Size S on Densities of θ_1 Estimates



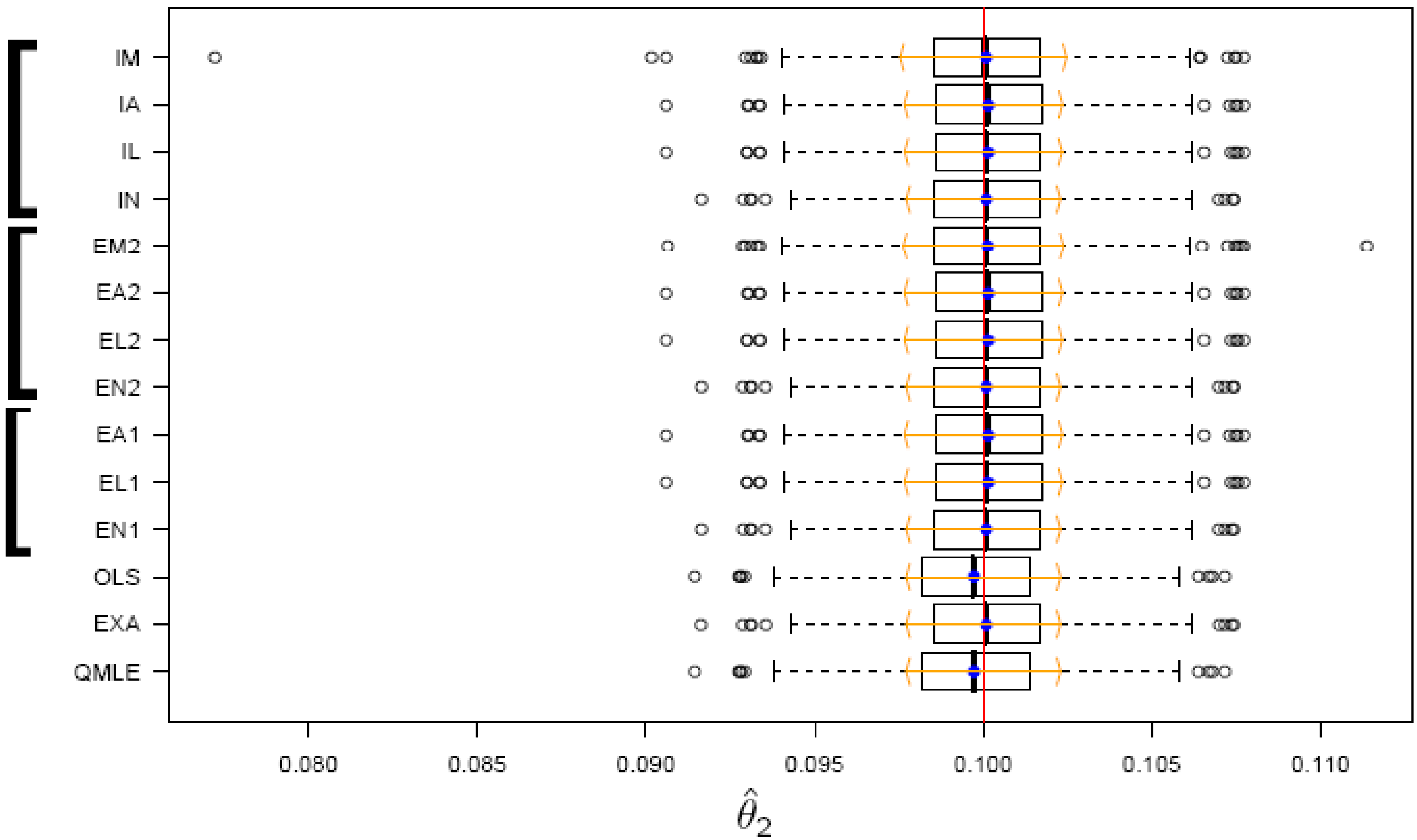
Distribution of $\hat{\theta}_1$: $\theta = (0.01, 0.1, 0.1)$, n=1000, $\Delta = 1/52$



Distribution of $\hat{\theta}_0$: $\theta = (0.01, 0.1, 0.1)$, n=1000, $\Delta = 1/52$



Distribution of $\hat{\theta}_2$: $\theta = (0.01, 0.1, 0.1)$, n=1000, $\Delta = 1/52$



Rejection frequencies of likelihood ratio type tests

at 5% nominal level of significance for 1000 Monte Carlo simulations.

$EN1$	$EL1$	$EA1$	$EN2$	$EL2$	$EA2$	$EM2$	IN	IL	IA	IM
True $\theta = (0.01, 0.1, 0.1)$, $n = 1000$, $\Delta = 1/52$.										
Joint Test										
0.767	0.760	0.728	0.129	0.114	0.116	0.419	0.121	0.112	0.113	0.415
Simple Test of $\theta_1 = 0.1$										
0.698	0.640	0.593	0.197	0.169	0.175	0.130	0.210	0.176	0.183	0.149

Research in Progress

- Fuleky, P., and Zivot, E. (2010). Further Evidence on Simulation Inference for Near Unit Root Processes with Implications for Term Structure Estimation. Manuscript in preparation.
- Fuleky, P., and Zivot, E. (2010). Indirect Inference Based on the Score. Manuscript in preparation.
- Fuleky, P., and Zivot, E. (2010). indirectInference: R package for indirect inference.

References

- 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.
- Gallant, A. and Tauchen, G. (1996). Which Moments to Match? *Econometric Theory*, 12(4):657-81.
- Lo, A. (1988). Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data. *Econometric Theory*, 4(2):231-247.
- Gourieroux, C. and Monfort, A. (1996). *Simulation-Based Econometric Methods*. Oxford University Press, USA.

References

- Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, 8:S85-S118.
- Phillips, P. and Yu, J. (2009). Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Handbook of Financial Time Series*.
- Smith Jr, A. (1993). Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions. *Journal of Applied Econometrics*, 8:S63-S84.