

NBER Summer Institute
What's New in Econometrics: Time Series
Lecture 8

July 15, 2008

Econometrics of DSGE Models

Outline

- 1) DSGE models and questions of interest
- 2) Model solution
- 3) Estimation: GMM
- 4) Estimation: Simulated GMM
- 5) Estimation: Maximum Likelihood
- 6) Estimation: Bayes
- 7) Inference, identification, and weak identification

1) DSGE Models and Questions of Interest

DSGEs can be used to address serious (real-world) empirical questions:

- Ask quantitative counterfactual policy questions
- Make conditional forecasts
- Examine effects of past policy changes (effects on means and variances, e.g. Great Moderation debate)
- Solve for optimal policies

A big breakthrough has been the development of numerical and conceptual methods that permit moving from calibration to estimation (Sargent (1989), Ireland (2000), Smets-Wouters (2003)).

This talk: the *econometrics* (estimation and inference) of DSGE modeling

- Briefly review specification and solution
- Focus will be on solved linearized models (models that have been written in state space form)

Overview of the standard DSGE methodology

1) Specify nonlinear optimization model

2) Obtain Euler equations

At this point, can estimate by GMM (single equation or system)

3) Log linearize

4) Solve the model (solve out expectations) – log linearization

5) Put into state space form

6) Estimation options:

a) Moment matching (match impulse response functions, possibly using simulation methods)

b) Maximum Likelihood

c) Bayes methods

7) Inference and evaluation

References to extensions will be given below

Overview of the standard DSGE methodology, ctd

There are some good recent references:

Textbooks/monographs:

Canova, F. (2007), *Methods for Applied Macroeconomic Research*,
Princeton: Princeton University Press.

Dejong, D.N. and C. Dave (2007), *Structural Macroeconomics*, Princeton:
Princeton University Press

Lecture Notes

Christiano L.J. (2007), “A Short Course on Estimation, Solution and
Policy Analysis using Equilibrium Monetary Models,” (extensive
slides are on Christiano’s Web site)

Outline

- 1) DSGE models and questions of interest
- 2) Model solution**
- 3) Estimation: GMM
- 4) Estimation: Simulated GMM
- 5) Estimation: Maximum Likelihood
- 6) Estimation: Bayes
- 7) Inference, identification, and weak identification

2) Model Solution

The steps:

Euler equations

Linearize

Example: the linearized Galí, López-Salido, Vallés (2003) model:

Calvo Pricing/NKPC:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t$$

intertemporal consumption:

$$x_t = -\sigma^{-1}(r_t - E_t \pi_{t+1} - rr_t^*) + E_t x_{t+1}$$

monetary policy:

$$r_t = (1-\alpha)\phi_\pi \pi_t + (1-\alpha)\phi_x x_t + \alpha r_{t-1} + u_t$$

natural interest rate:

$$rr_t^* = \rho \Delta a_t + (1+\phi)^{-1}(1-\lambda)\tau_t$$

processes for shocks:

$$\Delta a_t = \rho \Delta a_{t-1} + \eta_t^a$$

$$u_t = \delta u_{t-1} + \eta_t^u$$

$$\tau_t = \lambda \tau_{t-1} + \eta_t^\tau$$

Model parameters (12):

$$\theta = (\beta, \kappa, \sigma, \alpha, \phi_z, \phi_\pi, \rho, \delta, \lambda, \sigma_a^2, \sigma_u^2, \sigma_\tau^2)$$

Model solution, ctd.

Several solution methods are available for solving linearized models with rational expectations. For example, a model in the form,

$$\begin{pmatrix} x_{1t+1} \\ E_t x_{2t+1} \end{pmatrix} = A \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} + R x_{3t}$$

where x_{1t} are predetermined endogenous variables, x_{2t} are non-predetermined endogenous variables, and x_{3t} contains forcing variables (observed or unobserved) can be solved using Blanchard-Kahn (1980), see Dejong and Dave (2007, ch. 2) and Canova (2007, ch. 2).

Model solution, ctd.

The solved model will be linear and “one-sided” and can be written,

$$\textit{State equation:} \quad s_t = Fs_{t-1} + R\eta_t,$$

where: η_t are the system shocks, $E_{t-1}\eta_t = 0$, $E\eta_t\eta_t' = \Sigma_\eta$, s_t is the state vector and H , R , and Σ_ε depend on the original model parameters θ .

The system $s_t = Fs_{t-1} + R\eta_t$ is the state equation of the Kalman Filter, which is completed by adding the observer equation:

$$\textit{Observer equation:} \quad y_t = Hs_t + \varepsilon_t$$

where ε_t is i.i.d., $E\varepsilon_t = 0$, $E\varepsilon_t\varepsilon_t' = \Sigma_\varepsilon$, and $E\varepsilon_t\eta_t' = 0$. (ε_t is measurement error, and may or may not be present).

Model solution, ctd.

In general, the state space model,

$$\text{State equation:} \quad s_t = Fs_{t-1} + R\eta_t,$$

$$\text{Observer equation:} \quad y_t = Hs_t + \varepsilon_t$$

implies a VARMA representation for y . If there are more η 's and ε 's than observables, the VARMA errors will not in general be the state equation errors.

example (univariate permanent-transitory model):

$$s_t = s_{t-1} + \eta_t$$

$$y_t = \mu_t + \varepsilon_t$$

so $\Delta y_t = \eta_t + \Delta \varepsilon_t$. Thus $\text{var}(\Delta y_t) = \sigma_\eta^2 + 2\sigma_\varepsilon^2$, $\text{cov}(\Delta y_t, \Delta y_{t-1}) = -\sigma_\varepsilon^2$, and $\text{cov}(\Delta y_t, \Delta y_{t-h}) = 0$, $|h| > 1$, so y_t has the ARIMA(0,1,1) representation, $\Delta y_t = e_t + \theta e_{t-1}$

Model solution, ctd.

Using the KF (given the system matrices F , R , H , Σ_ε , Σ_η) and assuming normality, you can compute the entire joint distribution of the observables, which permits:

- computing conditional forecasts (conditional expectations)
- impulse responses (response of y_t to η_t)
- simulating data (draw η_t , ε_t and simulate)
- computing complicated moments (using simulated data)
- computing forecasts given data through date T , under changes in policy rules

Model solution, ctd.

For example, in the Gali, López-Salido, Vallés model above,

$$\text{State equation:} \quad s_t = Fs_{t-1} + R\eta_t,$$

$$\text{Observer equation:} \quad y_t = Hs_t + \varepsilon_t$$

If you observe r_t , π_t , and x_t , then $y_t = \begin{pmatrix} r_t \\ \pi_t \\ x_t \end{pmatrix}$, there is no ε_t , and H is a

matrix of zeros and ones selecting the appropriate elements of s_t :

$$s_t = \begin{pmatrix} \pi_t \\ x_t \\ r_t \\ rr_t \\ \Delta a_t \\ u_t \\ \tau_t \end{pmatrix} \text{ and } \eta_t = \begin{pmatrix} \eta_t^a \\ \eta_t^u \\ \eta_t^\tau \end{pmatrix},$$

Outline

- 1) DSGE models and questions of interest
- 2) Model solution
- 3) Estimation: GMM**
- 4) Estimation: Simulated GMM
- 5) Estimation: Maximum Likelihood
- 6) Estimation: Bayes
- 7) Inference, identification, and weak identification

3) Estimation: GMM

Note: henceforth we focus on models that are not stochastically singular (models with “at least as many shocks as observables” – linear models without singularity in spectral density). See e.g. Watson (1993) for one way to compare stochastically singular models to data.

GMM methods:

- (a) Euler equations, one equation at a time
- (b) Euler equations, as a system
- (c) GMM for matching other moments

(a) GMM, one equation at a time

Hybrid NKPC:

$$\pi_t = \lambda x_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \eta_t$$

Rational expectations:

$$E_{t-1}(\pi_t - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}) = 0$$

Instruments (need at least 2):

$$Z_t = \{ \pi_{t-1}, x_{t-1}, r_{t-1}, \dots \}$$

(a) GMM, one equation at a time, ctd.

We discussed GMM estimation in Lecture 4. For completeness, the GMM setup in the NKPC example is:

“errors”:

$$h(Y_t; \theta) = \pi_t - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}$$

errors \times instruments:

$$\phi_t(\theta) = h(Y_t; \theta) Z_t$$

GMM objective function:

$$S_T(\theta) = \left[T^{-1/2} \sum_{t=1}^T \phi_t(\theta) \right]' W_T \left[T^{-1/2} \sum_{t=1}^T \phi_t(\theta) \right]$$

GMM estimator:

$$\hat{\theta} \text{ minimizes } S_T(\theta)$$

Centered sample moments:

$$\Psi_T(\theta) = T^{-1/2} \sum_{t=1}^T (\phi_t(\theta) - E\phi_t(\theta))$$

Efficient GMM:

$$W_T = \hat{\Omega}^{-1}, \Omega = E[\Psi_T(\theta) \Psi_T(\theta)']$$

(a) GMM, one equation at a time, ctd.

Under conventional (strong identification) asymptotics, the feasible efficient GMM estimator (two-step, iterated, CUE):

$$\sqrt{T} (\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (D\Omega^{-1}D')^{-1})$$

where $D = E \left. \frac{\partial \phi_t(\theta)}{\partial \theta} \right|_{\theta_0}$.

Assessment tools:

- Test of overidentifying restrictions (*J*-test)
- Tests for stability (split-sample tests with a known break date, or GMM break tests with an unknown break date as in Andrews (1993) and Andrews-Ploberger (1994))
- forecast assessment (Lecture 10)

(b) System GMM estimation

In general system estimation improves efficiency over single-equation estimation (cross-equation restrictions, SUR reasons).

For example, in the GL-SV model, if the shocks are i.i.d.,

errors:
$$h(Y_t; \theta) = \begin{pmatrix} \pi_t - \beta\pi_{t+1} - \kappa x_t \\ x_t + \sigma^{-1}(r_t - E_t\pi_{t+1}) - x_{t+1} \\ r_t - (1 - \alpha)\phi_\pi\pi_t - (1 - \alpha)\pi_x x_t - \alpha r_{t-1} \end{pmatrix}$$

errors \times instruments:
$$\phi_t(\theta) = h(Y_t; \theta) \otimes Z_t$$

Or, different instruments could be used for different equations.

(b) System GMM estimation

Advantages of GMM estimation of Euler equations:

- model doesn't need to be solved (for estimation; however the model needs to be solved for applications)
- don't need to assume distribution for Euler equation errors – just martingale difference sequence with moments
- Estimation can proceed using nonlinear Euler equations (so nonlinear solution method can be applied only once, to the estimated parameters)
- Standard (strong identification) estimation and asymptotics proceeds as in the single-equation case; standard tools for assessing fit.
- tools are available for weak identification inference

Disadvantages:

- Efficiency loss relative to ML
- requires modification for unobserved serially correlated shocks (in which case expectational conditions won't hold using observables)

(c) GMM for matching other moments

- Most commonly used to match model impulse responses to estimated impulse responses (the moments are the empirical impulse responses).
- Examples (matching IRFs from structural VARs): Christiano, Eichenbaum, Evans (2005), Boivin and Giannoni (2006a) (recall that linearized SS form implies VARMA which is restricted VAR(∞))
- Motivation: limited information approach (see CEE (1995))
- Specifics: Compute sample IRFs from SVAR – call these $\mu_T(y)$
Compute implied IRFs from DSGE – call these $\mu(\theta)$
Choose θ to minimize the distance,
$$S_T(\theta) = [\mu(\theta) - \mu_T(y)]' W_T [\mu(\theta) - \mu_T(y)]$$
Usual GMM asymptotics apply
- This produces lots moments – which can result in bias & poor performance of sampling distribution. See Hall, Inoue, Nason, and Rossi (2007) for an IRF selection information criterion.

Outline

- 1) DSGE models and questions of interest
- 2) Model solution (references only)
- 3) Estimation: GMM
- 4) Estimation: Simulated GMM**
- 5) Estimation: Maximum Likelihood
- 6) Estimation: Bayes
- 7) Inference, identification, and weak identification

4) Estimation: Simulated GMM

Simulated GMM

McFadden (1989), Pakes and Pollard (1989)

- GMM matches moments implied by the model (for which we have explicit expressions) to sample counterparts based on the data, assuming that, in population, these match at a (unique) true value θ_0 .
- Simulated GMM addresses the case that the theoretical distribution (moments) implied by the model is difficult to derive analytically – so it is computed numerically instead, by simulation.
- Why DSGEs might not yield analytical GMM moment restrictions:
 - unobservables in Euler equations (e.g. missing variables, which can have dynamics under the model – so the expectational orthogonality condition can't be implemented)
 - nonlinear function of SS parameters not readily calculated

Simulated GMM, ctd.

A (very) simple SMM example:

model: y_t i.i.d. $N(\theta, \sigma_y^2)$, σ_y^2 known; want to estimate θ

data: (y_1, \dots, y_T)

GMM: $\min_{\theta} (\bar{y} - \theta)^2$, so $\hat{\theta} = \bar{y}$ and $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma_y^2)$

SMM:

(i) choose some θ and simulate x_1, \dots, x_m from the model $N(\theta, \sigma_y^2)$

(ii) compute the simulated theoretical moment, $\bar{x}(\theta)$

(iii) solve for the SMM estimator:

$$\hat{\theta}^{SMM} : \min_{\theta} (\bar{y} - \bar{x}(\theta))^2$$

you can do minimize this by grid search (or random search/simulated annealing or simplex) - no derivatives needed

Simple SMM example, ctd.

What is the distribution of the SMM estimator?

x_i are drawn from $N(\theta, \sigma_y^2)$ as $x_i = \theta + u_i$, where u_i is i.i.d $N(0, \sigma_y^2)$.

so, $\bar{x} = \theta + \bar{u}$, so the objective function can be rewritten,

$$\begin{aligned} S_T^{SMM}(\theta) &= (\bar{y} - \bar{x}(\theta))^2 = (\bar{y} - (\theta + \bar{u}))^2 \\ &= (\bar{y} - \bar{u}) - \theta)^2 \end{aligned}$$

so the SMM estimator is: $\hat{\theta}^{SMM} = \bar{y} - \bar{u}$

so
$$\begin{aligned} \sqrt{T}(\hat{\theta}^{SMM} - \theta) &= \sqrt{T}(\bar{y} - \theta) - \sqrt{T}\bar{u} = \sqrt{T}(\bar{y} - \theta) - \sqrt{\frac{T}{m}}\sqrt{m}\bar{u} \\ &\sim N(0, \sigma_y^2) + \sqrt{\frac{T}{m}}N(0, \sigma_y^2) = N\left(0, \left(1 + \frac{1}{\kappa}\right)\sigma_y^2\right), \text{ where } \frac{m}{T} \rightarrow \kappa \end{aligned}$$

Comments:

- The two normals are independent because \bar{y} and \bar{u} are independent
- You need to reuse the same u 's (same seed) for each trial θ (why?)
- variance is the same as efficient GMM – up to scale factor $1+1/\kappa$

SMM algorithm

(i) generate $m = \kappa T$ observations $x_1(\theta), \dots, x_m(\theta)$ from your model, using the same seed for different trial values of θ for smoothness of $S_T^{SMM}(\theta)$

(ii) compute simulated moments $\mu_m(x(\theta)) = \frac{1}{m} \sum_{t=1}^m h(x_t(\theta))$

in the simple example, $\mu_m(x(\theta)) = \bar{x}(\theta)$

(iii) compute sample moments $\mu_T(y) = \frac{1}{T} \sum_{t=1}^T h(y_t)$ (example: $\mu_T(y) = \bar{y}$)

(iv) Compute the SMM objective function,

$$S_T^{SMM}(\theta) = [\mu_m(x(\theta)) - \mu_T(y)]' W_T [\mu_m(x(\theta)) - \mu_T(y)]$$

(v) go to (i) and repeat until $S_T^{SMM}(\theta)$ is minimized

When $W_T = \Omega^{-1}$ (efficient GMM weighting matrix),

$$\sqrt{T} (\hat{\theta}^{SMM} - \theta) \xrightarrow{d} N \left(0, \left(1 + \frac{1}{\kappa} \right) (D' \Omega^{-1} D)^{-1} \right)$$

SMM, ctd.

$$\sqrt{T} (\hat{\theta}^{SMM} - \theta) \xrightarrow{d} N \left(0, \left(1 + \frac{1}{\kappa} \right) (D' \Omega^{-1} D)^{-1} \right)$$

Comments:

1. Same distribution as efficient GMM up to the scale factor $1+1/\kappa$
2. Subtleties arise when variables are discrete, see Gouriéroux and Monfort (1997) (continuity of the objective function; solutions can change the asymptotics)
3. Instead of drawing a single long path, $x_1(\theta), \dots, x_m(\theta)$, you might do better to draw κ paths of length T – then the simulated moments will better approximate any finite-sample bias inherent in using data of length T
4. For more on SMM see: Gouriéroux and Monfort (1997), Gallant and Tauchen (2001)

Outline

- 1) DSGE models and questions of interest
- 2) Model solution
- 3) Estimation: GMM
- 4) Estimation: Simulated GMM
- 5) Estimation: Maximum Likelihood**
- 6) Estimation: Bayes
- 7) Inference, identification, and weak identification

5) Estimation: Maximum Likelihood

With the assumption of normal errors, the Kalman Filter can be applied to the (linearized solved) state space model to yield the likelihood:

$$\text{State equation:} \quad s_t = Fs_{t-1} + R\eta_t, \quad \eta_t \text{ i.i.d. } N(0, \Sigma_\eta)$$

$$\text{Observer equation:} \quad y_t = Hs_t + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } N(0, \Sigma_\varepsilon); \quad \varepsilon_t, \eta_t \text{ independent}$$

The Kalman filter:

$$\text{State prediction:} \quad s_{t/t-1} = Fs_{t-1/t-1}, \quad P_{t/t-1} = FP_{t-1/t-1}F' + \Sigma_\eta,$$

$$y \text{ prediction:} \quad \mu_{t/t-1} = Hs_{t/t-1}, \quad \Sigma_{t/t-1} = HP_{t/t-1}H' + \Sigma_\varepsilon$$

$$\text{Updating:} \quad K_t = P_{t/t-1}H'\Sigma_{t/t-1}^{-1}, \quad s_{t/t} = s_{t/t-1} + K_t(y_t - \mu_{t/t-1}), \quad P_{t/t} = (I - K_t)P_{t/t-1}.$$

log-likelihood:

$$L(\theta, Y_T) = c + -0.5 \sum_{t=1}^T \left\{ \ln(\det(\Sigma_{t/t-1})) + (y_t - \mu_{t/t-1})' \Sigma_{t/t-1}^{-1} (y_t - \mu_{t/t-1}) \right\}$$

MLE, ctd.

$$\hat{\theta}^{MLE} : \max_{\theta} L(\theta; Y_T)$$

Comments

- Asymptotics of MLE (QMLE variance expression):

$$\sqrt{T} (\hat{\theta}^{MLE} - \theta_0) \xrightarrow{d} N(0, \mathbf{I}(\theta_0)^{-1} \mathbf{J}(\theta_0) \mathbf{I}(\theta_0)^{-1}),$$

$$\text{where } \mathbf{I}(\theta_0) = -E \frac{1}{T} \frac{\partial^2 L(\theta; Y_T)}{\partial \theta \partial \theta'} \Big|_{\theta_0} \text{ and } \mathbf{J}(\theta_0) = E \frac{1}{T} \frac{\partial L(\theta; Y_T)}{\partial \theta} \Big|_{\theta_0} \frac{\partial L(\theta; Y_T)}{\partial \theta} \Big|_{\theta_0}'$$

which holds “under suitable regularity conditions.” Essentially these regularity conditions are that the MLE is locally well-approximated by a quadratic with curvature that is nonrandom.

- Likelihoods can have multiple peaks (use multiple starting values; use random search algorithms) and cliffs (penalized likelihood if you stray outside the determinacy region (Dynare); try non-derivative based methods)

Outline

- 1) DSGE models and questions of interest
- 2) Model solution (references only)
- 3) Estimation: GMM
- 4) Estimation: Simulated GMM
- 5) Estimation: Maximum Likelihood
- 6) Estimation: Bayes**
- 7) Inference, identification, and weak identification

6) Estimation: Bayes

Bayes basics

Bayesian inference treats the parameters as random & conditions on the data.

Bayes law:
$$P(\theta|Y_T) = \frac{L(Y_T | \theta)\pi(\theta)}{f_Y(Y_T)}$$

where: L is the pdf of $Y_T|\theta$ (the likelihood)

$\pi(\theta)$ is the prior

f_Y is the marginal distribution of Y_T : $f_Y(Y_T) = \int L(Y_T | \theta)\pi(\theta)d\theta$

$P(\theta|Y_T)$ is the posterior distribution of θ given the data

Posterior mean:
$$E_{\theta|Y}g(\theta) = \int g(\theta)P(\theta|Y_T)d\theta = \frac{\int g(\theta)L(Y_T | \theta)\pi(\theta)d\theta}{\int L(Y_T | \theta)\pi(\theta)d\theta}$$

Bayes – implementation

$$E_{\theta|Y}g(\theta) = \int g(\theta)P(\theta | Y_T)d\theta = \frac{\int g(\theta)L(Y_T | \theta)\pi(\theta)d\theta}{\int L(Y_T | \theta)\pi(\theta)d\theta}$$

Comments

- Analytic integration isn't feasible/possible outside textbook examples
- Breakthrough in Bayesian statistics has been simulation-based methods (MCMC) for numerical integration, see Geweke (2005) for econometrics treatment (many other books on numerical methods in general statistics, statistical genetics, etc). MCMC methods were discussed in an earlier lecture and are implemented in Dynare.
- Inference tools: credible sets (Bayesian “confidence intervals”, posterior means; more later on model evaluation)
- Implementation in modern software is relatively painless so the relevant questions involve interpretation, not mechanics.
- An and Schorfheide (2007) provide a survey and primer

Why use Bayes methods? Some issues, old and new

1. First principles: The likelihood principle

(see Berger and Wolpert (1988)) The likelihood principle states that all the relevant information in the data is embodied in the likelihood function; two different mechanisms that yield the same likelihood function should yield identical inferences.

Classic example: Stopping rule for Bernoulli trials.

(a) Experiment #1: Give 5 patients an experimental drug, $Y_i = 1$ if they die from side effects. Observe 0,1,0,0,1

(b) Experiment #2: Give patients the experimental drug until 2 die. Observe 0,1,0,0,1

Bayes old and new, ctd

The data sets are the same. The frequentist holds that S/n is an unbiased estimator of p in (a), and N/s is an unbiased estimator of $1/p$ in (b). The likelihood principle says that inference about p should be the same. Said differently, the likelihood principle holds that experimental design doesn't matter if it yields the same likelihood

2. First principles: Decision theory (subjectivist Bayes)

Loss function: $L(\theta, \theta_0) = (\theta - \theta_0)^2$ (for example)

(Frequentist) risk function: $R(\theta, \hat{\theta}) = EL(\theta, \hat{\theta})$ (expectation over $f_{Y|\theta}$)

Bayes (integrated) risk: $r_{\pi}(\hat{\theta}) = \int R(\theta, \hat{\theta})\pi(\theta)d\theta$

Bayes decision rule: choose $\hat{\theta}$ to minimize $r_{\pi}(\hat{\theta})$

- If your (subjective) prior is $\pi(\theta)$, then use it for *your* decision
- Subject to technical conditions, the complete class theorem says that all admissible decision rules are Bayes or limiting Bayes

Bayes, old and new, ctd

3. Stein (1955)/James-Stein (1960) considerations

Suppose loss is

$$L(\theta, \theta_0) = (\theta - \theta_0)' (\theta - \theta_0)$$

- OLS is inadmissible if $\dim(\theta) \geq 3$ (extension of Stein (1955))
- Frequentist risk of the form can be reduced by using a Bayes estimator.
- This is (very) important for forecasting with many predictors – when bias (possibly large) in individual coefficients is irrelevant, and all that matters is forecast performance.
- This is not the right loss function if one is interested in inference about deep parameters

Bayes, old and new, ctd

4. Priors formalize the use of prior knowledge

- A reasonable point in principle. In experimental sciences this makes sense.
- In DSGE applications, one should distinguish between information gleaned from micro studies (e.g. micro studies of individual attitudes towards risk) and data based on prior empirical analysis of aggregate time series data. Running regressions on U.S. macro data then using the results as priors for analyzing the same macro data in DSGE doesn't really count as prior knowledge.

Bayes, old and new, ctd

5. It doesn't matter anyway – you get the same answers using Bayes and ML. This relies on the *Bernstein-von Mises theorem*.
- Loosely speaking, if θ is finite-dimensional, if the prior doesn't rule out any sets of θ , if the sample size is large, and θ is identified, then $\hat{\theta}^{MLE}$ and $\hat{\theta}^{Bayes}$ (the posterior mean) will be close. Moreover, the posterior distribution of θ around $\hat{\theta}^{Bayes}$ will be close to the distribution of $\hat{\theta}^{MLE}$ around θ_0 . Thus Bayes inference and MLE will give similar inferences (same point estimates, 95% credible sets and 95% confidence intervals coincide)
 - That is, enough data will overwhelm any nondogmatic prior. See Lehmann and Casella (1998, Chapter 6.8)

Bayes, old and new, ctd.

- Frequentist Wald tests and Bayes posterior odds tests also coincide (Andrews (1994))
- B-vM basically says the ML/Bayes choice is one of convenience.
- However, this result is rather delicate and doesn't hold in high dimensions (Freedman (1999)) and requires strong identification
- Applicability of this argument to DSGEs?

6. Priors solve the identification problem

- If your model is (nearly) unidentified then you can “solve” the identification problem by imposing a prior.
- This is calibration.

Bayes, old and new, ctd.

Futher comments:

1. ***Communication***, frequentist and Bayes.

2. ***Why this debate, while old, still matters.*** In many fields, statisticians have moved beyond these philosophical debates, but one reasons for doing so is that their data sets have evolved (expanded) while ours have not, see Efron (2005). We return to this in Lecture 11.

3. ***Posterior mode.*** MCMC (MH) might not visit full support, there might be cliffs, etc; if the posterior is not well estimated over the full support then the posterior mode can replace the posterior mean as the estimator. Sometimes this is called Bayes maximum likelihood.

Further comments, ctd.

4. *Robust Bayes.* Bayesian statisticians have developed a large set of tools for investigating and reporting sensitivity of Bayesian results to priors or parts of priors, including parametric methods with contaminated priors, or priors within density bounds, and mixtures (these might be the most easily implemented in the DSGE MCMC setup) and nonparametric methods. These and other methods are reviewed in Berger (1994), also see Geweke (2005, ch. 3.3).

5. *Empirical Bayes.* (Robbins (1955, 1964); see Maritz and Lwin (1989)). Empirical Bayes isn't really Bayes, it treats the parameters of the prior as the parameters to be estimated. This is most useful when there are hierarchical priors so that the prior distribution is tightly parameterized. We return to empirical Bayes in Lecture 11.

Outline

- 1) DSGE models and questions of interest
- 2) Model solution
- 3) Estimation: GMM
- 4) Estimation: Simulated GMM
- 5) Estimation: Maximum Likelihood
- 6) Estimation: Bayes
- 7) Inference, identification, and weak identification**

7) Inference, Identification, and Weak Identification

Threats to validity of inferences

1. Model misspecification

2. Structural breaks

3. Persistent regressors.

Not a new problem – but it is relevant here too (Li (2006))

4. Weak identification.

Weak Identification in DSGEs

- The identification conditions for asymptotic normality of likelihood inference are conceptual related to those for IV (but not the same in the details) – convergence of the objective function to a (local) quadratic with nonrandom curvature matrix.
- Part of the challenge is figuring out what parts of the model are strongly identified and what are not, then communicating that information
- There is no consensus about how to handle weak identification issues in DSGEs
- Tools for handling weak identification (from a frequentist perspective) are most fully developed to date for GMM, for which there are (limited) tests for weak identification and robust inference methods
- A promising approach is to get more information! (Boivin-Giannoni (2006b))

Weak identification, ctd.

Three examples:

- 1.Canova and Sala (2006)
- 2.Smets and Wouters (2003)
- 3.MC study in Ruge-Murcia (2007)
- 4.MC study of the Gali, López-Salido, Vallés (2003) model

Example #1: Canova and Sala (2006 version)

MM-IRF objective function contours of a small DSGE from Canova and Sala (2006)

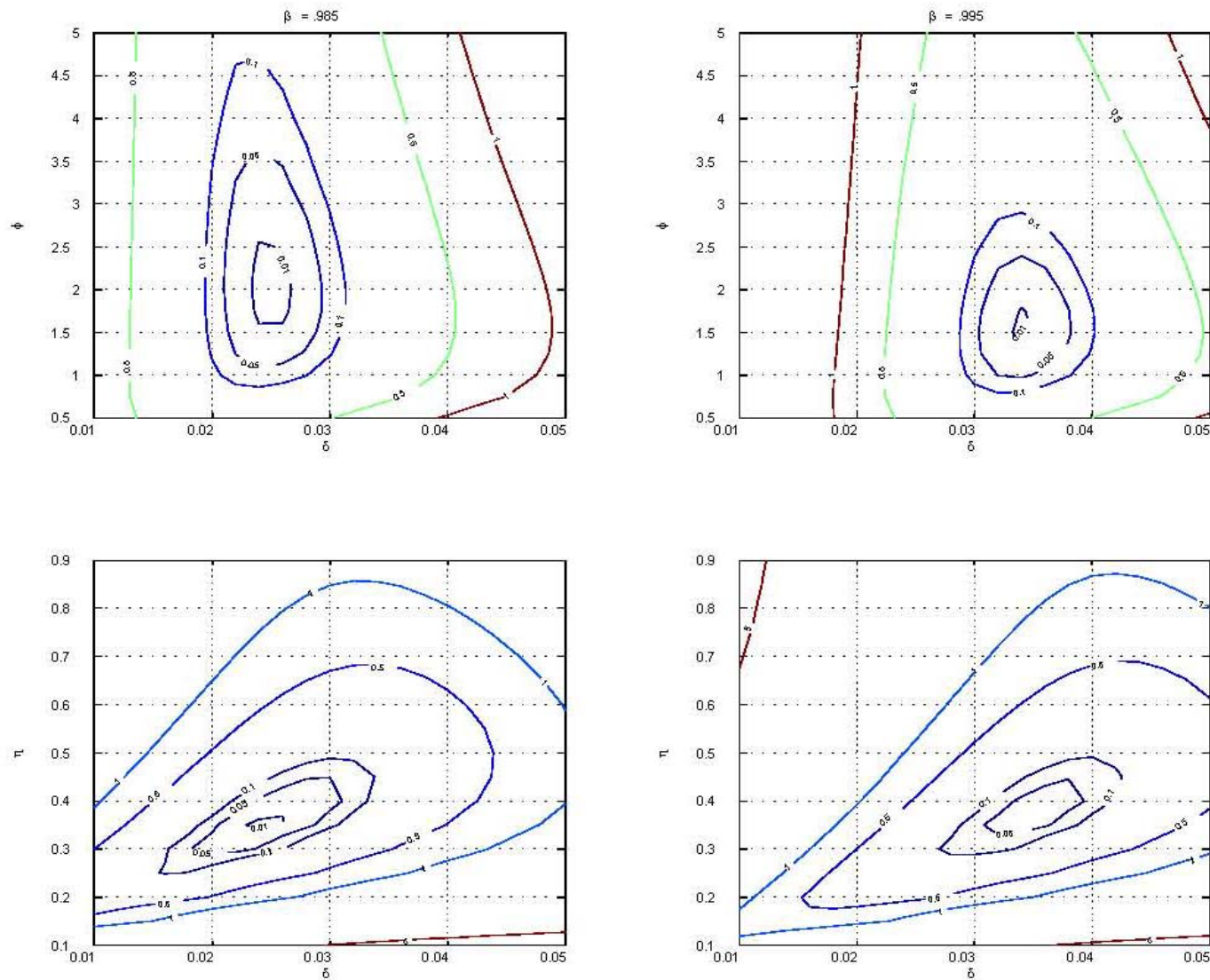
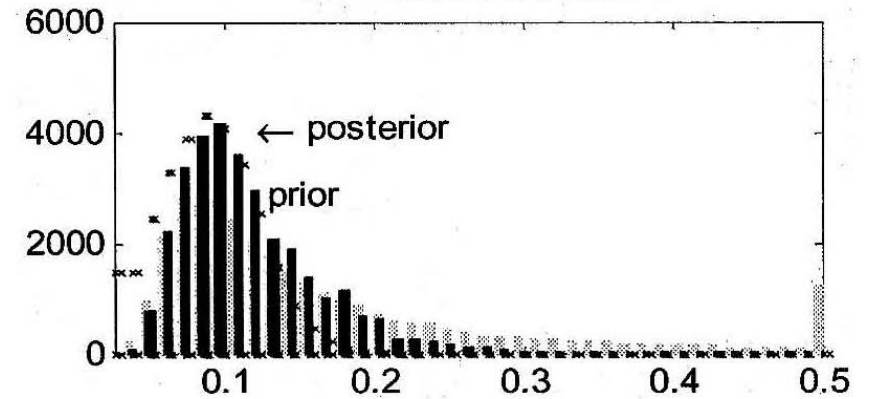
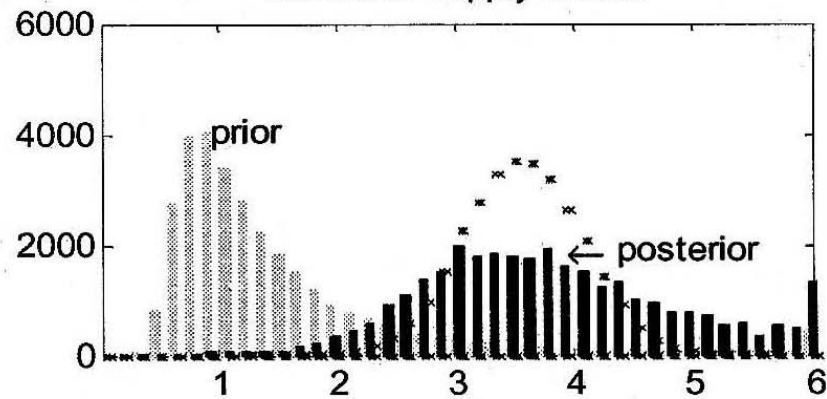
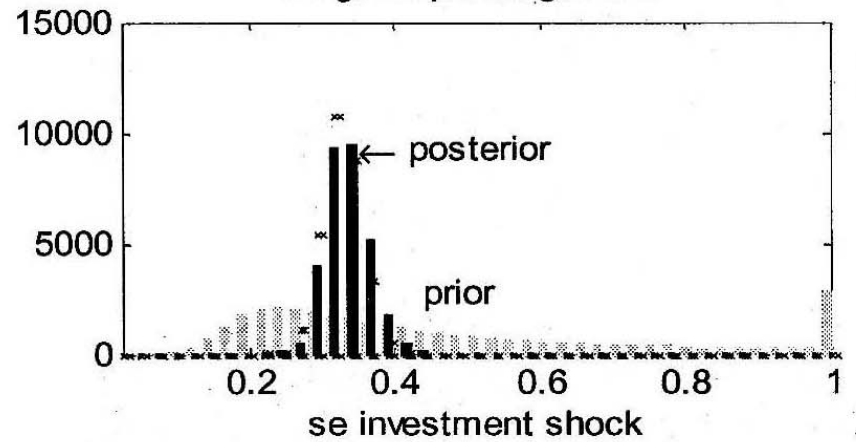
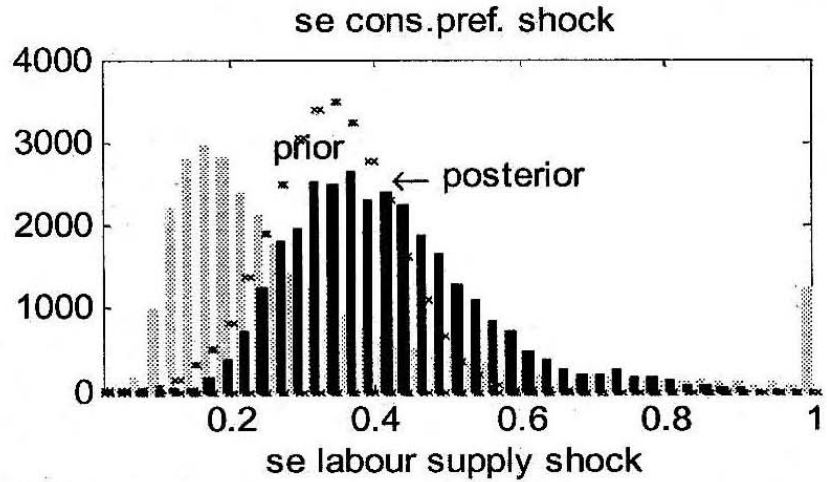
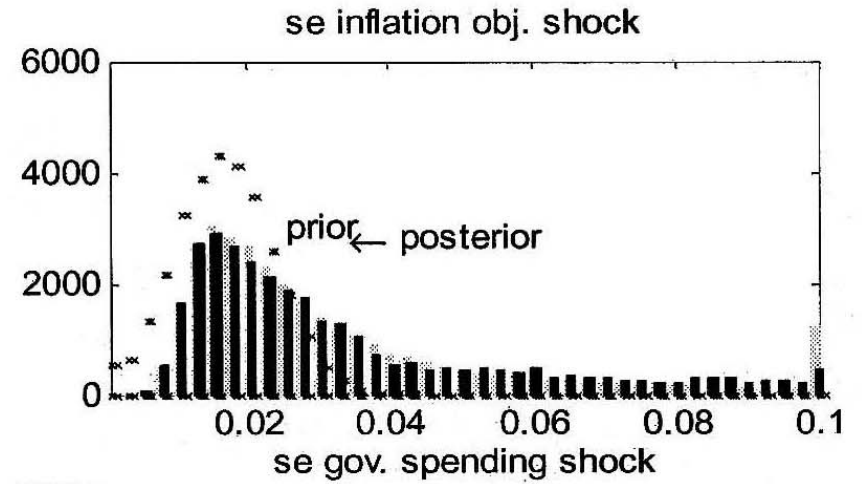
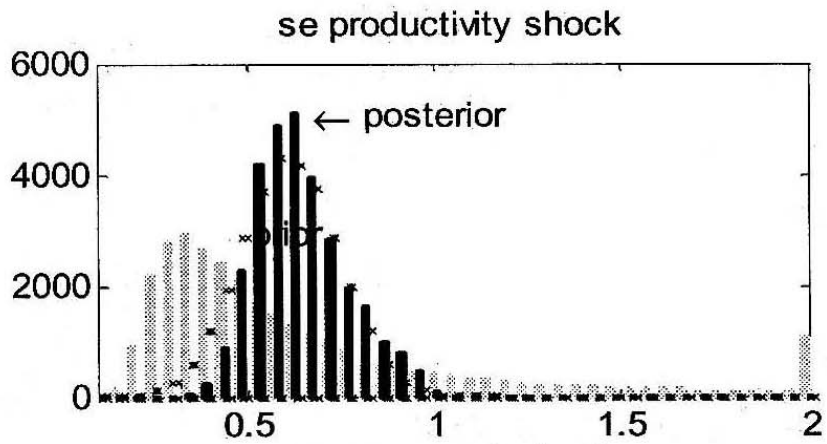


Figure 3: Contour Plots

Example #2: Smets and Wouters (2003)

Smets-Wouters (2003) Fig 1 plots priors and posteriors for their model, computed by MCMC



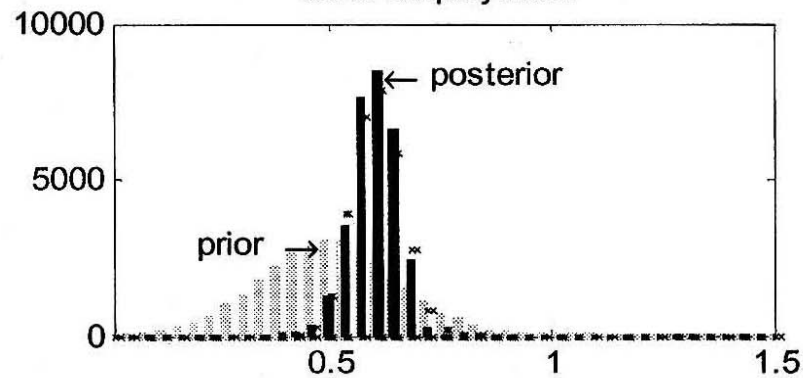
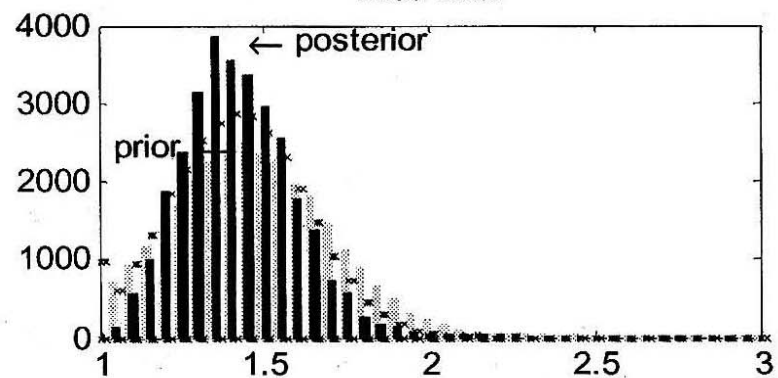
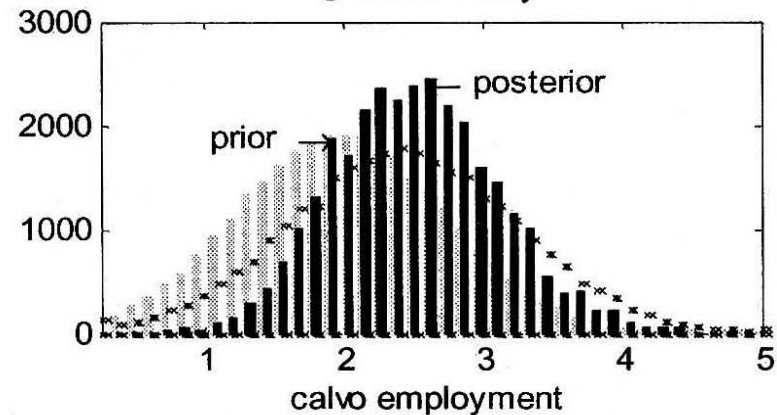
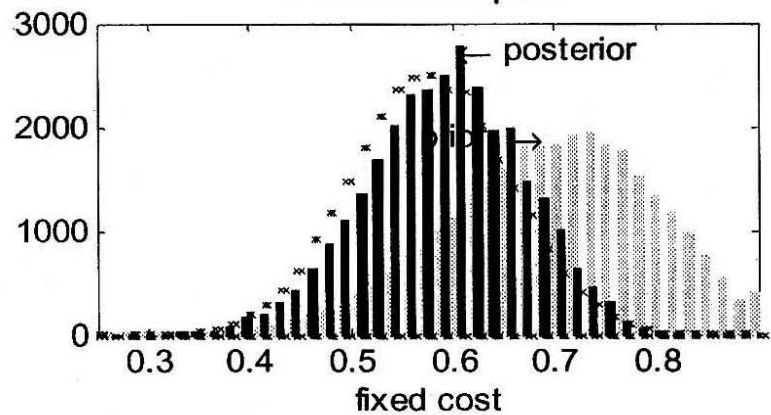
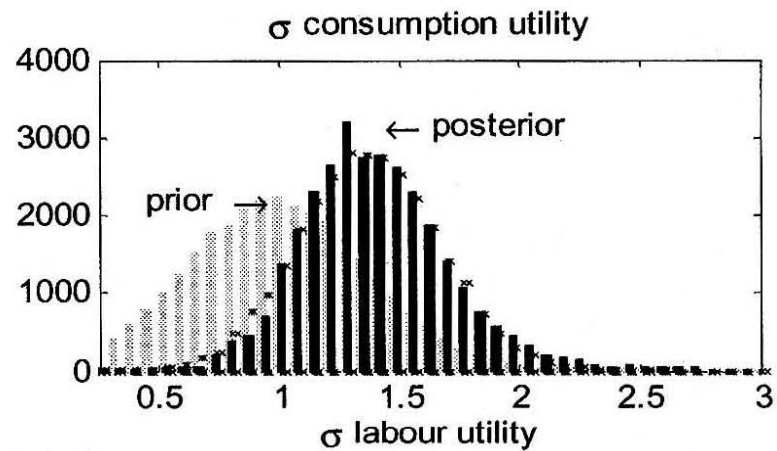
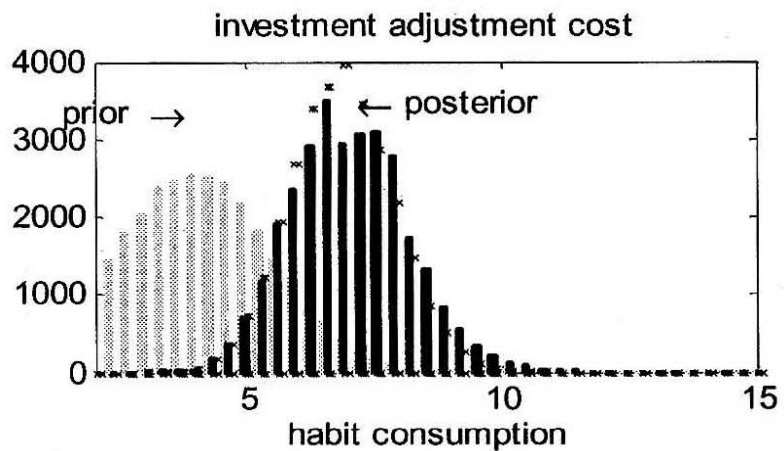
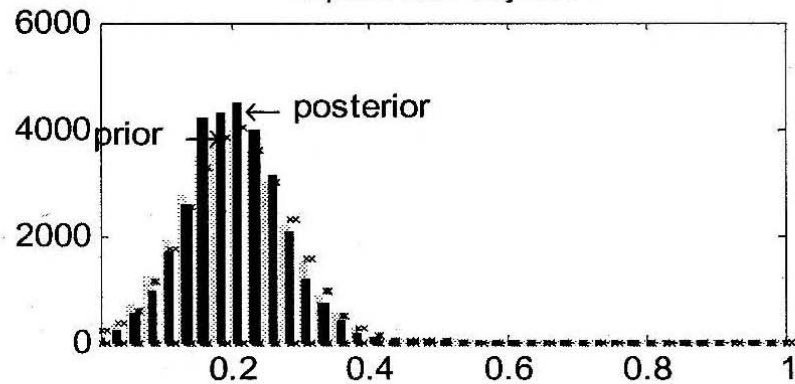
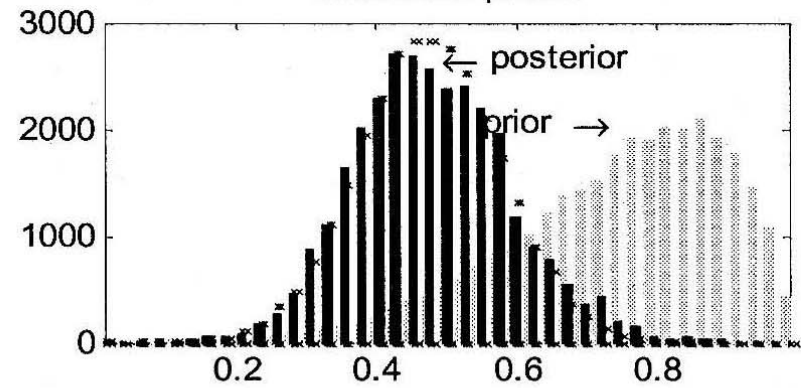
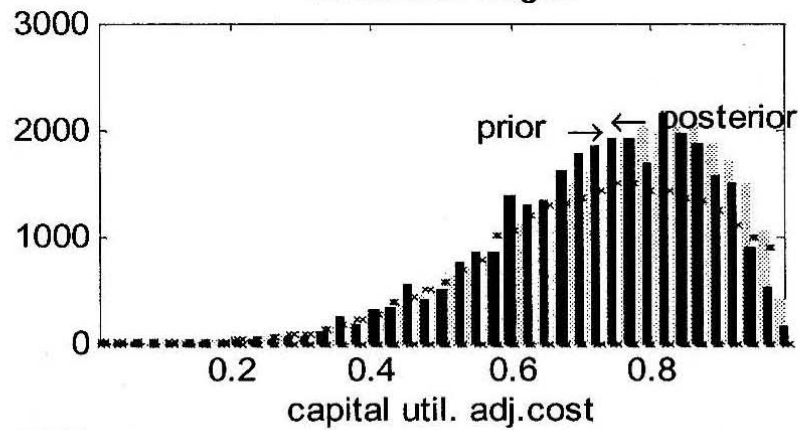
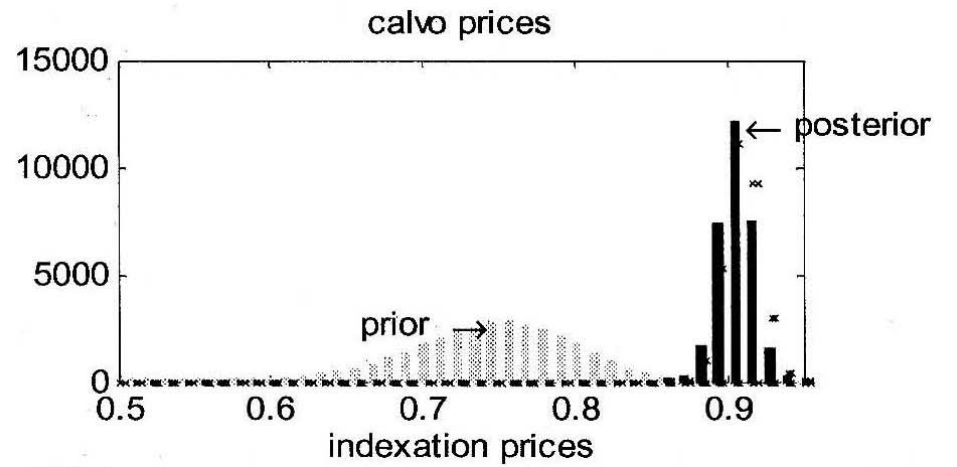
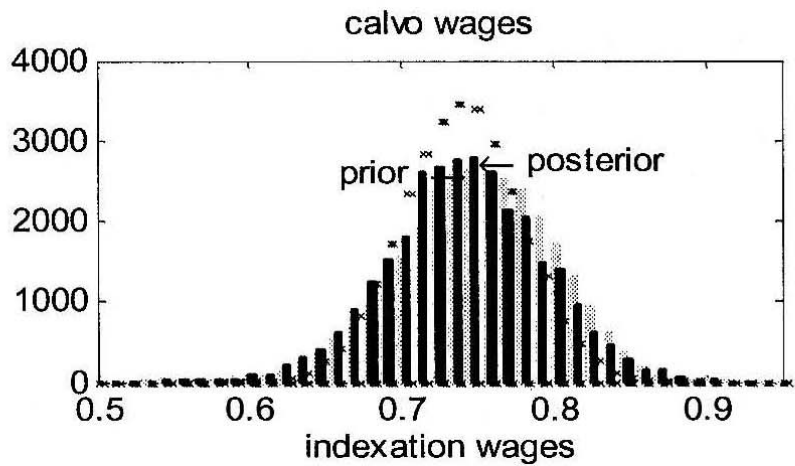


FIGURE 1B. Estimated Parameter Distribution



Example #3: MC study in Ruge-Murcia (2007)

- Investigates small model with stochastic singularity (two unobserved state variables, capital growth and a disturbance), and three observed variables, c_t , y_t , n_t .
- Finds reasonably good performance of GMM and moments matching (“good” here means that the normal approximation performs reasonably well).
- Difficult to generalize because of the stochastic singularity and the small size of the model.

Example #4: MC study of the Gali, López-Salido, Vallés (2003) model
(thanks to Anna Mikusheva)

GL-SV variant: add markup shock, eliminate taste shock

Calvo Pricing/NKPC: $\pi_t = \beta E_t \pi_{t+1} + \kappa x_t + \eta_t^\pi$

intertemporal consumption: $x_t = -\sigma^{-1}(r_t - E_t \pi_{t+1} - rr_t^*) + E_t x_{t+1}$

monetary policy: $r_t = (1-\alpha)\phi_\pi \pi_t + (1-\alpha)\phi_x x_t + \alpha r_{t-1} + u_t$

natural interest rate: $rr_t^* = \rho \Delta a_t$

processes for shocks: $\Delta a_t = \rho \Delta a_{t-1} + \eta_t^a$

$$u_t = \delta u_{t-1} + \eta_t^u$$

Model parameters (11): $\theta = (\beta, \kappa, \sigma, \alpha, \phi_z, \phi_\pi, \rho, \delta, \sigma_a^2, \sigma_u^2, \sigma_\pi^2)$

Observe: π_t, r_t, x_t

GL-SV MC study, ctd

Design details:

- Parameter values: $\beta = 0.99$, $\phi_x = 0.15$, $\varphi_\pi = 1.5$, $\alpha = 0.8$, $\rho = 0.9$, $\lambda = 0.5$, $\delta = 0.2$, $\phi = 1$, $\theta = 0.75$, $\sigma_a = \sigma_\tau = \sigma_u = \sigma_\pi = 1$.
- Two parameters fixed at true value (not estimated): $\beta = 0.99$ and $\sigma = 1$
 - so there are 9 estimated parameters
- $T = 160$
- 50 MC repetitions (need more)
- Statistics computed:
 - MLE of each element of θ , fixing others at true value (1-dimensional)
 - Unrestricted MLEs, t -statistics testing θ_0 , standard errors
 - LR statistic = $2 \times (\text{unrestricted L} - \text{restricted L})$, one variable at a time, all others estimated

Table A

One coefficient at a time – all others fixed at θ_0 (restricted ML)

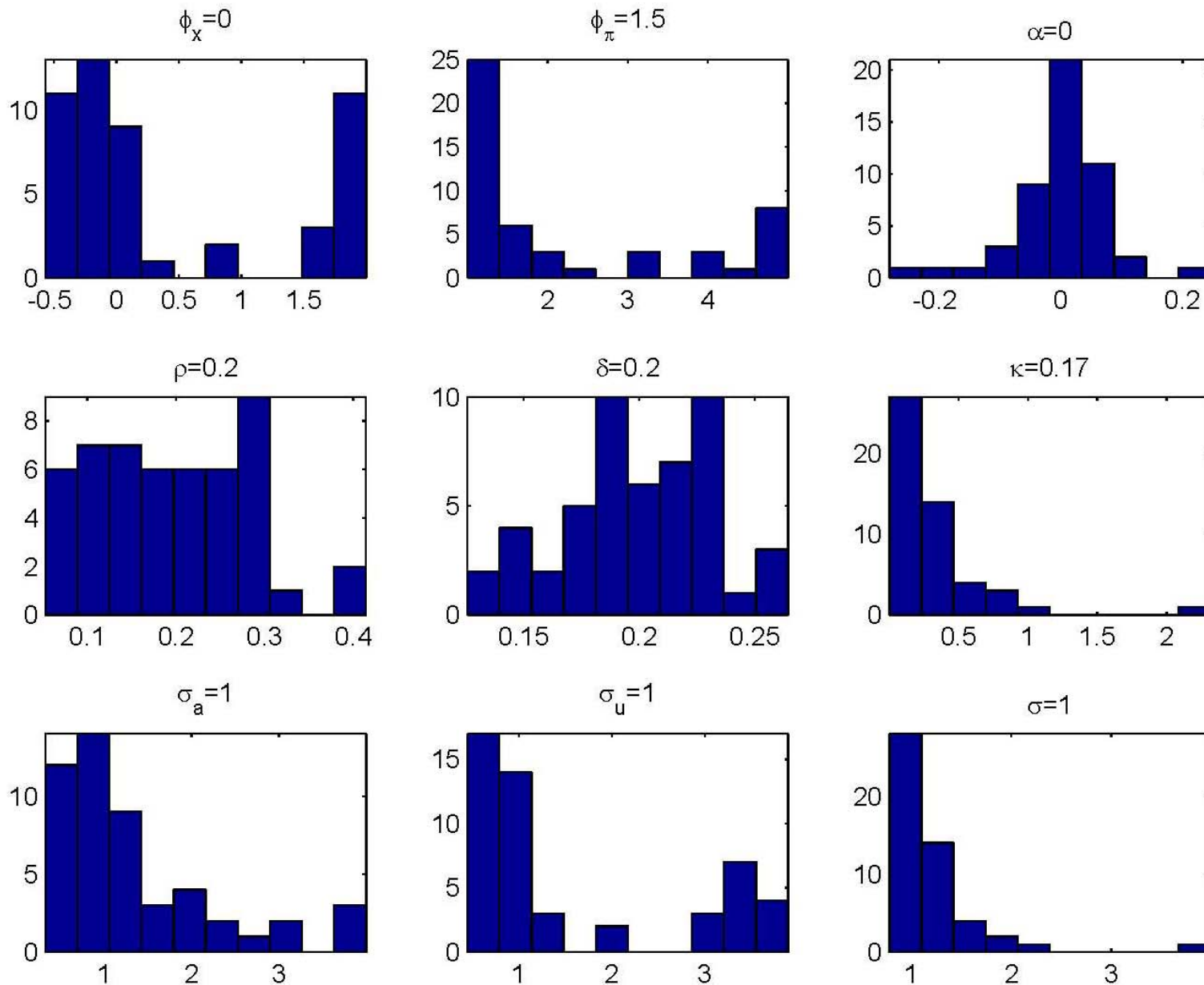
	ϕ_x	ϕ_π	α	ρ	δ	κ	σ_a	σ_u	σ_π
True values:	0.15	1.5	0.8	0.9	0.2	0.172	1	1	1
(a) LR test statistic									
LR test: Size	0.02	0.04	0.06	0.04	0.06	0.06	0.04	0.04	0.06
KS chi-squared? (p -value)	0.545	0.809	0.395	0.709	0.877	0.215	0.840	0.922	0.192
KS statistic	0.11	0.088	0.125	0.098	0.082	0.146	0.086	0.077	0.150
(b) Wald t statistic:									
Wald t size	0.02	0.04	0.06	0.04	0.06	0.06	0.04	0.06	0.06
KS normal? (p -value)	0.390	0.873	0.645	0.929	0.067	0.659	0.965	0.415	0.230
KS statistic	0.125	0.082	0.103	0.075	0.181	0.102	0.069	0.123	0.144
(c) Point estimator									
Bias	0.003	0.009	-0.001	0.000	-0.003	0.003	0.002	-0.009	-0.005
RMSE	0.0389	0.0651	0.0118	0.0087	0.0153	0.0371	0.0537	0.0533	0.0595

Table B

All coefficients estimated (full ML)

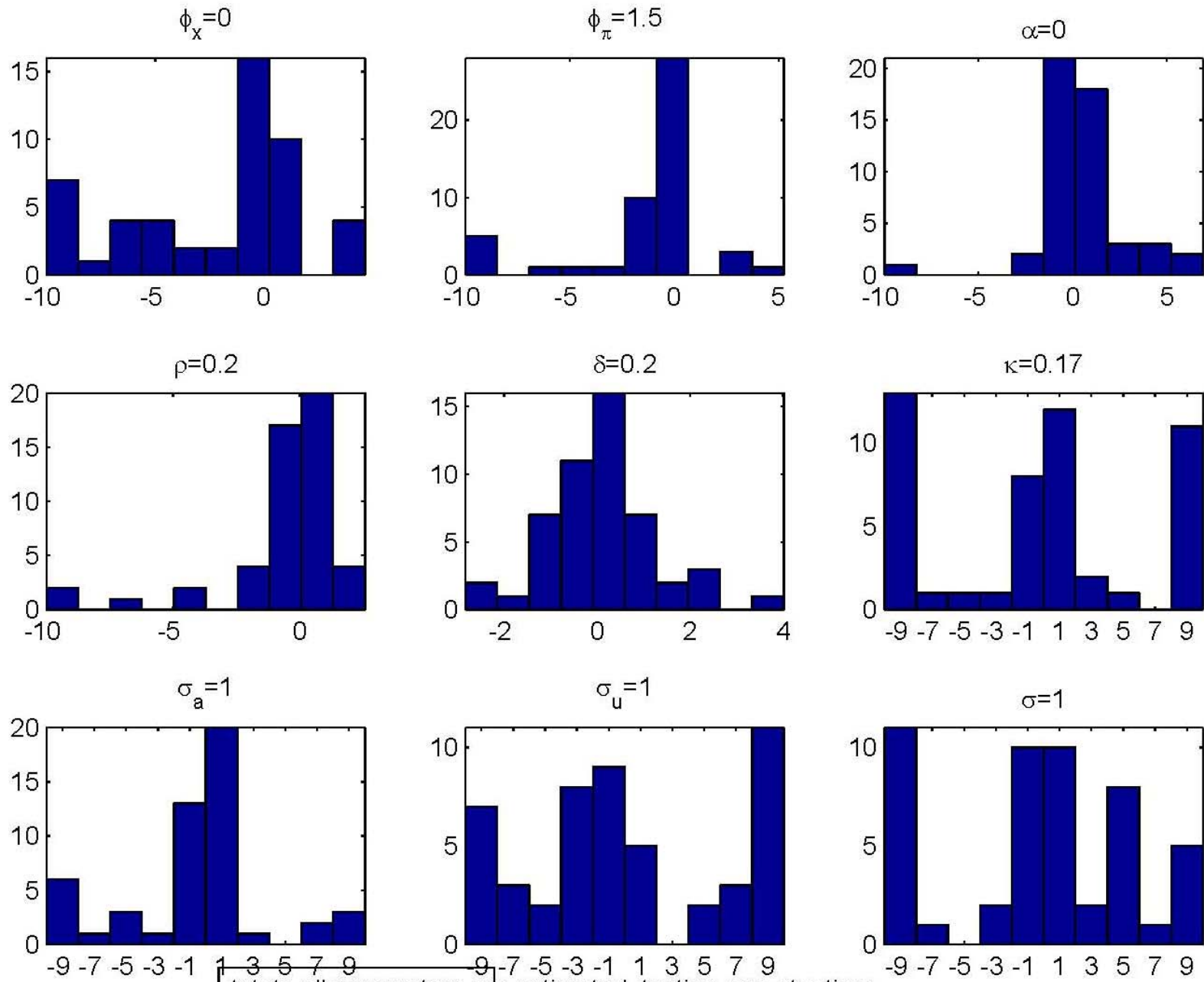
	ϕ_x	ϕ_π	α	ρ	δ	κ	σ_a	σ_u	σ_π
True values:	0.15	1.5	0.8	0.9	0.2	0.172	1	1	1
(a) LR test statistic									
LR test size	0	0	0.08	0.08	0	0.04	0.08	0	0.02
KS chi-squared? (p -value)	0.045	0.027	0.704	0.061	0.738	0.258	0.143	0.100	0.258
KS statistic	0.189	0.204	0.098	0.183	0.095	0.140	0.159	0.170	0.140
(b) Wald t statistic									
Wald t size	0.46	0.26	0.20	0.14	0.14	0.60	0.34	0.72	0.64
fraction $ t > 10$	0.14	0.10	0.02	0.04	0.00	0.44	0.18	0.36	0.26
(c) Point estimator									
p-value testing MLE normality (Lilliefors)	0	0	0.01	>0.5	>0.5	0	0	0	0
value of KS stat	0.275	0.253	0.145	0.080	0.067	0.208	0.183	0.278	0.214
bias	0.44	0.71	0.00	-0.00	0.00	0.13	0.36	0.57	0.18
sqrt(mean(SE ²))	4.379	7.693	0.655	0.154	0.282	0.221	1.825	1.106	0.291
RMSE	1.065	1.615	0.077	0.085	0.033	0.390	0.998	1.342	0.536

t -statistics, all parameters estimated (MC simulation)



ML-estimates: all parameters are estimated

MLE point estimates, all parameters estimated (MC simulation)



tstats all parameters are estimated. testing one at a time

Some final comments and references

Model comparison and assessment

- Bayesian model assessment/model comparison – see Rabanal and Rubio-Ramirez (2005) and Del Negro, Schorfheide, Smets, and Wouters (2007)
- Model assessment via forecasting – see Adolfson, Lindé, and Villani (2007) and Edge, Kiley, and Laforde (2008)
 - Methods for forecast comparisons will be covered in Lecture 14

Higher order solution methods (not just linearized).

- Higher order perturbations: Schmitt-Grohe and Uribe (2004)
- Particle filter: Fernandez-Villaverde and Rubio Ramirez (2007), An and Schorfheide (2007)

Some final comments and references, ctd.

- Some evidence that estimation is better behaved using higher order approximations:
 - An and Schorfheide (2007)
 - Fernandez Fernandez-Villaverde and Rubio Ramirez (2007)
 - DeJong and Dave (2007), ch. 11.3, comparison of MLEs using KF, a nonlinear approximation, and an exact (particle filter) solution (6 parameter optimal growth model)