Estimation of moment-based models with latent variables

work in progress

Raffaella Giacomini and Giuseppe Ragusa

UCL/Cemmap and UCI/Luiss

UPenn, 5/4/2010

Dynamic latent variables in macroeconomic models

- E.g., time-varying parameters, structural shocks, stochastic volatility etc.
- Typical parametric setting: $X^T = (X_1, ..., X_T) = (Y^T, Z^T)$, Y^T observable, Z^T latent
- Joint density $p(X, \theta_0) = p(Y^T|Z^T, \theta_0)p(Z^T, \theta_0) \Longrightarrow$ estimation of θ_0 based on integrated likelihood

$$\widehat{\theta} = \arg\max_{\theta} \int p(Y^T|Z^T, \theta)p(Z^T, \theta)dZ^T$$

- Integrated likelihood computed by state-space methods

Existing state-space methods

- State equation $\longrightarrow p(Z^T, \theta)$
  - known in closed form
- Observation equation $\longrightarrow p(Y^T | Z^T, \theta)$ "filtering" density
  - known in closed form (e.g. Kalman filter) or easy to simulate
- Integral can be computed by MCMC methods

State-space methods for limited information models?

- We consider the following scenario:
- $p(Z^T, \theta)$ known $\rightarrow$ state equation same as before
- $p(Y^T|Z^T, \theta)$ unknown. Only information about θ is in the form of (non-linear) moment conditions

$$E_{t-1}\left[g(Y_t, Z_t, \theta)\right] = 0$$

  - $\rightarrow$ substitute observation equation with moment conditions

Applications. GMM with time-varying parameters

- Example #1. Time-varying "structural" parameters:

$$E\left[g(Y_t, \beta_t)\right] = 0$$

$$\beta_t = \Phi\beta_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iidN(0, \Sigma)$$

- $E\left[\cdot\right]$ defined with respect to joint distribution of Y_t and β_t
- Want to estimate $\theta = (\Phi, \Sigma)$ and sequence of "smoothed" β_t
- Application: Cogley and Sbordone's (2005) analysis of stability of structural parameters in a Calvo model of inflation

Applications. "Robust" stochastic volatility estimation

- Example #2.

$$Y_t = \sigma_t \varepsilon_t$$

$$\log\left(\sigma_t^2\right) = \alpha + \beta \log\left(\sigma_{t-1}^2\right) + v_t, \;\; v_t \sim iidN(0,1)$$

- Existing estimation methods require distributional assumption on ε_t (typically $N(0,1)$)
- Problem: does not capture "fat tails" of financial data $\Longrightarrow$ include jumps or use fat-tailed distribution for ε_t (not as straightforward as in GARCH case)
- Our method is robust to misspecification in distribution of ε_t

Applications. Nonlinear DSGE models

- Example #3. Prototypical DSGE model. Optimality conditions:

$$\begin{aligned} E_{t-1}\left[m(Y_t, S_t, Z_t, \beta)\right] &= 0 \\ S_t &= f(S_{t-1}, Y_t, Z_t, \beta) \\ Z_t &= \Phi Z_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iidN(0, \Sigma) \end{aligned}$$

- Want to estimate $\theta = (\beta, \Phi, \Sigma)$
- Y_t = observable variables
- S_t = endogenous latent variables
- Z_t = exogenous latent variables
- $m(\cdot)$ and $f(\cdot)$ known

An and Schorfheide (2007) DSGE model

- In AS model, the endogenous latent variable equation has a simple form:

$$S_t = f(Y_t, Z_t, \beta) \tag{1}$$

- Can substitute S_t and rewrite the equilibrium conditions as

$$E_{t-1}\left[g(Y_t, Z_t, \beta)\right] = 0$$

$$Z_t = \Phi Z_{t-1} + \varepsilon_t,\ \ \varepsilon_t \sim iid\ N(0, \Sigma)$$

- Warning: not all DSGEs fit this framework

Existing approaches to estimation of DSGE models

1. Theory does not provide likelihood $\longrightarrow$ must use approximation methods
2. Linearize around steady state (Smets and Wouters, 2003; Woodford, 2003)
   - *Solve the model* to find policy functions $Y_t = h(S_t, Z_t)$
   - Construct likelihood by Kalman filter
3. Nonlinear approximations (Fernandez-Villaverde and Rubio-Ramirez, 2005)
   - Solve the model (numerically or analytically in the case of second order approximations around steady state) to find policy functions
   - Construct likelihood by nonlinear state-space methods (e.g., particle filter)

Drawbacks of existing likelihood-based approaches

1. Linearization = possible loss of information (Fernandez-Villaverde and Rubio-Ramirez, 2005)
2. Must impose structure to solve the model
    1. Add "shocks"/measurement error to avoid stochastic singularity
    2. Restrict parameters to rule out indeterminacy (multiple rational expectations solutions)
3. Nonlinear state-space methods computationally intensive (must solve the model for each parameter draw) $\Longrightarrow$ so far mostly applied to simple models

Relationship with simulation-based method of moments

- GMM, SMM, EMM, Indirect inference (eg, Ruge-Murcia, 2010)
- Difference: requires knowledge of $p(Y^T|Z^T)$ or focuses on moments of the type

$$E_{\mathbf{Y}}\left[g\left(Y,\beta\right)\right]=0, \tag{2}$$

  where $g\left(Y,\beta\right)$ can be computed by simulation
- In our case, the model gives

$$E_{\mathbf{Y},\mathbf{Z}}\left[m\left(Y,Z,\beta\right)\right]=0$$

  $\Longrightarrow$ can be written as (2) only if $p(Z|Y)$ known
- Unlike these methods, we directly obtain estimates of the smoothed latent variables

The idea

- Propose methods for estimating non-linear moment-based models that "exploit" the information contained in the moment conditions
- Methods are:

1. Computationally convenient
2. Classical or Bayesian

Key elements of methodology

- Recall problem we want to solve (e.g., classical framework)

$$\max_{\theta} \int p(Y^T|Z^T,\theta) \;\cdot\; p(Z^T,\theta)dZ^T$$

$\uparrow$ unknown $\quad \uparrow$ known

- Two steps:
  1. Approximate the unknown likelihood $p(Y^T|Z^T,\theta)$
  2. Integrate out the latent variables using classical or Bayesian methods
  3. For DSGEs: from an exact likelihood of the approximate model.... to an approximate likelihood of the exact model

Approximate likelihoods

- We consider two different approximation strategies
- Both use projection theory (for no latent variables, Kim (2002), Chernozhukov and Hong (2003), Ragusa (2009)): out of all probability measures satisfying the moment conditions, choose the one that minimizes the Kullback-Leibler information distance
- Method 1 does not require solving the model (but not applicable to models with dynamic latent endogenous variables)
- Method 2 applicable to all models but requires solution of (approximate) model

Approximate likelihoods - Method 1

- Find density that satisfies moment conditions and minimizes distance from the true density: gives approximate likelihood

$$\widetilde{p}(Y^T|Z^T,\theta) \propto$$

$$\exp\left\{-\frac{1}{2}g_T'\left(Y^T,Z^T,\theta\right)V_T^{-1}\left(Y^T,Z^T,\theta\right)g_T\left(Y^T,Z^T,\theta\right)\right\}$$

$$\begin{aligned} g_T\left(Z^T,\theta\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(Y_t,Z_t,\theta) * w_{t-1} \\ V_T\left(Y^T,Z^T,\theta\right) &= Var(g_T\left(Y^T,Z^T,\theta\right)),\ w_{t-1} \text{ instruments} \end{aligned}$$

Approximate likelihoods - Method 1

$$\widetilde{p}(Y^T|Z^T,\theta) \propto$$
$$\exp\left\{-\frac{1}{2}g_T'\left(Y^T,Z^T,\theta\right)V_T^{-1}\left(Y^T,Z^T,\theta\right)g_T\left(Y^T,Z^T,\theta\right)\right\}$$

- $\widetilde{p}(Y^T|Z^T,\theta)$ is a simple transformation of the GMM objective function.
- Intuition:
  - When (Z^T,θ) is consistent with the model $g_T\left(Y^T,Z^T,\theta\right)\approx 0$ $\Longrightarrow \widetilde{p}(Y^T|Z^T,\theta)$ close to max value of 1.
  - When (Z^T,θ) is inconsistent with the moment conditions $\Longrightarrow$ large values of $g_T'\left(Y^T,Z^T,\theta\right)V_T^{-1}\left(Y^T,Z^T,\theta\right)g_T\left(Y^T,Z^T,\theta\right)\Longrightarrow$ $\widetilde{p}(Y^T|Z^T,\theta)\approx 0$.

Approximate likelihoods - Method 2

- Write $p(Y^T|Z^T) = \Pi_{t=1}^T p(Y_t|Z^t, Y^{t-1})$
- Choose approximate density $\widehat{p}(Y_t|Z^t, Y^{t-1}, \theta)$ (does not need to satisfy moment condition but easy to calculate) -
- For DSGEs, e.g., linearize model around steady state and apply Kalman filter $\Longrightarrow \widehat{p}(Y_t|Z^t, Y^{t-1}, \theta)$ are the filtered densities

Approximate likelihoods - Method 2

- "Tilt" $\widehat{p}(Y_t|Z^t, Y^{t-1}, \theta)$ towards moment condition $E_{t-1}\left[g(Y_t, Z_t, \theta)\right] = 0 \Leftrightarrow$ new density $\widetilde{p}(\cdot)$ satisfies moment condition and minimizes Kullback Leibler distance from $\widehat{p}(\cdot)$:
- Solve problem:

$$\min_{h \in \mathcal{H}} \int \int \log \left(\frac{h(Y_t|Z^t, Y^{t-1})}{\widehat{p}\left(Y_t|Z^t, Y^{t-1}, \theta\right)} \right) \widehat{p}\left(Y_t|Z^t, Y^{t-1}, \theta\right) dY_t dF\left(Z^t,\right.$$

$$s.t. \int \int g(Y_t, Z_t, \theta) h(Y_t|Z^t, Y^{t-1}) dY_t dF\left(Z_t\right) = 0$$

Approximate likelihoods - Method 2

- Under regularity conditions the solution is

$$\begin{aligned} & \widetilde{p}(Y_t|Z^t, Y^{t-1}, \theta) \\ = \; & \exp\left\{\eta_t + \lambda_t g\left(Y_t, Z_t, \theta\right)\right\} \widehat{p}(Y_t|Z^t, Y^{t-1}, \theta) \end{aligned}$$

- where

$$(\eta_t, \lambda_t) = \arg\min_{\eta,\lambda} \int \exp\left\{\eta + \lambda g\left(Y_t, Z_t, \theta\right)\right\} \widehat{p}(Y_t|Z^t, Y^{t-1}, \theta) dY$$

- $\lambda_t =$ "weights for each moment condition"; $\eta_t =$ integration constant
- (η_t, λ_t) are functions of Z^t, Y^{t-1}, θ

Approximate likelihoods - Method 2

- $$\widetilde{p}(Y_t|Z^t, Y^{t-1}, \theta) = \exp\{\eta_t + \lambda_t g(Y_t, Z_t, \theta)\}\,\widehat{p}(Y_t|Z^t, Y^{t-1},$$

- In practice, approximate integral and compute (η_t, λ_t) by simulating N times from $\widehat{p}(Y_t|Z^t, Y^{t-1}, \theta) \Longrightarrow$

$$(\eta_t, \lambda_t) = \arg\min_{\eta,\lambda} \frac{1}{N}\sum_{i=1}^{N} \exp\left\{\eta + \lambda g\left(Y_t^{(i)}, Z_t, \theta\right)\right\}$$

- Well-behaved objective function $\Longrightarrow$ for DSGEs, small additional computational cost relative to Kalman filter (cf. particle filter?)

The two methods in a simple case

- No latent variables, $Y^T = (Y_1, ..., Y_T)$ mean μ_0, variance σ_0^2
- Moment condition identifying parameters are

$$
\begin{aligned}
g_1(Y_t, \mu, \sigma^2) &= Y_t - \mu \\
g_2(Y_t, \mu, \sigma^2) &= Y_t^2 - \sigma^2
\end{aligned}
$$

- Method 1:

$$
\begin{aligned}
&\left(\widehat{\mu}, \widehat{\sigma}^2\right) \\
&= \arg \max_{\theta=(\mu,\sigma^2)} \exp\left\{-\frac{1}{2} g_T'\left(Y^T, \theta\right) V_T^{-1}\left(Y^T, \theta\right) g_T\left(Y^T, \theta\right)\right\}
\end{aligned}
$$

$\Longrightarrow$ our estimator is same as GMM (Chernozhukov and Hong (2003))

The two methods in a simple case

- Method 2: Start from pdf of $N(\overline{\mu}, \overline{\sigma}^2)$: $\widehat{p}\left(Y_t\right) = \frac{1}{\sqrt{2\pi\overline{\sigma}}} \exp\left\{-\frac{1}{2\overline{\sigma}}\left(Y_t - \overline{\mu}\right)^2\right\}$ and "tilt it" towards moment conditions

$$
\begin{aligned}
\widetilde{p}\left(Y_t\right) &= \exp\left\{\eta + \lambda_1\left(Y_t - \mu\right) + \lambda_2\left(Y_t^2 - \sigma^2\right)\right\} \frac{1}{\sqrt{2\pi\overline{\sigma}}} e^{-\frac{1}{2}(Y_t} \\
\lambda_1 &= \frac{\mu_0}{\sigma_0} - \frac{\overline{\mu}}{\overline{\overline{\sigma}}}; \\
\lambda_2 &= \frac{1}{2\overline{\sigma}} - \frac{1}{2\sigma_0}
\end{aligned}
$$

- No tilting if $\overline{\mu} = \mu_0, \overline{\sigma}^2 = \sigma_0^2$
- In this case $\widetilde{p}\left(Y_t\right) \sim N(\mu_0, \sigma_0^2) \Longrightarrow$ our estimator is the same as (Q)MLE
- Normality here is a special result - $\widetilde{p}\left(\cdot\right)$ no longer normal if e.g., $g\left(\cdot\right)$ non-linear

Step 2. Integrate out latent variables

- Classical estimation approach: solve

$$\widehat{\theta} = \max_{\theta} \int \widetilde{p}(Y^T | Z^T, \theta) p(Z^T, \theta) dZ^T$$

  - using Jacquier, Johannes and Polson (2007) to compute integral here works well in our limited experience

- Bayesian estimation approach: assume prior for θ (and Z_0), $\pi(\theta)$ and calculate the approximate posterior

$$\widetilde{p}(\theta, Z^T | Y^T) \propto \widetilde{p}(Y^T | Z^T, \theta) p(Z^T | \theta) \pi(\theta)$$

- Integration of latent variables step is the same as previous literature

Econometric properties

- For method 2 (tilted density), can show that MLE based on approximate integrated likelihood $\widetilde{p}(Y^T, \theta)$ is consistent for

$$\theta^* = \arg\min_{\theta} \int \log\left(\frac{\widetilde{p}(Y^T,\theta)}{p(Y^T)}\right) p(Y^T) dY^T$$

- $\theta^* =$ parameter that sets the approximate density that is consistent with the moment conditions as close as possible to true density
- In particular if moment condition uniquely identifies parameter θ_0, by construction $\theta^* = \theta_0$

Econometric properties

- Back to simple example: $Y_t \sim iid(\mu_0, \sigma_0^2)$, $g(Y_t, \theta) = (Y_t - \mu, Y_t^2 - \sigma^2)$, initial density $\widehat{p} \sim N(\overline{\mu}, \overline{\sigma}^2)$
- If tilt towards both moments, approximate density $\widetilde{p} \sim N(\mu_0, \sigma_0^2) \Longrightarrow$ our estimator (=QMLE) consistent for true parameters
- What if tilt towards only one moment condition?
    - E.g., only use $g_2(Y_t, \theta) = Y_t^2 - \sigma^2 \Longrightarrow \widetilde{p} \sim N(\frac{\overline{\mu}}{\overline{\sigma}}\sigma_0, \sigma_0^2)$
    - Variance estimated consistently; mean not estimated consistently
    - Suggests that not using moments can cause distortions $\Longrightarrow$ need to understand tradeoffs between too many/too few moments

Econometric properties

- Hypothesis testing, model selection relatively straightforward for method 2
- E.g., could test whether λ (or individual components) $= 0 \Leftrightarrow$ understand importance of non-linearities in DSGE models
- Open issue: identification (here assumed but challenging because of presence of latent variables + nonlinearity of moment conditions)

Method 1 in a simple example

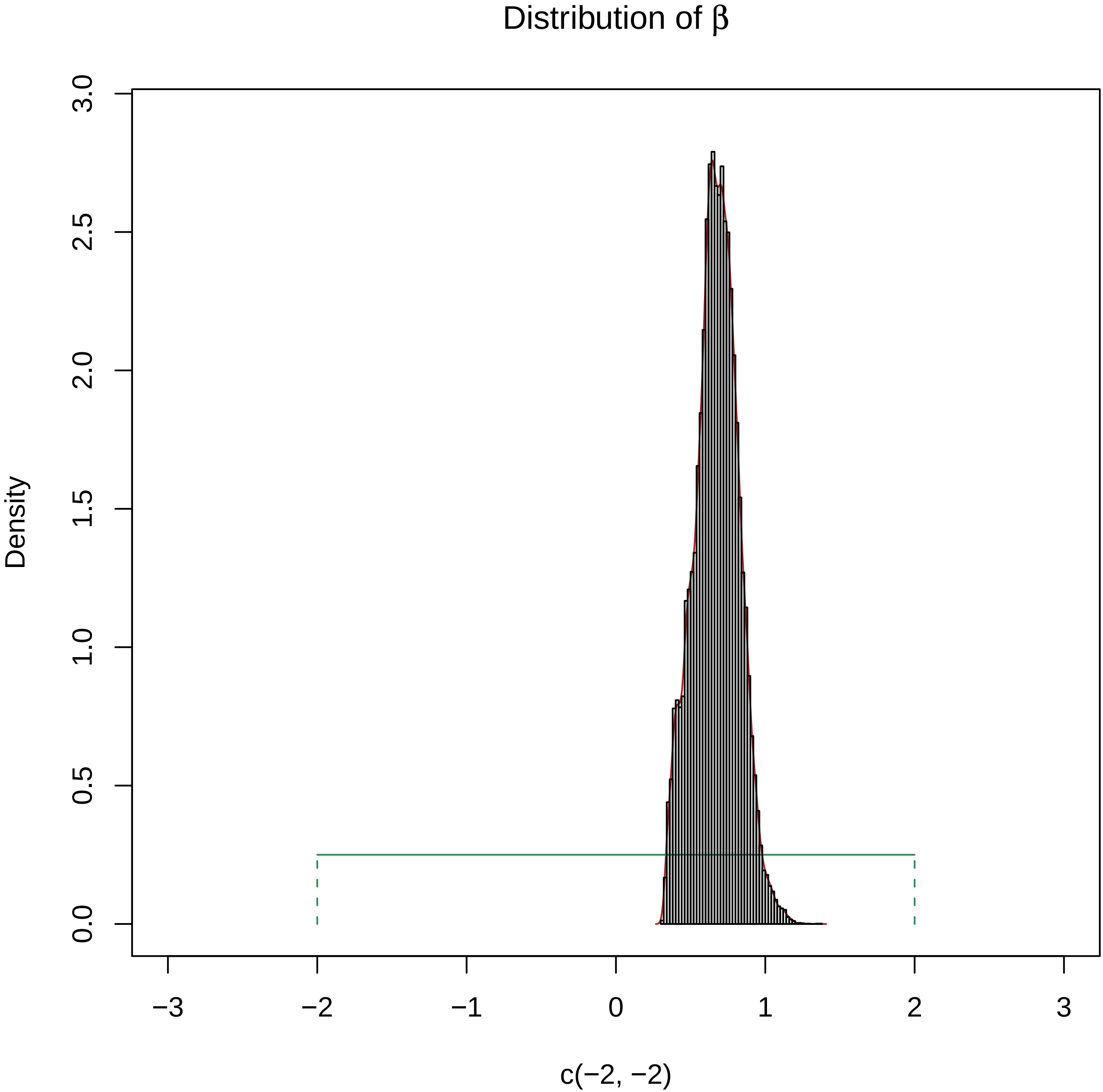
- Data-generating process

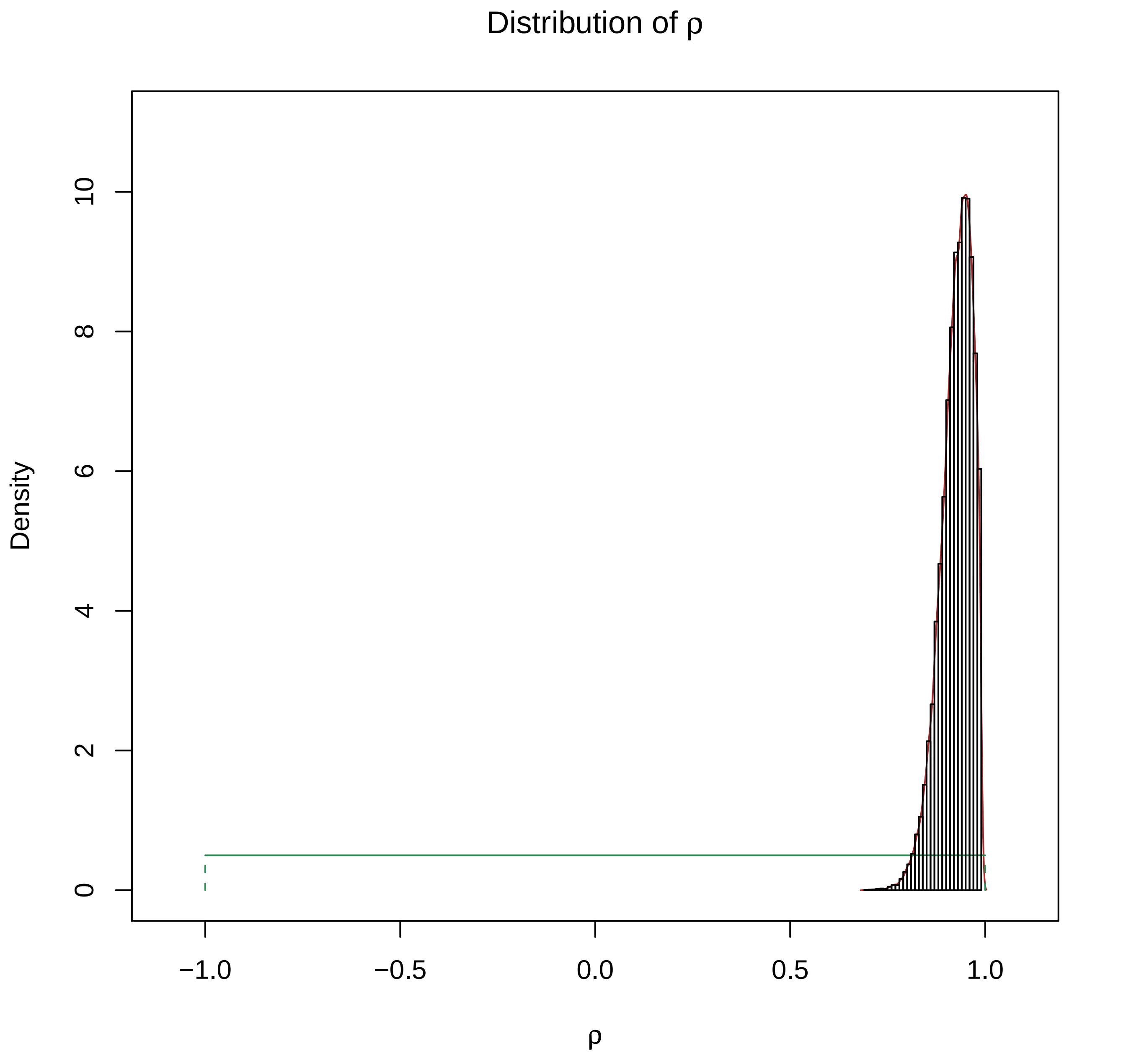
$$\begin{aligned} Y_t &= .9Z_t + v_t \sim iid\ N(0,1) \\ Z_t &= .9Z_{t-1} + \varepsilon_t \sim iid\ N(0,1) \end{aligned}$$

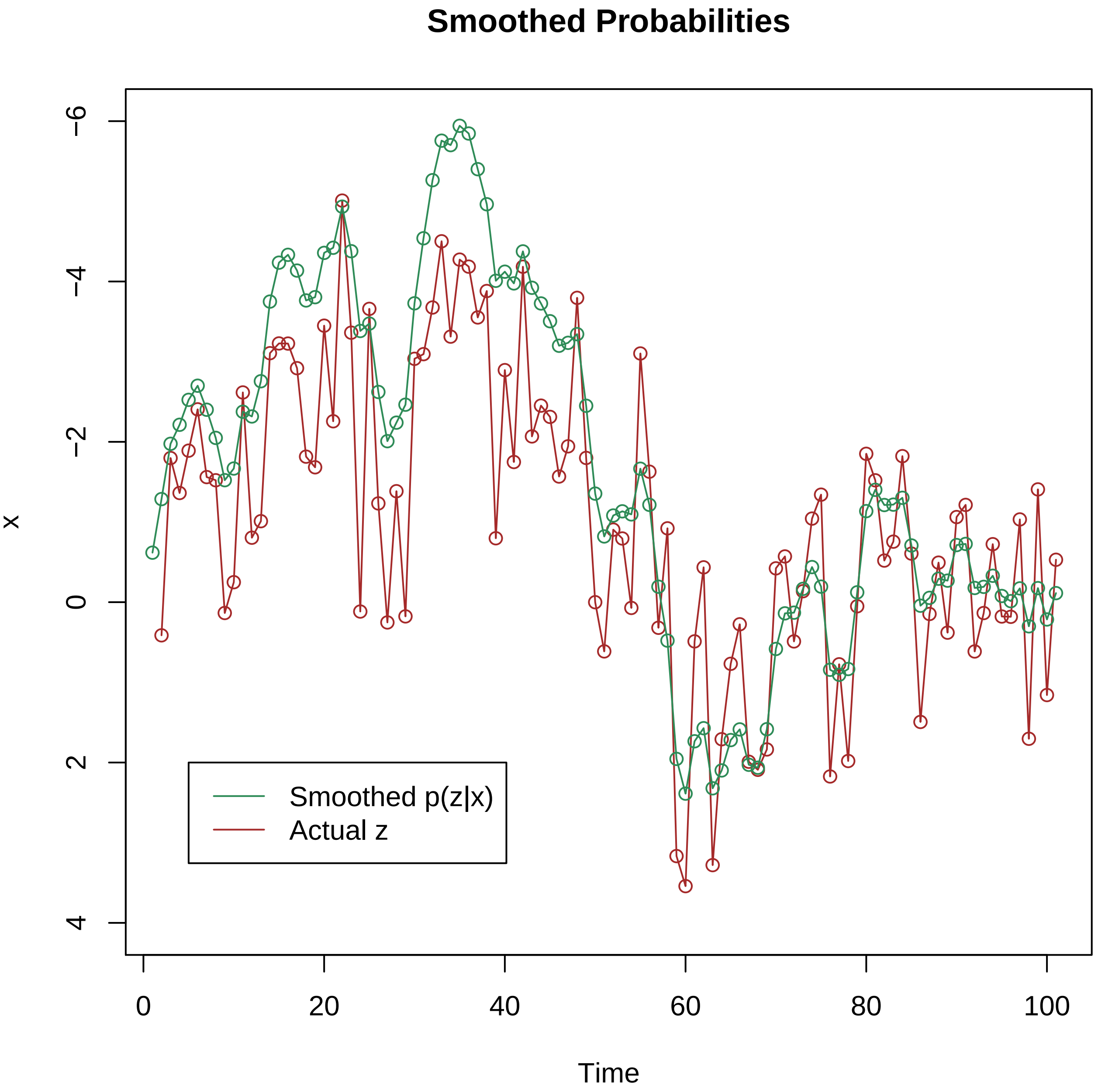
- Moment condition

$$\begin{aligned} E[Z_t(Y_t - \beta Z_t)] &= 0 \\ Z_t &= \rho Z_{t-1} + \varepsilon_t \sim iid\ N(0,1) \end{aligned}$$

- $g(Y_t, Z_t, \beta) = Z_t(Y_t - \beta Z_t)$
- Priors: $\beta \sim U(0,2)$, $\rho \sim U(0,1)$, $Z_0 \sim N(0, \frac{1}{1-\rho^2})$, $T = 100$
- Use Jacquier et al. (2007)


Distribution of β
Density
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−3
−2
−1
0
1
2
3
c(−2, −2)


Distribution of ρ
Density
ρ
0
2
4
6
8
10
−1.0
−0.5
0.0
0.5
1.0



Smoothed Probabilities
Smoothed p(z|x)
Actual z
x
Time
−6
−4
−2
0
2
4
0
20
40
60
80
100

Simulation: AS New Keynesian model

$$1 = \beta E_t \left[e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{R}_t - \hat{z}_{t+1} - \hat{\pi}_{t+1}} \right] \quad (3)$$

$$\begin{aligned} \frac{1-\nu}{\nu \phi \pi^2}(e^{\tau \hat{c}_t} - 1) &= (e^{\hat{\pi}_t} - 1)\left\{ \left[1 - \frac{1}{2\nu}\right] e^{\hat{\pi}_t} + \frac{1}{2\nu} \right\} \quad (4) \\ & \quad -\beta E \left[(e^{\hat{\pi}_{t+1}} - 1) e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{y}_{t+1} - \hat{y}_t + \hat{\pi}_{t+1}} \right] \\ e^{\hat{c}_t - \hat{y}_t} &= e^{-\hat{g}_t} - \frac{\phi \pi^2}{2}(e^{\hat{\pi}_t} - 1)^2 \quad (5) \end{aligned}$$

$$\begin{aligned} \hat{R}_t &= \rho_r \hat{R}_{t-1} + (1-\rho_r)\psi_1 \hat{\pi}_t + (1-\rho_r)\psi_2(\hat{y}_t - \hat{g}_t) + \sigma_R \varepsilon_{R,t} \quad (6) \\ & \hat{z}_t = \rho_z \hat{z}_{t-1} + \sigma_z \varepsilon_{z,t} \\ & \hat{g}_t = \rho_g \hat{g}_{t-1} + \sigma_g \varepsilon_{g,t} \end{aligned}$$

ε's independent $N(0,1)$

AS New Keynesian model

- Observable variables: $Y_t = (X_t, \pi_t, R_t)'$ (output, inflation and interest rate), where

$$\begin{aligned} X_t &= \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t) \\ \pi_t &= \pi^{(A)} + 400\hat{\pi}_t \\ R_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t. \end{aligned}$$

  $\hat{y}_t, \hat{R}_t, \hat{\pi}_t =$ deviation from steady state

- Endogenous latent variable: $S_t = \widehat{c}_t =$ deviation from steady state of consumption
- Exogenous latent variables: $Z_t = (\widehat{z}_t, \widehat{g}_t)' =$ technology and government spending

AS model in compact form

- (4) implies expression for S_t as a function of Y_t and $Z_t \Longrightarrow$ substitute into moment conditions
- Write policy rule as moment conditions
- Choose instruments to transform $E_t[\cdot]$ into $E[\cdot]$
- Write model as

$$E\left[g(Y_{t+1}, Y_t, Z_{t+1}, Z_t, \theta)\right] = 0$$

$$Z_t = \begin{pmatrix} \rho_z & 0 \\ 0 & \rho_g \end{pmatrix} Z_{t-1} + \varepsilon_t, \varepsilon_t \sim iidN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_g^2 \end{pmatrix}\right)$$

$g(\cdot)$ is 11×1, $\theta =$
$(\tau, \nu, \phi, 1/g, \psi_1, \psi_2, \rho_R, \sigma_R, \pi^{(A)}, \gamma^{(Q)}, r^{(A)}, \rho_z, \rho_g, \sigma_z, \sigma_g)$

AS model posterior

- Approximate posterior

$$\widetilde{p}(\theta, Z^T | Y^T) \propto \exp\left\{ -\frac{1}{2} g_T' \left(Y^T, Z^T, \theta \right) V_T^{-1} \left(Y^T, Z^T, \theta \right) g_T \left(Y \right. \right.$$
$$\times \prod_{t=1}^{T} p(z_t | z_{t-1}, \theta) \prod_{t=1}^{T} p(g_t | g_{t-1}, \gamma) p(z_0, g_0 | \gamma)$$

- z_0 and g_0 drawn from their stationary distributions

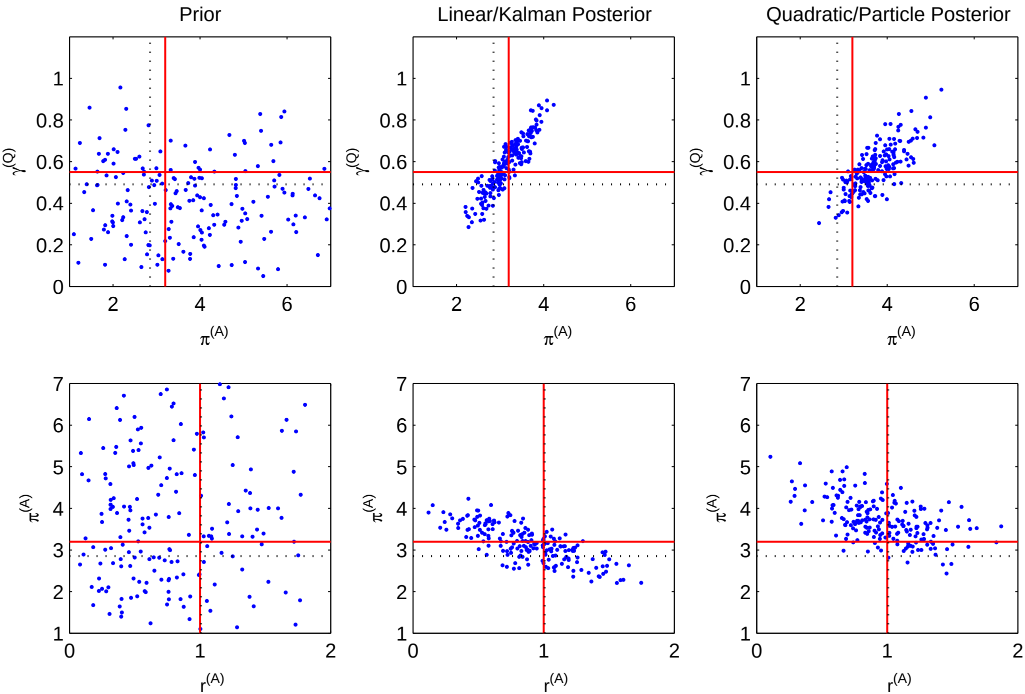
Simulation exercise

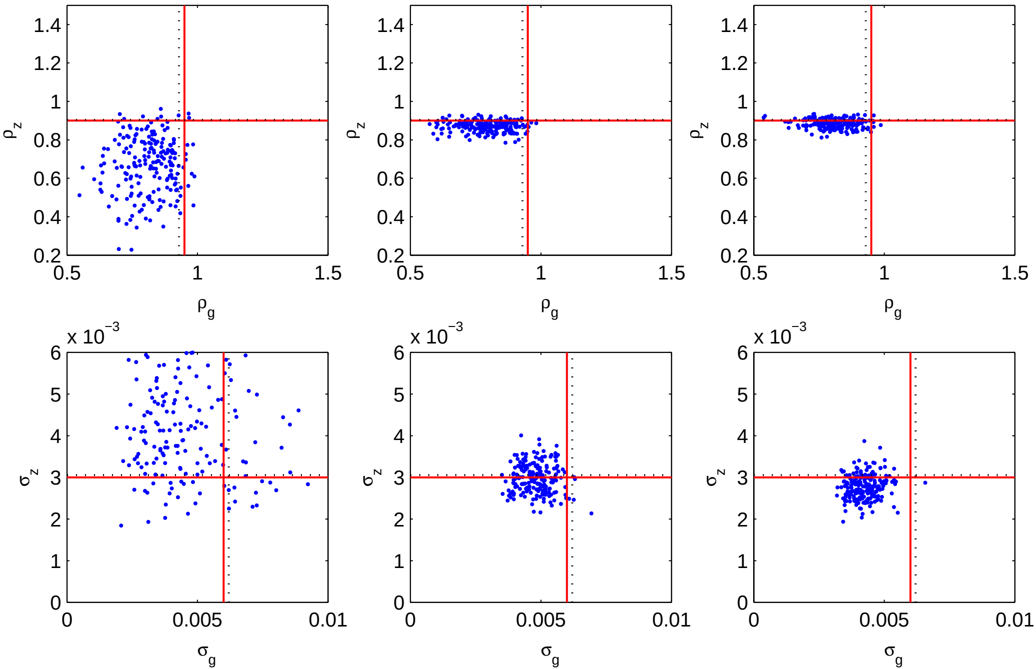
- Same DGP as AS:
  - Generate a time series ($T = 80$) from a second order approximation to the model
  - Parameters and priors as in AS
- Compare posteriors for θ obtained by our method to those in AS (both linear and nonlinear solution methods)

AS estimation results

- Draws from priors and posteriors for parameters $\pi^{(A)}, \gamma^{(Q)}, r^{(A)}, \rho_z, \rho_g, \sigma_z, \sigma_g$
- Red lines = true parameter values
- Estimation time: 100,000 MCMC draws $\approx$ 6 days

Figure 17: Posterior Draws: Linear versus Quadratic Approximation II

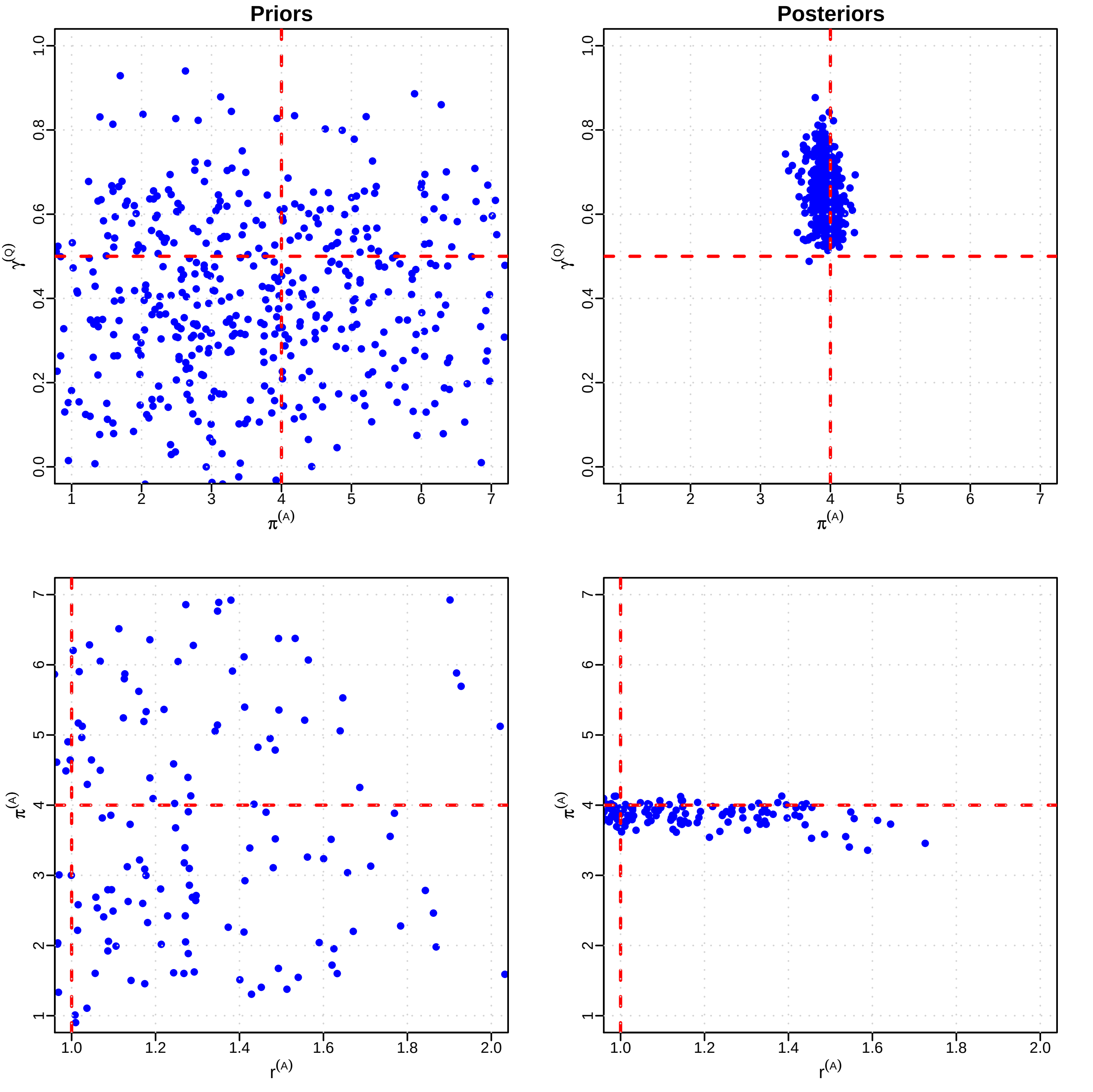


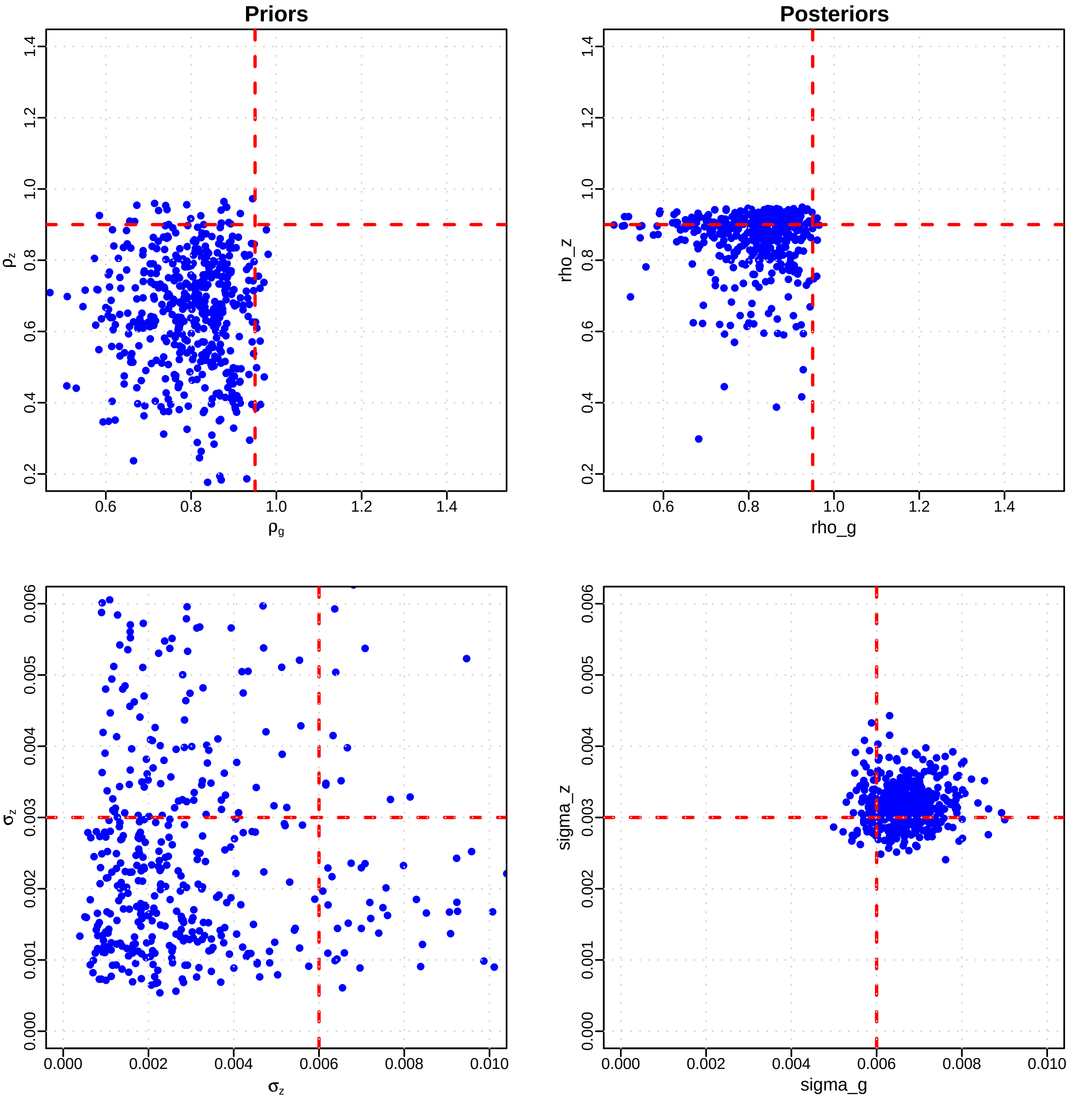


1.4
1.2
1
0.8
0.6
0.4
0.2
ρz
0.5
1
1.5
ρg
x 10−3
6
5
4
3
2
1
0
σz
0
0.005
0.01
σg

Our estimation results

- Draws from priors and posteriors for parameters $\pi^{(A)}, \gamma^{(Q)}, r^{(A)}, \rho_z, \rho_g, \sigma_z, \sigma_g$
- Red lines = true parameter values
- Estimation time: 2 million MCMC draws $\approx$ 4-5 hours


Priors
Posteriors
γ(Q)
π(A)
r(A)
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
1.2
1.4
1.6
1.8
2.0


Priors
Posteriors
ρ_z
ρ_g
rho_z
rho_g
σ_z
σ_z
sigma_z
sigma_g
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.008
0.010

Conclusion

- Two new methods for estimating structural parameters in moment-based models that depend on dynamic latent variables
- Projection-based approximate likelihoods that satisfy the moment conditions
- Marries the computational convenience of MCMC in high-dimensional problems with the ability of GMM to handle nonlinear moment conditions
- Directly delivers "smoothed" latent variables
- Potential for estimating realistic models and understanding importance of non-linearities