Bayesian Methods for DSGE models

Fabio Canova
EUI and CEPR
September 2012

Outline

- Bayes Theorem.

- Prior Selection.

- Posterior Simulators.

- Robustness.

- Estimation of DSGE.

- Topics: Prior elicitation, DSGE-VAR, data rich DSGE, data selection, dealing with trends, non-linear DSGE.

References

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer and Verlag.

Bauwens, L., M. Lubrano and J.F. Richard (1999) *Bayesian Inference in Dynamics Econometric Models*, Oxford University Press.

Gelman, A., J. B. Carlin, H.S. Stern and D.B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, London.

Poirier, D. (1995) Intermediate Statistics and Econometrics, MIT Press.

Casella, G. and George, E. (1992) Explaining the Gibbs Sampler *American Statistician*, 46, 167-174.

Chib, S. and Greenberg, E. (1995) Understanding the Hasting-Metropolis Algorithm, *The American Statistician*, 49, 327-335.

Chib, S. and Greenberg, E. (1996) Markov chain Monte Carlo Simulation methods in Econometrics, *Econometric Theory*, 12, 409-431.

Geweke, J. (1995) Monte Carlo Simulation and Numerical Integration in Amman, H., Kendrick, D. and Rust, J. (eds.) *Handbook of Computational Economics* Amsterdaam, North Holland, 731-800.

Kass, R. and Raftery, A (1995), Empirical Bayes Factors, *Journal of the American Statistical Association*, 90, 773-795.

Sims, C. (1988) " Bayesian Skepticism on unit root econometrics", *Journal of Economic Dynamics and Control*, 12, 463-474.

Tierney, L (1994) Markov Chains for Exploring Posterior Distributions (with discussion), *Annals of Statistics*, 22, 1701-1762.

Del Negro, M. and F. Schorfheide (2003), " Priors from General Equilibrium Models for VARs", *International Economic Review*, 45, 643-673.

Del Negro M. and Schorfheide, F. (2008) Forming priors for DSGE models (and how it affects the assessment of nominal rigidities), *Journal of Monetary Economics*, 55, 1191-1208.

Kadane, J., Dickey, J., Winkler, R. , Smith, W. and Peters, S., (1980), Interactive elicitation of opinion for a normal linear model, *Journal of the American Statistical Association*, 75, 845-854.

Beaudry, P. and Portier, F (2006) Stock Prices, News and Economic Fluctuations, *American Economic Review*, 96, 1293-1307.

Boivin, J. and Giannoni, M (2006) DSGE estimation in data rich environments, University of Montreal working paper

Canova, F., (1998), "Detrending and Business Cycle Facts", *Journal of Monetary Economics*, 41, 475-540.

Canova, F., (2010), "Bridging DSGE models and the data", manuscript.

Canova, F., and Ferroni, F. (2011), "Multiple filtering device for the estimation of DSGE models", *Quantitative Economics*, 2, 73-98.

Chari, V., Kehoe, P. and McGratten, E. (2009) "New Keynesian models: not yet useful for policy analysis, *American Economic Journal: Macroeconomics*, 1, 242-266.

Guerron Quintana, P. (2010), "What you match does matter: the effects of data on DSGE estimation", *Journal of Applied Econometrics*, 25, 774-804.

Ireland, P. (2004) A method for taking Models to the data, *Journal of Economic Dynamics and Control*, 28, 1205-1226.

Stock, J. and Watson, M. (2002) Macroeconomic Forecasting using Diffusion Indices, Journal of Business and Economic Statistics, 20, 147-162.

Smets, F. and Wouters, R (2003), An estimated dynamic stochastic general equilibrium model of the euro area, *Journal of European Economic Association*, 1, 1123–1175.

Watson, M. (1993) "Measures of Fit for Calibrated Models", *Journal of Political Economy*, 101, 1011-1041.

1 Preliminaries

Classical and Bayesian analysis differ on a number of issues

Classical analysis:

• Probabilities = limit of the relative frequency of the event.

• Parameters are fixed, unknown quantities.

• Unbiased estimators useful because average value of sample estimator converge to true value via some LLN. Efficient estimators preferable because they yield values closer to true parameter.

• Estimators and tests are evaluated in repeated samples (to give correct result with high probability).

Bayesian analysis:

• Probabilities = degree of (typically subjective) beliefs of a researcher in an event.

• Parameters are random with a probability distributions.

• Properties of estimators and tests in repeated samples uninteresting: beliefs not necessarily related to relative frequency of an event in large number of hypothetical experiments.

• Estimators are chosen to minimize expected loss functions (expectations taken with respect to the posterior distribution), conditional on the data. Use of probability to quantify uncertainty.

In large samples (under appropriate regularity conditions):

• Posterior mode $\alpha^* \xrightarrow{P} \alpha_0$ (Consistency)

• Posterior distribution converges to a normal with mean α_0 and variance $(T \times I(\alpha_0))^{-1}$, where $I(\alpha)$ is Fisher's information matrix (Asymptotic normality).

Classical and Bayesian analyses differ in small samples and for dealing with unit root processes.

Bayesian analysis requires:

- Initial information $\longrightarrow$ Prior distribution.
- Data $\longrightarrow$ Likelihood.
- Prior and Likelihood $\longrightarrow$ Bayes theorem $\longrightarrow$ Posterior distribution.
- Can proceed recursively (mimic economic learning).

2 Bayes Theorem

Parameters of interest $\alpha \in A$, A compact. Prior information $g(\alpha)$. Sample information $f(y|\alpha) \equiv \mathcal{L}(\alpha|y)$.

- Bayes Theorem.

$$g(\alpha|y) = \frac{f(y|\alpha)g(\alpha)}{f(y)} \propto f(y|\alpha)g(\alpha) = \mathcal{L}(\alpha|y)g(\alpha) \equiv \grave{g}(\alpha|y)$$

$f(y) = \int f(y|\alpha)g(\alpha)d\alpha$ is the unconditional sample density (Marginal likelihood), and it is constant from the point of view of $g(\alpha|y)$; $g(\alpha|y)$ is the posterior density, $\grave{g}(\alpha|y)$ is the posterior kernel, $g(\alpha|y) = \frac{\grave{g}(\alpha|y)}{\int \grave{g}(\alpha|y)d\alpha}$.

• $f(y)$ it is a measure of fit. It tells us how good the model is in reproducing the data, not at a single point, but on average over the parameter space.

• α are regression coefficients, structural parameters, etc.; $g(\alpha|y)$ is the conditional probability of α, given what we observe, y.

• Theorem uses rule: $P(A, B) = P(A|B)P(B) = P(B|A)P(A)$. It says that if we start from some beliefs on α, we may modify them if we observe y. It does not says what the initial beliefs are, but how they should change is data is observed.

To use Bayes theorem we need:

a) Formulate prior beliefs, i.e. choose $g(\alpha)$.

b) Formulate a model for the data (the conditional probability of $f(y|\alpha)$).

Then, after observing the data, we treat the model as the likelihood of α conditional on y, and update beliefs about α.

• Bayes Theorem with two (N) samples.

Suppose $y_t = [y_{1t}, y_{2t}]$ and that y_{1t} is independent of y_{2t}. Then

$$\breve{g} \equiv f(y_1, y_2|\alpha)g(\alpha) \;=\; f_2(y_2|\alpha)f_1(y_1|\alpha)g(\alpha) \propto f_2(y_2|\alpha)g(\alpha|y_1) \quad (1)$$

Posterior for α is obtained finding first the posterior of using y_{1t} and then, treating it as a prior, finding the posterior using y_{2t}.

- Sequential learning.

- Can use data from different regimes.

- Can use data from different countries.

2.1 Likelihood Selection

• It should reflect an economic model.

• It must represent well the data. Misspecification problematic since it spills across equations and makes estimates uninterpretable.

• For the purpose of this class, the likelihood is simply the DSGE model you use

2.2 Prior Selection

• Three methods to choose priors in theory. Two not useful for DSGE models since are designed for for models which are linear in the parameters.

1) Non-Informative subjective. Choose **reference priors** because they are invariant to the parametrization.

- Location invariant prior: $g(\alpha)$ =constant (=1 for convenience). Scale invariant prior $g(\sigma) = \sigma^{-1}$.

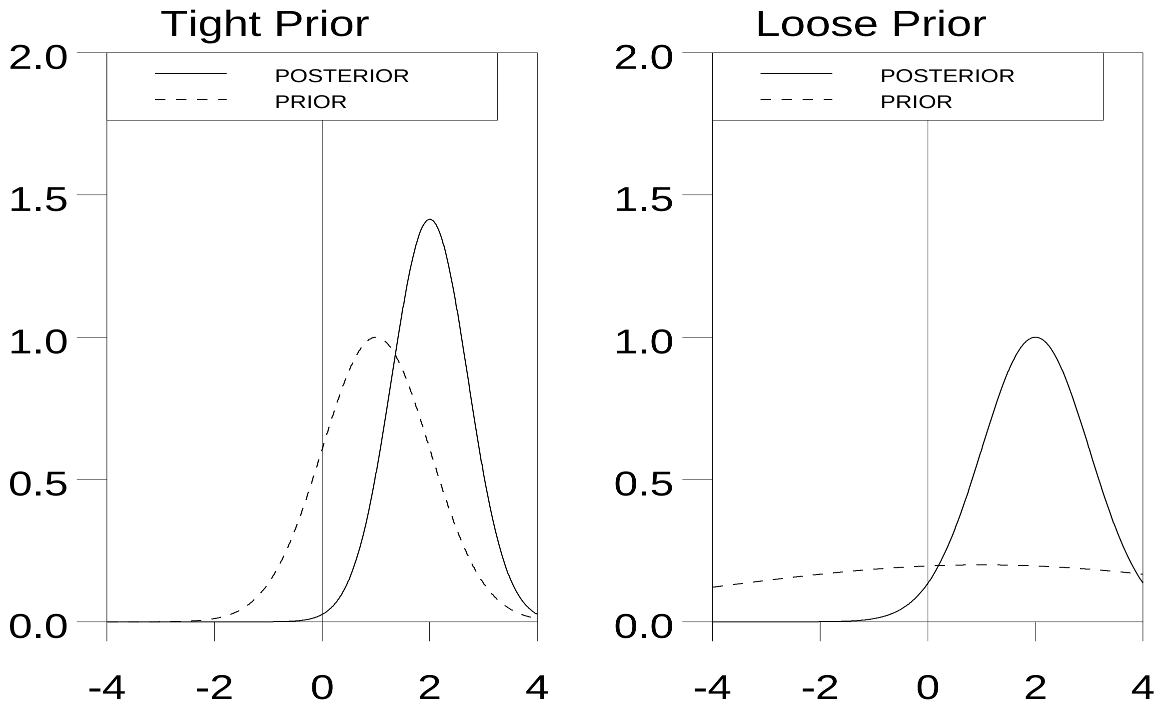
- Location-scale invariant prior : $g(\alpha, \sigma) = \sigma^{-1}$.

• Non-informative priors useful because many classical estimators (OLS, ML) are Bayesian estimators with non-informative priors

2) Conjugate Priors

A prior is conjugate if the posterior has the same form as the prior. Hence, the form posterior will be analytically available, only need to figure out its posterior moments.

• Important result in linear models: Posterior moments $=$ weighted average of sample and prior information. Weights $=$ relative precision of sample and prior informations.



Tight Prior
Loose Prior
POSTERIOR
PRIOR
2.0
1.5
1.0
0.5
0.0
-4
-2
0
2
4

3) Objective priors and ML-II approach. Based on:

$$f(y) = \int \mathcal{L}(\alpha|y)g(\alpha)d\alpha \equiv \mathcal{L}(y|g) \tag{2}$$

Since $\mathcal{L}(\alpha|y)$ is fixed, $\mathcal{L}(y|g)$ reflects the plausibility of g in the data.
If g_1 and g_2 are two priors and $\mathcal{L}(y|g_1) > \mathcal{L}(y|g_2)$, there is better support for g_1. Hence, can estimate the "best" g using $\mathcal{L}(y|g)$.

In practice, set $g(\alpha) = g(\alpha|\theta)$, where $\theta=$ hyperparameters (e.g. the mean and the variance of the prior). Then $\mathcal{L}(y|g) \equiv \mathcal{L}(y|\theta)$.

The θ that maximizes $\mathcal{L}(y|\theta)$ is called ML-II estimator and $g(\alpha|\theta_{ML})$ is ML-II based prior.

Important:

- $y_1, \dots y_T$ **should not** be the same sample used for inference.

- $y_1, \dots y_T$ could represent past time series information, cross sectional/ cross country information.

- Typically $y_1, \dots y_T$ is called "Training sample".

4) Priors for DSGE - similar to MLII priors.

• Assume that $g(\alpha) = g_1(\alpha_1)g_2(\alpha_2)....g_q(\alpha_q)$.

• Use a conventional format for the distributions: a Normal, Beta and Gamma for individual parameters. Choose moments in data based fashion.

• Mean $=$ calibrated parameters, variance: subjective.

• Careful about circularity: priors too much data based!!

• Careful about standard errors: multivariate priors often too tight!!

See Later del Negro and Schorfheide (2008) for formally choosing data based priors.

Summary

Inputs of the analysis: $g(\alpha)$, $f(y|\alpha)$.

Outputs of the analysis:

$g(\alpha|y) \propto f(y|\alpha)g(\alpha)$ (posterior),

$f(y) = \int f(y|\alpha)g(\alpha)$ (marginal likelihood), and

$f(y^{T+\tau}|y^T)$ (predictive density of future observations).

Likelihood should reflect data/ economic theory.

Prior could be non-informative, conjugate, data based (objective).

- In simple examples $f(y)$ and $g(\alpha|y)$ can be computed analytically.

- In general, can only be computed numerically by Monte Carlo methods.

- If the likelihood is a (log-linearized) DSGE model: always need numerical computations.

3 Posterior simulators

Objects of interest for Bayesian analysis: $E(h(\alpha)) = \int h(\alpha)g(\alpha|y)d\alpha$. Occasionally, can evaluate the integral analytically. In general, it is impossible.

If $g(\alpha|y)$ were available: we could compute $E(h(\alpha))$ numerically:

- Draw α^l from $g(\alpha|y)$. Compute $h(\alpha^l)$

- Repeat draw L times. Average $h(\alpha^l)$ over draws.

Example 3.1 *Suppose we are interested in computing $Pr(\alpha > 0)$. Draw α^l from $g(\alpha|y)$. If $\alpha^l > 0$, set $h(\alpha^l) = 1$, else set $h(\alpha^l) = 0$. Repeat the draw L times and average $h(\alpha^l)$ over draws. The average is an estimate of $Pr(\alpha > 0)$.*

• Approach works because draws are iid and the law of large numbers (LLN) insures that sample averages converge to population averages (ergodicity).

• By a central limit theorem (CLT) the difference between sample and population averages has a normal distribution with zero mean and some variance as L grows (numerical standard errors can be used a a measure of accuracy).

- However $g(\alpha|y)$ is not analytically available. Use a $g^{AP}(\alpha|y)$, which is similar to ($g(\alpha|y)$, and easy to draw from.

Many possibilities:

• Normal Approximation

• Basic Posterior simulators (Acceptance and Importance sampling).

• Markov Chain Monte Carlo (MCMC) methods

3.1 Normal posterior analysis

If T is large $g(\alpha|y) \approx f(\alpha|y)$. If $f(\alpha|y)$ is unimodal, roughly symmetric, α^* (mode) is in the interior of A then

$$\log g(\alpha|y) \approx \log g(\alpha^*|y) + 0.5(\alpha - \alpha^*)'[\frac{\partial^2 \log g(\alpha|y)}{\partial\alpha\partial\alpha'}|_{\alpha=\alpha^*}](\alpha - \alpha^*) \quad (3)$$

Since $g(\alpha^*|y)$ is constant, letting $\Sigma_{\alpha^*} = -[\frac{\partial^2 \log g(\alpha|y)}{\partial\alpha\partial\alpha'}{}^{-1}|_{\alpha=\alpha^*}]$

$$g(\alpha|y) \approx N(\alpha^*, \Sigma_{\alpha^*}) \quad (4)$$

- An approximate 100(1-ρ)% highest credible set is $\alpha^* \pm \Phi(\rho/2)I(\alpha^*)^{-0.5}$ where $\Phi(.)$ the CDF of a standard normal.

• Approximation is valid under regularity conditions when $T \to \infty$ or when the posterior kernel is roughly normal. It is highly inappropriate when:

- Likelihood function flat in some dimension ($I(\alpha^*)$ badly estimated).

- Likelihood function is unbounded (no posterior mode exists).

- Likelihood function has multiple peaks.

- α^* is on the boundary of A (quadratic approximation wrong).

- $g(\alpha) = 0$ in a neighborhood of α^* (quadratic approximation wrong).

How do we construct a normal approximation?

A) Find the mode of the posterior.

$$\max \log g(\alpha|y) = \max(\log L(\alpha|y) + \log g(\alpha))$$

- Problem is identical to the one of finding the maximum of a likelihood. Simply the function is different.

Two mode finding algorithms:

i) Newton algorithm

- Let $L = \log g(\alpha|y)$ or $L = \log \breve{g}(\alpha|y)$. Choose α_0.

- Calculate $L' = \frac{\partial L}{\partial \alpha}(\alpha_0)$ $L^{''} = \frac{\partial^2 L}{\partial \alpha \partial \alpha'}(\alpha_0)$. Approximate L quadratically.

- Set $\alpha^l = \alpha^{l-1} - \gamma(L^{''}(\alpha^{l-1}|y))^{-1}(L'(\alpha^{l-1}|y))$ $\gamma \in (0, 1)$.

- Iterate until convergence i.e. until $||\alpha^l - \alpha^{l-1}|| < \iota$, ι small.

Fast and good if α_0 is good and L close to quadratic. Bad if $L^{''}$ not positive definite.

ii) Conditional maximization algorithm.

Let $\alpha = (\alpha_1, \alpha_2)$. Start from some $(\alpha_{10}, \alpha_{20})$. Then

- Maximize $g(\alpha_1, \alpha_2)$ with respect to α_1 keeping α_{20} fixed. Let α_1^* the maximizer.

- Maximize $g(\alpha_1, \alpha_2)$ with respect to α_2 keeping $\alpha_1 = \alpha_1^*$ fixed. Let α_2^* the maximizer.

- Iterate on two previous steps until convergence.

- Start from different $(\alpha_{10}, \alpha_{20})$, check if maximum is global.

B) Compute the variance covariance matrix at the mode

- Use the Hessian $\Sigma_{\alpha^*} = -[\frac{\partial^2 \log g(\alpha|y)}{\partial\alpha\partial\alpha'}^{-1}|_{\alpha=\alpha^*}]$

C) Approximate the posterior density around the mode.

- If only one peak is present: $g^{AP}(\alpha|y) = \mathbb{N}(\alpha^*, \Sigma_{\alpha^*})$.

- If multiple modes are present, find an approximation to each mode, and set $g^{AP}(\alpha|y) = \sum_i \varrho_i \mathbb{N}(\alpha_i^*, \Sigma_{\alpha_i^*})$ where $0 \leq \varrho_i \leq 1$. If modes are clearly separated select $\varrho_i = g(\alpha_i^*|y)|\Sigma_{\alpha_i^*}|^{-0.5}$.

- If the sample is small use a t-approximation i.e. $g^{AP}(\alpha|y) = \sum_i \varrho_i g(\tilde{\alpha}|y)[\nu + (\alpha - \alpha_i^*)'\Sigma_{\alpha_i}(\alpha - \alpha_i^*)]^{-0.5(k+v)}$ with small ν.

(If $\nu = 1$ t-distribution=Cauchy distribution, large overdispersion. Typically $\nu = 4, 5$ appropriate).

D) To conduct inference, draw α^l from $g^{AP}(\alpha|y)$.

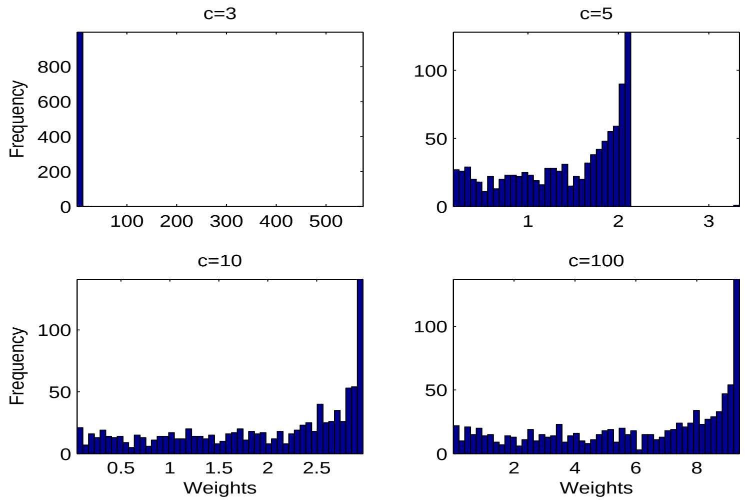
If draws are iid, $E(h(\alpha)) = \frac{1}{L}\sum_l h(\alpha^l)$. Use LLN to approximate any posterior probability contours of $h(\alpha)$, e.g. a 16-84 range is $[h(\alpha^{16}), h(\alpha^{84})]$.

E) Check accuracy of approximation.

Compute *Importance Ratio* $IR^l = \frac{\breve{g}(\alpha^l|y)}{g^{AP}(\alpha^l|y)}$. Accuracy is good if IR^l is constant across l. If not, need to use other techniques.

Note: Importance ratios are not automatically computed in Dynare. Need to do it yourself.

Example 3.2 *True:* $g(\alpha|y)$ *is t(0,1,2). Approximation: N(0,c), where* $c = 3, 5, 10, 100$.



Horizontal axis=importance ratio weights, vertical axis= frequency of the weights.

- Posterior has fat tails relative to a normal (poor approximation).

3.2 Basic Posterior Simulators

- Draw from a general $g^{AP}(\alpha|y)$ (not necessarily normal).

- Non-iterative methods - $g^{AP}(\alpha|y)$ is fixed across draws.

- Work well when IR^{l} is roughly constant across draws.

A) Acceptance sampling

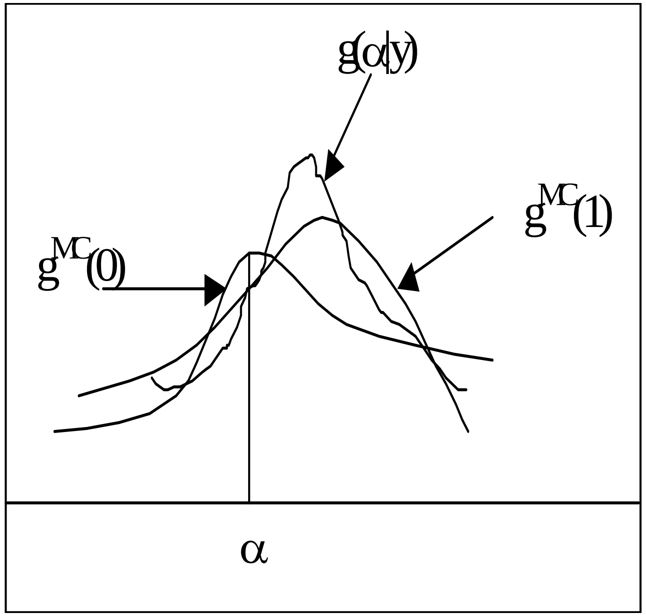
B) Importance sampling

3.3 Markov Chain Monte Carlo Methods

• Problem with basic simulators: approximating density selected once and for all. If mistakes are made, they stay. With MCMC location of approximating density changes as iterations progress.

• Idea: Suppose n states $(x_1, \ldots x_n)$. Let $P(i,j) = Pr(x_{t+1} = x_j | x_t = x_i)$ and let $\mu(t) = (\mu_{1t}, \ldots \mu_{nt})$ be the unconditional probability at t of each state n. Then $\mu(t+1) = P\mu(t) = P^t\mu(0)$ and μ is an equilibrium (ergodic, steady state, invariant) distribution if $\mu = \mu P$.

Set $\mu = g(\alpha|y)$, choose $\mu(0)$ some initial density and P some transition across states. If conditions are right, iterate on $\mu(0)$ and limiting distribution will be $g(\alpha|y)$, the unknown posterior.



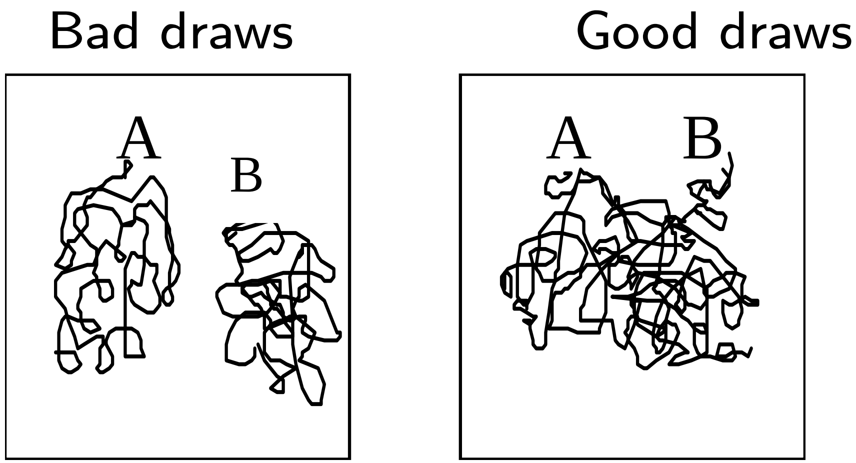
• Under general conditions, the ergodicity of P insures consistency and asymptotic normality of estimates of any $h(\alpha)$.

Need a transition $P(\alpha, A)$, where A is some set, such that $||P(\alpha, A) - \mu(\alpha)|| \rightarrow 0$ in the limit. For this need that the chain associated with P :

• is irreducible, i.e. it has no absorbing state

• is aperiodic, i.e. it does not cycle across a finite number of states.

• it is Harris recurrent, i.e. each cell is visited an infinite number of times with probability one.



Result 1: A reversible Markov chain, has an ergodic distribution (existence). (if $\mu_i P_{i,j} = \mu_j P_{j,i}$ then $(\mu P)_j = \sum_j \mu_i P_{i,j} = \sum_i \mu_j P_{j,i} = \mu_j \sum_i P_{j,i} = \mu_j$.)

Result 2: (Tierney (1994)) (uniqueness) If a Markov chain is Harris recurrent and has a proper invariant distribution. $\mu(\alpha)$, $\mu(\alpha)$ is unique.

Result 3: (Tierney(1994)) (convergence) If a Markov chain with invariant $\mu(\alpha)$ is Harris recurrent and aperiodic, for all $\alpha_0 \in A$ and all A, as $L \to \infty$.

- $||P^L(\alpha_0, A) - \mu(\alpha)|| \to 0$, $||.||$ is the total variation distance.

- For all $h(\alpha)$ absolutely integrable with respect to $\mu(\alpha)$.

- $\lim_{L\to\infty} \frac{1}{L}\sum_{l=1}^{L} h(\alpha^l) \overset{a.s.}{\to} \int h(\alpha)\mu(\alpha)d\alpha$.

If chain has a finite number of states, sufficient for the chain to be irreducible, Harris recurrent and aperiodic: is that $P(\alpha^l \in A_1|\alpha^{l-1} = \alpha_0, y) > 0$, all $\alpha_0, A_1 \in A$.

- Can dispense with the finite number of state assumption.

- Can dispense with the first order Markov assumption.

General simulation strategy:

- Choose starting values α_0, choose a P with the right properties.
- Run MCMC simulations.
- Check convergence.
- Summarize results, after you have discarded some initial set of draws.

1) MCMC methods generate draws which are *correlated* (with normal/basic simulators, posterior draws are iid).

2) MCMC methods generate draws from posterior only after a burn-in period (with normal/basic simulators, first draw is from the posterior).

3) MCMC Can be used to explore intractable likelihoods using "data augmentation" technique (non-bayesian method).

4) MCMC methods only need $\breve{g}(\alpha|y)$ (no knowledge of the normalizing constants is needed).

3.3.1 Metropolis-Hastings algorithm

MH is a general purpose MCMC algorithm that can be used when faster methods (such as the Gibbs sampler) are either not usable or difficult to implement.

Starts from an arbitrary transition function $q(\alpha^{\dagger}, \alpha^{l-1})$, where $\alpha^{l-1},\ \alpha^{\dagger} \in A$ and an arbitrary $\alpha^{0} \in A$. For each $l = 1, 2, \ldots L$.

- Draw $\alpha^{\dagger}$ from $q(\alpha^{\dagger}, \alpha^{l-1})$ and draw $\varpi \sim U(0, 1)$.

- If $\varpi < \mathfrak{E}(\alpha^{l-1}, \alpha^{\dagger}) = [\frac{\breve{g}(\alpha^{\dagger}|Y)q(\alpha^{\dagger},\alpha^{l-1})}{\breve{g}(\alpha^{l-1}|Y)q(\alpha^{l-1},\alpha^{\dagger})}]$, set $\alpha^{\ell} = \alpha^{\dagger}$.

- Else set $\alpha^{\ell} = \alpha^{\ell-1}$.

These iterations define a mixture of continuous and discrete transitions:

$$\begin{aligned} P(\alpha^{l-1},\ \alpha^{l}) &= q(\alpha^{l-1},\alpha^{l})\mathfrak{E}(\alpha^{l-1},\alpha^{l}) \qquad \text{if } \alpha^{l}\neq\alpha^{l-1} \\ &= 1-\int_{A} q(\alpha^{l-1},\alpha)\mathfrak{E}(\alpha^{l-1},\alpha)d\alpha \ \text{ if } \alpha^{l}=\alpha^{l-1} \end{aligned} \tag{5}$$

$P(\alpha^{l-1},\ \alpha^{l})$ satisfies the conditions needed for existence, uniqueness and convergence.

• Idea: Want to sample from highest probability region but want to visit as much as possible the parameter space. How to do it? Choose an initial vector and a candidate, compute kernel of posterior at the two vectors. If you go uphill keep draw. If not, keep the draw with probability ϖ.

If $q(\alpha^{l-1}, \alpha^\dagger) = q(\alpha^\dagger, \alpha^{l-1})$, (Metropolis version of the algorithm) $\mathfrak{E}(\alpha^{l-1}, \alpha^\dagger) = \frac{\breve{g}(\alpha^{l-1}|Y)}{\breve{g}(\alpha^\dagger|Y)}$. If $\mathfrak{E}(\alpha^{l-1}, \alpha^\dagger) > 1$ the chain moves to $\alpha^\dagger$. Hence always move uphill; if the draw moves downhill stay at α^{l-1} with probability $1 - \mathfrak{E}(\alpha^{l-1}, \alpha^\dagger)$ and explore new areas with probability equal to $\mathfrak{E}(\alpha^{l-1}, \alpha^\dagger)$.

Important: $q(\alpha^{l-1}, \alpha^\dagger)$ is not necessarily equal (proportional) to posterior - histograms of draws not equal to the posterior. This is why we need to use a scheme which accepts more in the regions of high probability.

How do you choose $q(\alpha^{l-1}, \alpha^{\dagger})$ (the transition probability)?

- Typical choice: random walk chain. $q(\alpha^{\dagger}, \alpha^{l-1}) = q(\alpha^{\dagger} - \alpha^{l-1})$, and $\alpha^{\dagger} = \alpha^{l-1} + v$ where $v \sim \mathbb{N}(0, \sigma_v^2)$. To get "reasonable" acceptance rates adjust σ_v^2. Often $\sigma_v^2 = c * \Omega_{\alpha}, \Omega_{\alpha} = [-g^{''}(\alpha^*|y)]^{-1}$. Choose c.

- Reflecting random walk: $\alpha^{\dagger} = \mu + (\alpha^{l-1} - \mu) + v$

- Independent chain $q(\alpha^{\dagger}, \alpha^{l-1}) = \bar{q}(\alpha^{\dagger})$,$\mathfrak{E}(\alpha^{l-1}, \alpha^{\dagger}) = min[\frac{w(\alpha^{\dagger})}{w(\alpha^{l-1})}, 1]$, where $w(\alpha) = \frac{g(\alpha|Y)}{\bar{q}(\alpha)}$. Monitor both the location and the shape of $\bar{q}$ to insure reasonable acceptance rates. Standard choices for $\bar{q}$ are normal and t.

- General rule for selecting q. A good q must:

a) be easy to sample from

b) be such that it is easy to compute $\mathfrak{E}$.

c) each move goes a reasonable distance in parameter space but does not reject too frequently (ideal rejection rate 30-50%).

Implementation issues

A) How to draw samples?

- Produce one sample (of dimension $n * L + \bar{L}$). Throw away initial $\bar{L}$ observations. Keep only elements $(L, 2L, \ldots, n * L)$ (to eliminate the serial correlation of the draws).

- Produces n samples of $\bar{L} + L$ elements. Use last L observations in each sample for inference.

- Dynare setup to produce n samples, keep the last 25 percent of the draws. **Careful: Need to make sure that with 75 percent of the draws the chain has converged**.

B) How long should be $\bar{L}$? How do you check convergence?

- Start from different α^0. Check if sample you keep, for a given $\bar{L}$, has same properties (Dynare approach).

- Choose two points, $\bar{L}_1 < \bar{L}_2$; compute distributions/moments of α after these points. If visually similar, algorithm has converged at $\bar{L}_1$. Could this recursively $\rightarrow$ CUMSUM statistic for mean, variance, etc.(checks if it settles down, no testing required).

- Fix $\bar{L}$, compute distributions/moments using $n_2 > n_1$ sampled values. If converged, distributions/moments should be similar.

For simple problems $\bar{L} \approx 50$ and $L \approx 200$.

For DSGEs $\bar{L} \approx 100,000 - 200,000$ and $L \approx 500,000$.

C) Inference : easy.

- Weak Law of Large Numbers $E(h(\alpha)) \approx \frac{1}{j}\sum_{j=1}^{n} h(\alpha^{jL})$ where α^{jL} is the $j * L$-th observation drawn after $\bar{L}$ iterations are performed.

- $E(h(\alpha)h(\alpha)') = \sum_{-J(L)}^{J(L)} w(\tau) ACF_h(\tau)$; $ACF_h(\tau) =$ autocovariance of $h(\alpha)$ for draws separated by τ periods; $J(L)$ function of L, $w(\tau)$ a set of weights.

- Marginal density $(\alpha_k^1, \ldots \alpha_k^L)$: $g(\alpha_k|y) = \frac{1}{L}\sum_{j=1}^{L} g(\alpha_k|y, \alpha_{k'}^j, k' \neq k)$.

- Predictive inference $f(y_{t+\tau}|y_t) = \int f(y_{t+\tau}|y_t, \alpha) g(\alpha|y_t) d\alpha$.

- Model comparisons: compute marginal likelihood numerically.

4 Robustness

• Typically prior chosen to make calculation convenient. How sensitive are results to prior choice?

• Typical approach (brute force): Repeat estimation for different priors (inefficient).

• Alternative.

i) Select an alternative prior $g_1(\alpha)$ with support included in $g(\alpha)$.

ii) Let $w(\alpha) = \frac{g(\alpha)}{g_1(\alpha)}$. Then $h_1(\alpha) = \int (h(\alpha) w(\alpha) dg_1(\alpha)$ can be approximated using $h_1(\alpha) \approx \frac{\frac{1}{L}\sum_l w(\alpha^l) h(\alpha^l)}{\sum_l w(\alpha^l)}$

Example 4.1 $y_t = x_t\alpha + u_t$ $u_t \sim (0, \sigma^2)$. *Suppose* $g(\alpha)$ *is* $\mathbb{N}(0, 10)$. *Then* $g(\alpha|Y)$ *is normal with mean* $\tilde{\alpha} = \tilde{\Sigma}^{-1}(0.1 + \sigma^{-2}x'x\alpha_{ols})$ *and variance* $\tilde{\Sigma} = 0.1 + \sigma^{-2}x'x,$. *If one wishes to examine how forecasts of the model change when the prior variances changes (for example to 5) two alternatives are possible:*

(a) draw from normal $g(\alpha|Y)$ *which has mean* $\tilde{\alpha}_1 = \tilde{\Sigma}_1^{-1}(0.2+\sigma^{-2}x'x\alpha_{ols})$ *and variance* $\tilde{\Sigma} = 0.2 + \sigma^{-2}x'x,$ *and compare forecasts.*

(b) Weight draws from the initial posterior distribution by $\frac{g(\alpha)}{g_1(\alpha)}$ *where* $g_1(\alpha)$ *is* $N(0, 5)$.

5 Bayesian estimation of DSGE models

Why using Bayesian methods to estimate DSGE models?

1) Hard to include non-sample information in classical ML (a part from range of possible values).

2) Classical ML is justified only if the model is the GDP of the actual data. Can use Bayesian methods for misspecified models (economic inferencemay be problematic, no problem for statistical inference).

3) Can incorporate prior uncertainty about parameters and models.

General Principles:

• Use the fact that (log-)linearized DSGE models are state space models whose reduced form parameters α are nonlinear functions of structural θ. Compute the likelihood via the Kalman filter.

• Posterior of θ can be obtained using MH algorithm.

• Use posterior output to compute the marginal likelihood, Bayes factors and any posterior function of the parameters (impulse responses, ACF, turning point predictions, forecasts, etc.).

• Check robustness to the choice of prior.

General algorithm: Given θ_0

[1.] Construct a log-linear solution of the DSGE economy. Add measurement errors if estimation performed on a vector of variables which is larger than the one of the state space.

[2.] Specify prior distributions $g(\theta)$.

[3.] Transform the actual data ($\rightarrow$ filter, detrend or use data transformations) to make sure that is conformable with the model.

[4.] Compute likelihood via Kalman filter.

[5.] Draw sequences for θ using MH algorithm. Check convergence.

[6.] Compute marginal likelihood numerically. Compute marginal likelihood for candidate alternative models. Compute Bayes factors.

[7.] Construct statistics of interest from the draws (after burned-out period). Use loss-based evaluation of discrepancy model/data.

[8.] Perform robustness exercises.

Step 1.: can have nonlinear state space models (see later and e.g. Amisano and Tristani (2006), Rubio and Villaverde (2009)) and can avoid adding measurement errors (useful only to reduce computational problems). Problem: computations much more complex.

In Step 2. typically the mean of the prior is centered around calibrated values. Standard errors reflect subjective or objective ideas (to cover the range of existing estimates). Form of the prior chosen for convenience.

In step 5. Given θ^l

i) Draw a $\theta^\dagger$ from the $\mathfrak{P}(\theta^\dagger \theta^l)$. Solve the model.

ii) Use the KF to compute the likelihood.

iii) Evaluate the posterior kernel at the draw $\breve{g}(\theta^\dagger|y) = f(y|\theta^\dagger)g(\theta^\dagger)$.

iv) Evaluate the posterior kernel at θ^l i.e $\breve{g}(\theta^0|y) = f(y|\theta^l)g(\theta^l)$.

v) Compute $IR = \frac{\breve{g}(\theta^\dagger)}{\breve{g}(\theta^l)}\frac{\mathfrak{P}(\theta^l,\theta^\dagger)}{\mathfrak{P}(\theta^\dagger,\theta^l)}$. If $IR > 1$ set $\theta^{l+1} = \theta^\dagger$.

vi) Else draw $\varpi \sim U(0,1)$. If $\varpi < IR$ set $\theta^{l+1} = \theta^\dagger$ otherwise set $\theta^{l+1} = \theta^l$.

vii) Repeat i)-vi) $\bar{L} + nL$ times. Throw away $\bar{L}$ draws. Keep one every n for inference.

viii) Estimate marginal/ joint posteriors using kernel methods. Compute point estimate and credible sets.

ix) Compute continuous functions $h(\theta)$ of interest. Set up a loss function. Compare models using the risk function.

In Step 6. use a modified harmonic mean estimator i.e. approximate $\mathcal{L}(y_t|\mathcal{M}_i)$ using $[\frac{1}{L}\sum_l \frac{f(\alpha_l^i)}{\mathcal{L}(y_t|\alpha_l^i,\mathcal{M}_i)g(\alpha_l^i|\mathcal{M}_i)}]^{-1}$ where α_i^l is the draw l of the parameters α of model i and f is a density with tails thicker than a normal. If $f(\alpha_l^i) = 1$ we have a simple harmonic mean estimator.

Competitors could be a) more densely parametrized structural model (nesting the interested one).

b) more densely parametrized reduced form model (e.g. VAR or a BVAR). Bayes factors can be computed numerically or via Laplace approximations (to decrease computational burden in large scale systems).

In step 8. Reweight the draws appropriately.

Example 5.1 *(One sector growth model)*

- *Analytic solution if* $U(c,l) = \ln c$ *and* $\delta = 1$. *Equations are:*

$$
\begin{aligned}
K_{t+1} &= (1-\eta)\beta A K_t^{1-\eta}\zeta_t + u_{1t} && (6) \\
GDP_t &= A K_t^{1-\eta}\zeta_t + u_{2t} && (7) \\
c_t &= \eta\beta GDP_t + u_{3t} && (8) \\
r_t &= (1-\eta)\frac{GDP_t}{K_t} + u_{4t} && (9)
\end{aligned}
$$

- ζ_t *technology shock,* u_{jt} *measurement errors added to break singularity of the data.*

Parameters: β*: is the discount factor,* $1-\eta$*: the share of capital in production,* σ^2*: variance of technology shock, A: constant in the production function.*

Simulate 1000 points from using $k_0 = 100.0$ *using* $A = 2.86; 1 - \eta = 0.36; \beta = 0.99, \sigma^2 = 0.07$.

Assume $u_{1t} \sim \mathbb{N}(0, 0.1)$; $u_{2t}^m \sim \mathbb{N}(0, 0.06); u_{3t}^m \sim \mathbb{N}(0, 0.02);$
$u_{4t}^m \sim \mathbb{N}(0, 0.08)$; *(Note: lots of measurement error!)*

- Keep last 160 as data (to mimic about 40 years of quarterly data).

Interested in $(1 - \eta), \beta$ *i.e (treat* σ^2, A *as fixed).*

Use (8)-(9) to identify the parameters from the data.

Priors: $(1-\eta) \sim$ *Beta(3,5);* $\beta \sim$ *Beta(98,2) (NOTATION DIFFERENT FROM DYNARE)*

Mean of a Beta(a,b) is $(a/a+b)$ *and the variance of a Beta(a,b)is* $ab/[(a+b)^2 * (a+b+1)]$*. Thus prior mean of* $1-\eta = 0.37$*, prior variance* 0.025*; prior mean of* $\beta = 0.98$*, prior variance* 0.0001*.*

Let $\theta = (1-\eta, \beta)$ *Use random walk to draw* $\theta^{\dagger}$*, i.e.* $\theta^{\dagger} = +\theta^{l-1} + e^{\dagger}$*,* μ *is the mean and* e_i *is* $U(-0.08, 0.08)$ *for* β *and* $U(-0.06, 0.06)$ *for* η *(roughly about 28% acceptance rate).*

Draw 10000 replications from the posterior kernel. Convergence is fast.

Keep last 5000; use one every 5 for inference.

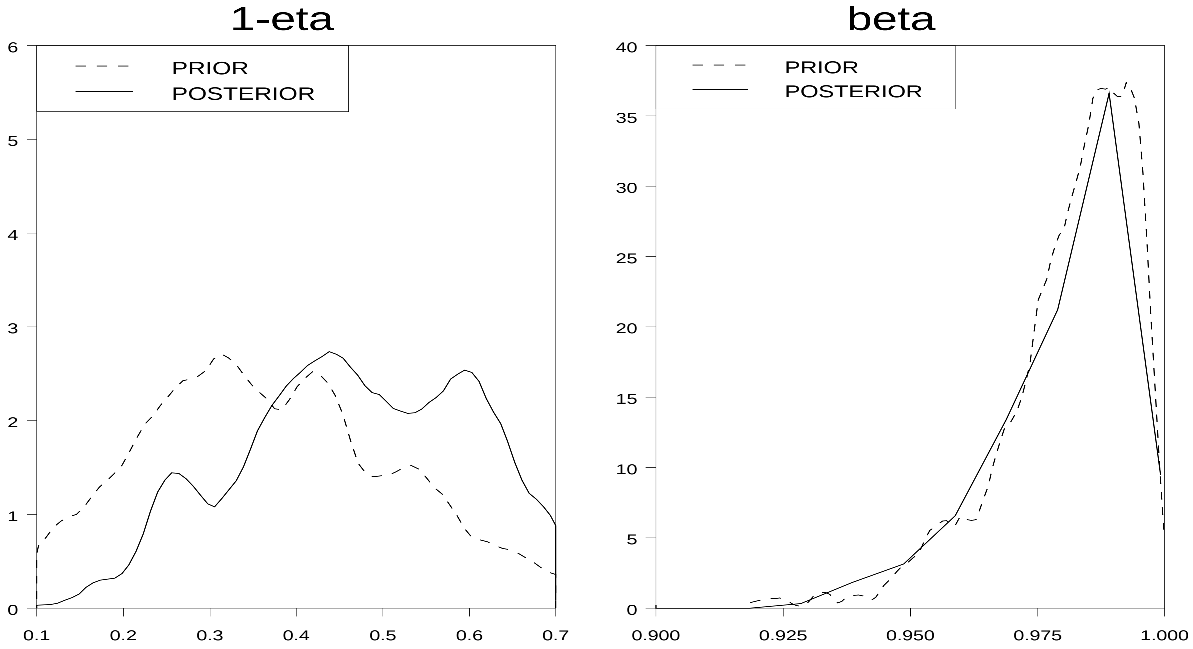


Figure 4: Priors and Posteriors, RBC model

- Prior for β sufficiently loose, posterior similar, data is not very formative.

-Posteriors centered around the true parameters,large dispersion.

Variances/covariances

	true	*posterior 68% range*
var(c)	*0.24*	*[0.11, 0.27]*
var(y)	*0.05*	*[0.03, 0.11]*
cov(c,y)	*0.0002*	*[0.0003, 0.0006]*

Wrong model

- Simulate data from model with habit $\gamma = 0.8$

- Estimate model conditioning on $\gamma = 0$.

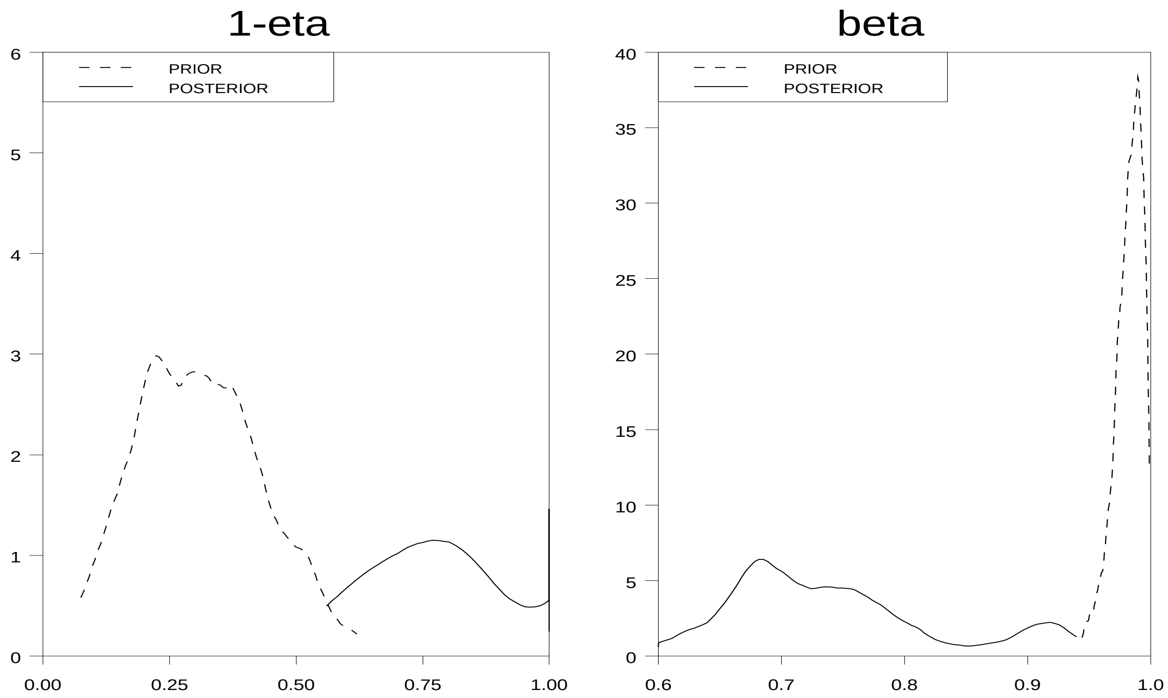


Figure 5: Priors and Posteriors, wrong model

Example 5.2 *(New Keynesian model)*

$$
\begin{aligned}
gap_t &= E_t gap_{t+1} - \frac{1}{\varphi}(r_t - E_t \pi_{t+1}) + g_t && (10)\\
\pi_t &= \beta E_t \pi_{t+1} + \kappa gap_t + v_t && (11)\\
r_t &= \phi_r r_{t-1} + (1-\phi_r)(\phi_\pi \pi_{t-1} + \phi_{gap} gap_{t-1}) + e_t && (12)
\end{aligned}
$$

$\kappa = \frac{(1-\zeta_p)(1-\beta\zeta_p)(\varphi+\vartheta_N)}{\zeta_p}$; ζ_p = *degree of (Calvo) stickiness,* β = *discount factor,* φ = *risk aversion,* ϑ_N = *elasticity of labor supply.* g_t *and* v_t *are AR(1) with persistence* ρ_g, ρ_v *and variances* σ_g^2, σ_v^2; $e_t \sim iid(0, \sigma_r^2)$.

$\theta = (\beta, \varphi, \vartheta_l, \zeta_p, \phi_\pi, \phi_{gap}, \phi_r, \rho_g, \rho_v, \sigma_v^2, \sigma_g^2, \sigma_r^2)$.

Assume $g(\theta) = \prod g(\theta_i)$ *and* $\beta \sim Beta(98, 3), \varphi \sim \mathbb{N}(1, 0.375^2), \vartheta_N \sim \mathbb{N}(2, 0.75^2),\ \ \zeta_p \sim Beta(9, 3), \phi_r \sim Beta(6, 2), \phi_\pi \sim Normal(1.5, 0.1^2),$ $\phi_{gap} \sim \mathbb{N}(0.5, 0.05^2),\ \rho_g \sim Beta(17, 3), \rho_v \sim Beta(17, 3)\ \sigma_i^2 \sim IG(2, 0.01), i = g, v, r.$

Use random walk MH algorithm to draw candidates.

Use US linearly detrended data from 1948:1 to 2002:1 to estimate the model.

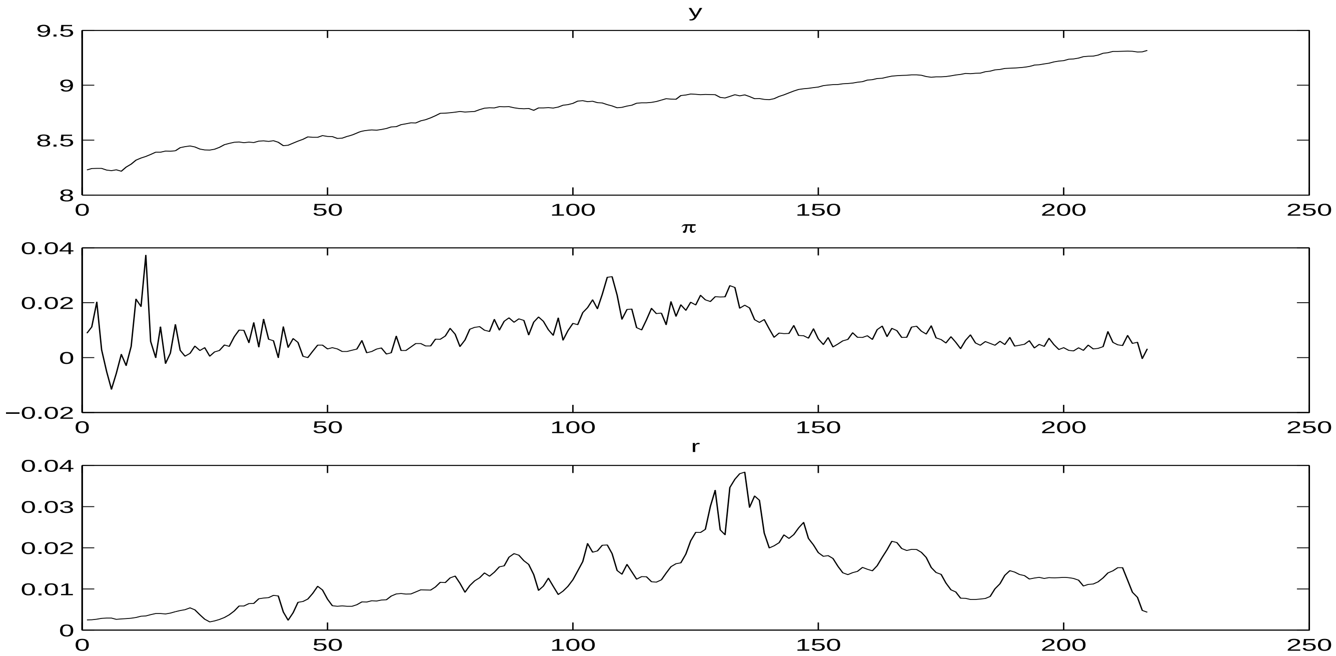


Figure 6: Raw Time series

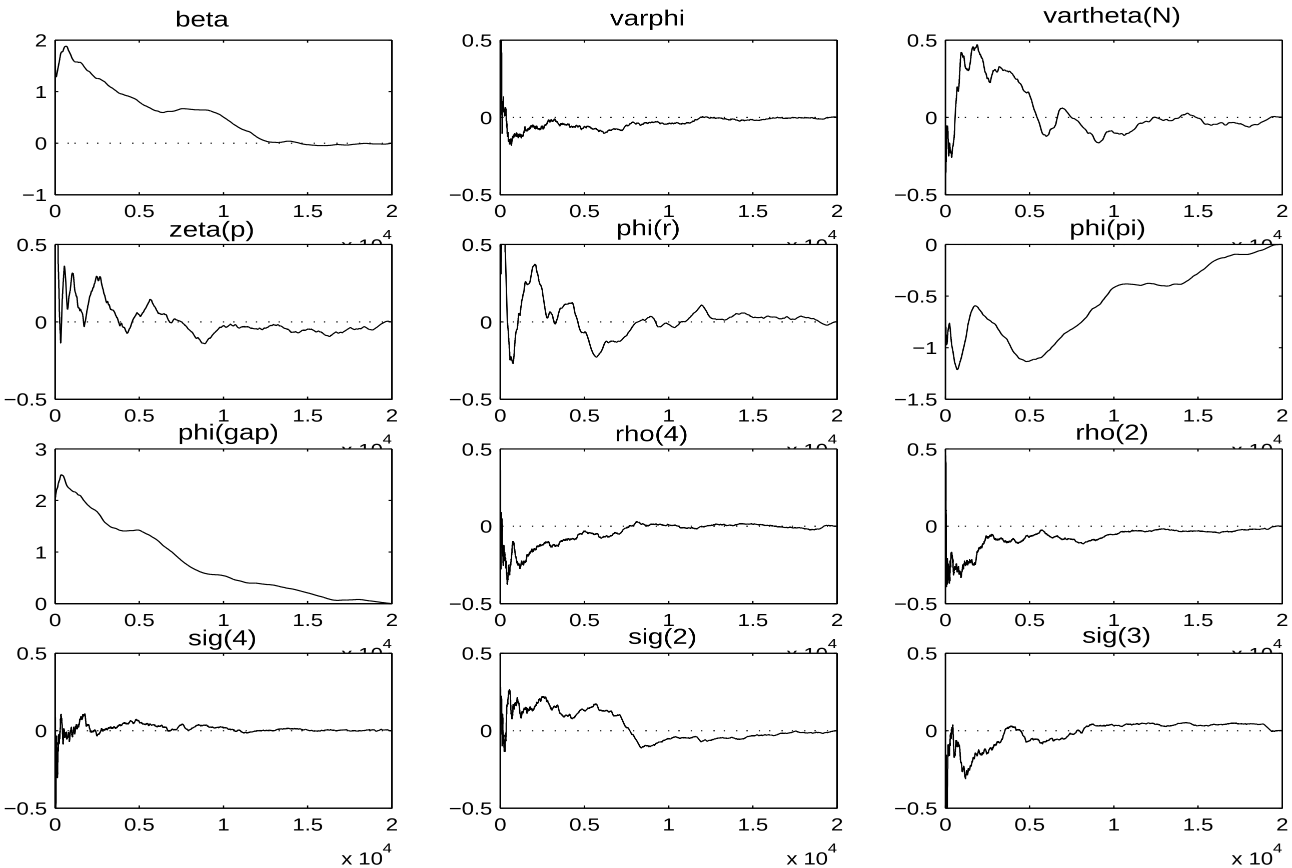


Figure 7: CUMSUM statistics

Prior and Posterior statistics

	Prior		Posterior				
	mean	*std*	*median*	*mean*	*std*	*max*	*min*
β	0.98	0.01	0.992	0.991	0.003	0.999	0.998
φ	1.00	0.37	0.826	0.843	0.123	1.262	0.425
ϑ_N	2.00	0.75	1.825	1.884	0.768	3.992	0.145
ζ_p	0.75	0.12	0.743	0.696	0.195	0.997	0.141
ϕ_r	0.75	0.14	0.596	0.587	0.154	0.959	0.102
ϕ_π	1.50	0.10	1.367	1.511	0.323	2.33	1.042
ϕ_{gap}	0.5	0.05	0.514	0.505	0.032	0.588	0.411
ρ_g	0.85	0.07	0.856	0.854	0.036	0.946	0.748
ρ_u	0.85	0.07	0.851	0.851	0.038	0.943	0.754
σ_g	0.025	0.07	0.025	0.025	0.001	0.028	0.021
σ_v	0.025	0.07	0.07	0.07	0.006	0.083	0.051
σ_r	0.025	0.07	0.021	0.021	0.005	0.035	0.025

- Little information in the data for some parameters (prior and posterior overlap).

- For parameters of the policy rule: posteriors move and not more concentrated.

-Posterior distributions roughly symmetric except for ϕ_π and ζ_p (mean and median coincide).

-Posterior distribution of economic parameters reasonable (except φ).

- Posterior for the AR parameters has a high mean, but no pile up at one.

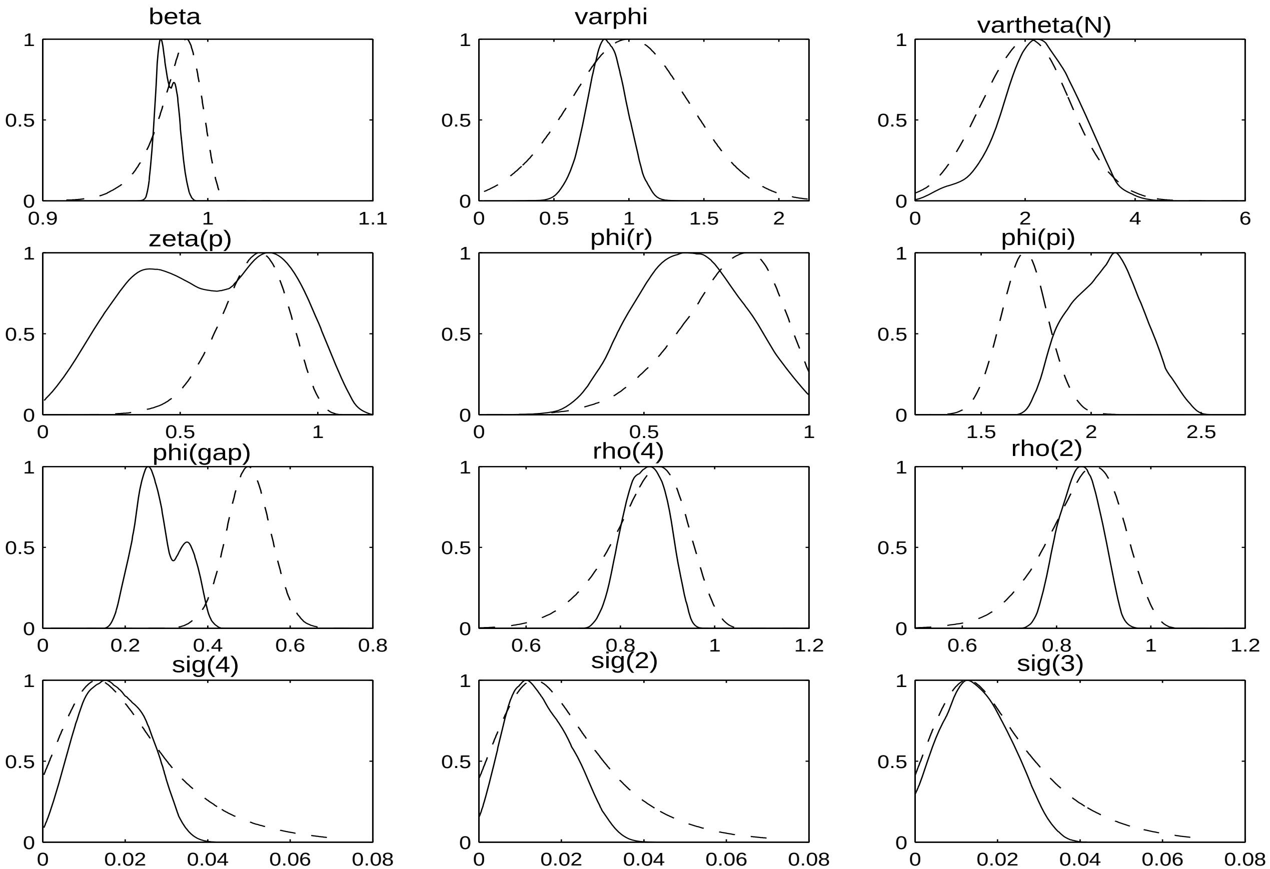


Figure 8: Priors and Posteriors, NK model

Model comparisons

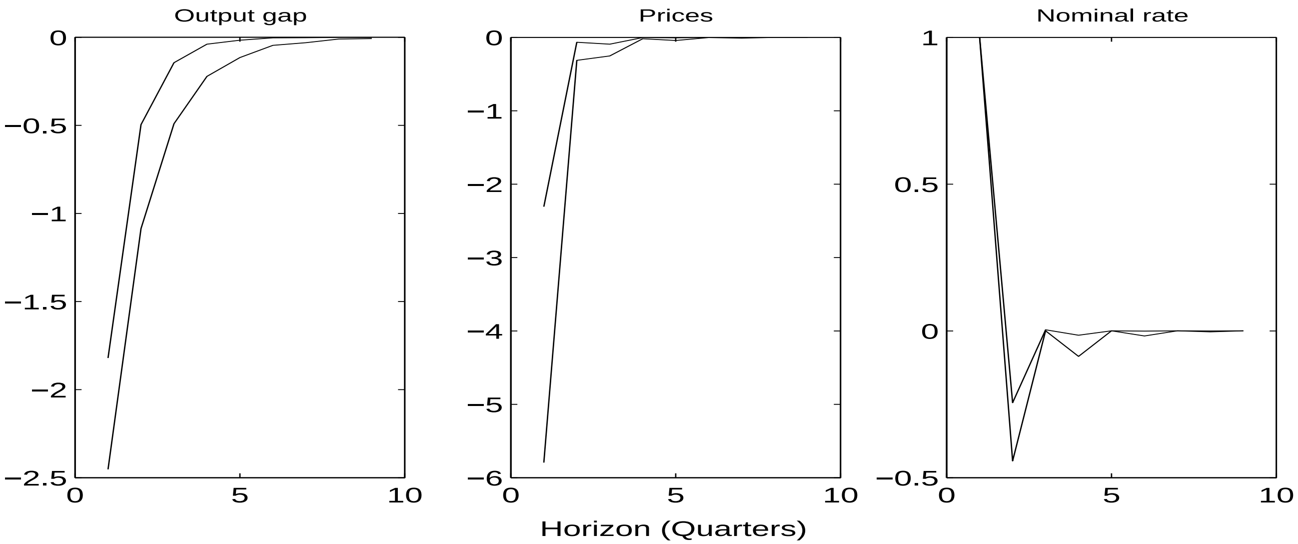
In-sample forecasting race against VAR(3) or a BVAR(3) with Minnesota prior and standard parameters (tightness=0.1, linear lag decay and weight on other variables equal 0.5), both with a constant.

Bayes factor are very small $\approx$ 0.02 in both cases.

Conclusion: The restrictions the model imposes are false. Need to add features that make dynamics of the model more similar to those of a VAR(3).

Posterior analysis

How do responses to monetary shocks look like? No persistence!



How much of the output gap and inflation variance explained by monetary shocks? Almost all!!

5.1 Interpreting results

- Most of the shocks of DSGE models are non-structural (alike to measurement errors). Careful with interpretation and policy analyses with these models (see Chari et al. (2009)).

- A model where "measurement errors" explain a large portion of main macro variables is very suspicious (e.g. in Smets and Wouters (2003) markup shocks dominate).

- If the standard error of one the shocks is large relative to the others: evidence of misspecification.

- Compare estimates with standard calibrated values. Are they sensible? Often yes, but because of tight priors are centered at calibrated values.

5.2 Bayesian methods and identification

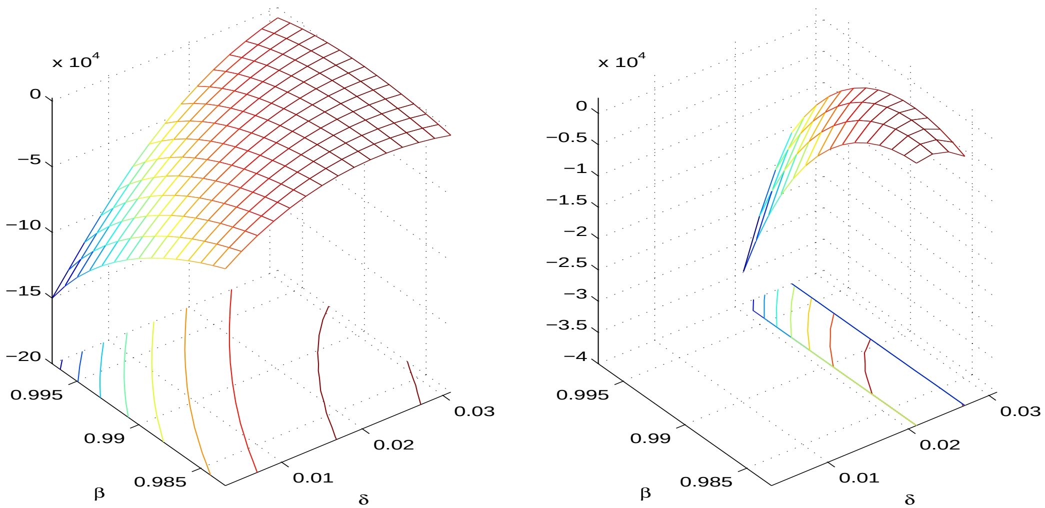
Likelihood of a DSGE typically flat. Could be due to marginalization (use only a subset of economic relationships), or to lack of information. Difficult to say a-priori which parameters is underidentified and which is not.

Standard remedy: fix some (non-identifiable) parameters. Problem if parameter not fixed at a consistent estimator $\rightarrow$ biases!

Alternative: add a prior. This increases the curvature of the likelihood $\rightarrow$ underidentification may be hidden!.

In general if $\mathcal{L}(\theta_1, \theta_2|Y^T) = \bar{\mathcal{L}}(\theta_1|Y^T)$ then $g(\theta_1, \theta_2|Y^T) = g_1(\theta_1|Y^T)$ $g(\theta_2|\theta_1)$, i.e no updating of conditional prior of θ_2.

However, updating possible even if no information is present if θ_1, θ_2 are linked by economic or stability conditions!!



Likelihood and Posterior, δ and β in a RBC model

If prior $\approx$ posterior: weak identification or too much data based prior?

6 Topics

6.1 Eliciting Priors from existing information

- Prior distributions for DSGE parameters often arbitrary.

- Prior distribution for individual parameters assumed to be independent: the joint distribution may assign non-zero probability to " unreasonable" regions of the parameter space.

- Prior sometimes set having some statistics in mind (the prior mean is similar to the one obtained in calibration exercises).

- Same prior is used for the parameters of different models. Problem: same prior may generate very different dynamics in different models. Hard to compare the outputs.

- Del Negro and Schorfheide (2008): elicit prior consistent with some distribution for the actual data (or statistics of it) (see also Kadane et al. (1980)). Basic idea:

i) Let θ be a set of DSGE parameters. Let S_T be a set of statistics obtained in the data with T observations and σ_S be the standard deviation of these statistics (which can be computed using asymptotic distributions or small sample devices, such as bootstrap or MC methods).

ii) Let $S_N(\theta)$ be the same set of statistics which are measurable from the model once θ is selected using N observations. Then

$$S_T = S_N(\theta) + \eta \quad \eta \sim (0, \Sigma_{TN}) \tag{13}$$

where η is a set of measurement errors.

Note

i) in calibration exercises $\Sigma_{TN} = 0$ and S_T are averages of the data.

ii) in SMM: $\Sigma_{TN} = 0$ and S_T are moments of the data.

Then $L(S_N(\theta)|S_T) = p(S_T|S_N(\theta))$, where the latter is the conditional density in (13).

Given any other prior information $\pi(\theta)$ which is not based on S_T, the prior for θ would be

$$p(\theta|S_T) \propto L(S_N(\theta)|S_T)\pi(\theta) \tag{14}$$

- $dim(S_T) \geq dim(\theta)$: overidentification is possible.

- Even if Σ_{TN} is diagonal, $S_N(\theta)$ will induce correlation across θ_i.

-Information used to construct S_T should be **different** than information used to estimate the model. Could be data in a training sample or could be data from a different country or a different regime (see e.g. Canova and Pappa (2007)).

- Assume that η are normal why? Make life easy, Could also use other distributions, e.g. uniform.

- Need to choose what are the S_T: could be steady states, could be autocorrelation functions. What S_T is depends on where the parameters enters.

Example 6.1

$$\max_{(c_t, K_{t+1}, N_t)} \quad E_0 \sum_t \beta^t \frac{(c_t^{\vartheta}(1-N_t)^{1-\vartheta})^{1-\varphi}}{1-\varphi} \tag{15}$$

$$G_t + c_t + K_{t+1} = GDP_t + (1-\delta)K_t \tag{16}$$

$$\ln \zeta_t = \bar{\zeta} + \rho_z \ln \zeta_{t-1} + \epsilon_{1t} \quad \epsilon_{1t} \sim (0, \sigma_z^2) \tag{17}$$

$$\ln G_t = \bar{G} + \rho_g \ln G_{t-1} + \epsilon_{4t} \quad \epsilon_{4t} \sim (0, \sigma_g^2) \tag{18}$$

$$GDP_t = \zeta_t K_t^{1-\eta} N_t^{\eta} \tag{19}$$

K_0 are given, c_t is consumption, N_t is hours, K_t is the capital stock. Let G_t be financed with lump sum taxes and λ_t the Lagrangian on (16).

The FOC are ((23) and (24) equate factor prices and marginal products)

$$\lambda_t = \vartheta c_t^{\vartheta(1-\varphi)-1}(1-N_t)^{(1-\vartheta)(1-\varphi)} \tag{20}$$

$$\lambda_t \eta \zeta_t k_t^{1-\eta} N_t^{\eta-1} = -(1-\vartheta)c_t^{\vartheta(1-\varphi)}(1-N_t)^{(1-\vartheta)(1-\varphi)-1} \tag{21}$$

$$\lambda_t = E_t \beta \lambda_{t+1}[(1-\eta)\zeta_{t+1}K_{t+1}^{-\eta}N_{t+1}^{\eta} + (1-\delta)] \tag{22}$$

$$w_t = \eta \frac{GDP_t}{N_t} \tag{23}$$

$$r_t = (1-\eta)\frac{GDP_t}{K_t} \tag{24}$$

Using (20)-(21) we have:

$$-\frac{1-\vartheta}{\vartheta}\frac{c_t}{1-N_t} = \eta \frac{GDP_t}{N_t} \tag{25}$$

Log linearizing the equilibrium conditions

$$\hat{\lambda}_t - (\vartheta(1-\varphi)-1)\hat{c}_t + (1-\vartheta)(1-\varphi)\frac{N^{ss}}{1-N^{ss}}\hat{N}_t = 0 \quad (26)$$

$$\hat{\lambda}_{t+1} + \frac{(1-\eta)(GDP/K)^{ss}}{(1-\eta)(GDP/K)^{ss}+(1-\delta))}(\widehat{GDP}_{t+1} - \hat{K}_{t+1}) = \hat{\lambda}_t \quad (27)$$

$$\frac{1}{1-N^{ss}}\hat{N}_t + \hat{c}_t - \widehat{gdp}_t = 0 \quad (28)$$

$$\hat{w}_t - \widehat{GDP}_t + \hat{n}_t = 0 \quad (29)$$

$$\hat{r}_t - \widehat{GDP}_t + \hat{k}_t = 0 \quad (30)$$

$$\widehat{GDP}_t - \hat{\zeta}_t - (1-\eta)\hat{K}_t - \eta\hat{N}_t = 0 \quad (31)$$

$$(\frac{g}{GDP})^{ss}\hat{g}_t + (\frac{c}{GDP})^{ss}\hat{c}_t + (\frac{K}{GDP})^{ss}(\hat{K}_{t+1} - (1-\delta)\hat{K}_t) - \widehat{GDP}_t = 0 \quad (32)$$

(31) and (32) are the production function and resource constraint.

Four types of parameters appear in the log-linearized conditions:

i.) Technological parameters (η, δ).

ii) Preference parameters $(\beta, \varphi, \vartheta)$.

iii) Steady state parameters $(N^{ss}, (\frac{c}{GDP})^{ss}, (\frac{K}{GDP})^{ss}, (\frac{g}{GDP})^{ss})$.

iv) Parameters of the driving process $(\rho_g, \rho_z, \sigma_z^2, \sigma_g^2)$.

Question: How do we set a prior for these 13 parameters?

The steady state of the model (using (22)-(25)-(16)) is:

$$\frac{1-\vartheta}{\vartheta}\left(\frac{c}{GDP}\right)^{ss} = \eta\frac{1-N^{ss}}{N^{ss}} \quad (33)$$

$$\beta\left[(1-\eta)\left(\frac{GDP}{K}\right)^{ss} + (1-\delta)\right] = 1 \quad (34)$$

$$\left(\frac{g}{GDP}\right)^{ss} + \left(\frac{c}{GDP}\right)^{ss} + \delta\left(\frac{K}{GDP}\right)^{ss} = 1 \quad (35)$$

$$\frac{GDP}{wc} = \eta \quad (36)$$

$$\frac{K}{i} = \delta \quad (37)$$

Five equations in 8 parameters!! Need to choose.

For example: (33)-(37) determine $(N^{ss}, (\frac{c}{GDP})^{ss}, (\frac{K}{GDP})^{ss}, \eta, \delta)$ *given* $((\frac{g}{GDP})^{ss}, \beta, \vartheta)$.

Set $\theta_2 = [(\frac{g}{GDP})^{ss}, \beta, \vartheta]$ *and* $\theta_1 = [N^{ss}, (\frac{c}{GDP})^{ss}, (\frac{K}{GDP})^{ss}, \eta, \delta]$

Then if S_{1T} *are steady state relationships, we an use (33)-(37) to construct a prior distribution for* $\theta_1|\theta_2$.

How do we measure uncertainty in S_{1T}*?*

- Take a rolling window to estimate S_{1T} *and use uncertainty of the estimate to calibrate var*(η).

- Bootstrap S_{1T}*, etc.*

How do we set θ_2? Use additional information (statistics)!

- $(\frac{g}{GDP})^{ss}$ *should be the average G/Y in the data.*

- $\beta = (1 + r)^{-1}$ *and typically* $r^{ss} = [0.0075, 0.0150]$ *per quarter.*

- ϑ *is related to Frish elasticity of labor supply: use estimates of labor supply elasticity.*

If S_{2T} are these statistics, possible to back out a prior for θ_2 consistent with available information. Uncertainty could be data based or across studies (meta uncertainty).

Parameters of the driving process $(\rho_g, \rho_z, \sigma_z^2, \sigma_g^2)$ *do not enter the steady state. Call them* θ_3*. How do we choose a prior for them?*

- ρ_z, σ_z^2 *can be backed out from moments of Solow residual i.e. estimate the variance and the AR(1) of* $\hat{z} = \ln GDP_t - (1-\eta)K_t - \eta N_t$*, once* η *is chosen. Prior for* η *induce a distribution for* $\hat{z}$

- ρ_g, σ_g^2 *backed out from moments government expenditure data.*

- For φ (coefficient of relative risk aversion (RRA) is $1 - \vartheta(1 - \varphi)$) one has two options:

(a) appeal to existing estimates of RRA. Construct a prior which is consistent with the cross section of estimates (e.g. a $\chi^2(2)$ would be ok).

(b) select an interesting moment, say $var(c_t)$ then

$$var(c_t) = var(c_t(\varphi)|\theta_1, \theta_2, \theta_3) + \eta \tag{38}$$

For some parameters (call them θ_5) we have no moments to match but may have some micro evidence. Then $p(\theta_5) = \pi(\theta_5)$ could be estimated from the histogram of the estimates which are available.

In sum, the prior for the parameters is

$$\begin{aligned} p(\theta) &= p(\theta_1|S_{1T})p(\theta_2|S_{2T})p(\theta_3|S_{3T})p(\theta_4|S_{4T}) \\ &\quad \pi(\theta_1)\pi(\theta_2)\pi(\theta_3)\pi(\theta_4)\Pi(\theta_5) \end{aligned} \tag{39}$$

- If we had used a different utility function, the prior e.g. for θ_1, θ_4 would be different. **Prior for different models/parameterizations should be different**.

- To use the priors we have created, need a normalizing constant in (14). Need a RW metropolis to draw from the priors we have produced.

- Careful about multidimensional ridges: e.g. steady states are 5 equations, and there are 8 parameters - solution not unique, impossible to invert the relationship.

- Careful about choosing θ_3 and θ_4 when there are weak and partial identification problems.

6.2 Choice of data and estimation

- Does it matter which variables are used to estimate the parameters? Yes.

i) Omitting relevant variables may lead to distortions.

ii) Adding variables may improve the fit, but also increase standard errors if added variables are irrelevant.

iii) Different variables may identify different parameters (e.g. with aggregate consumption data and no data on who own financial assets may be very difficult to get estimate the share of rule-of-thumb consumers).

Example 6.2

$$y_t = a_1 E_t y_{t+1} + a_2(i_t - E_t \pi_{t+1}) + v_{1t} \quad (40)$$

$$\pi_t = a_3 E_t \pi_{t+1} + a_4 y_t + v_{2t} \quad (41)$$

$$i_t = a_5 E_t \pi_{t+1} + v_{3t} \quad (42)$$

Solution:

$$\begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_2 \\ a_4 & 1 & a_2 a_4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix}$$

- a_1, a_3, a_5 *disappear from the solution.*

- *Different variables identify different parameters (*i_t *identify nothing!!)*

iv) Likelihood function (Posterior) may change shape depending on the variables. Bimodality or multimodality may be present if important variables are omitted (e.g. if y_t is excluded in above example).

- Using the same model and the same econometric approach Levin et al (2005, NBER macro annual) find habit in consumption is 0.30; Fernandez and Rubio (2008, NBER macro annual) find habit in consumption is 0.88. Why? They use different data sets to estimate the same model!

Can we say something systematic? Difficult.

Guerron-Quintana (2010); use Smets and Wouters model and different combinations of observable variables. Finds:

- Internal persistence of the model change if nominal rate, inflation and real wage are absent.

- Duration of price spells affected by the omission of consumption and real wage data.

- Responses of inflation, investment, hours and real wage sensitive to the choice of variables.

- " Best combination" of variables (use in-sample prediction and out-of-sample MSE): use $Y_t, C_t, I_t, R_t, H_t, \pi_t, W$.

Parameter	Wage stickiness	Price Stickiness	Slope Phillips
Data	Median (s.d.)	Median (s.d.)	Median (s.d.)
Basic	0.62 (0.54,0.69)	0.82 (0.80, 0.85)	0.94 (0.64,1.44)
Without C	0.80 (0.73,0.85)	0.97 (0.96, 0.98)	2.70 (1.93,3.78)
Without Y	0.34 (0.28,0.53)	0.85 (0.84, 0.87)	6.22 (5.05,7.44)
Without C,W	0.57 (0.46,0.68)	0.71 (0.63, 0.78)	2.91 (1.73,4.49)
Without R	0.73 (0.67,0.78)	0.81 (0.77, 0.84)	0.74 (0.53,1.03)

(in parenthesis 90% probability intervals)

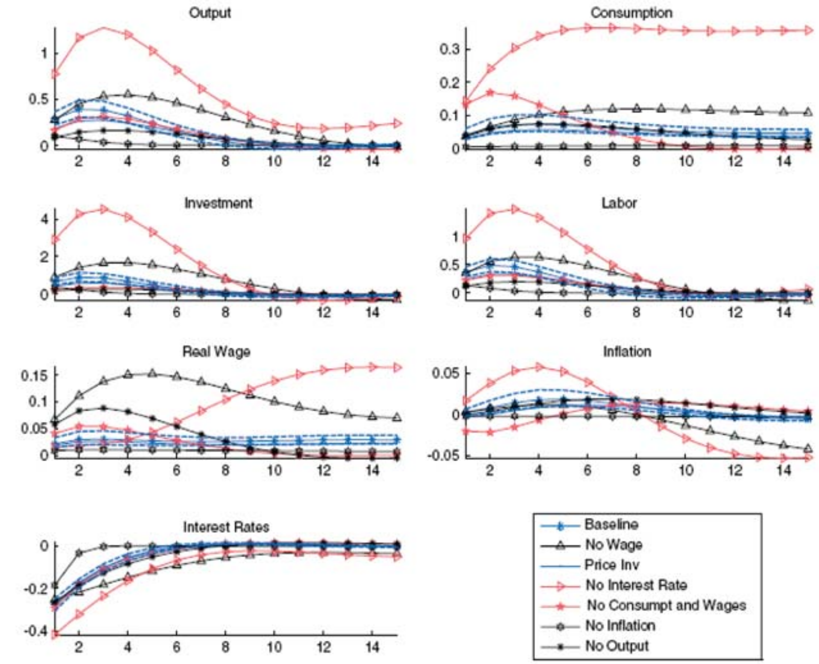


Figure 1. Responses to an expansionary monetary shock. This figure is available in color online at www.interscience.wiley.com/journal/jae

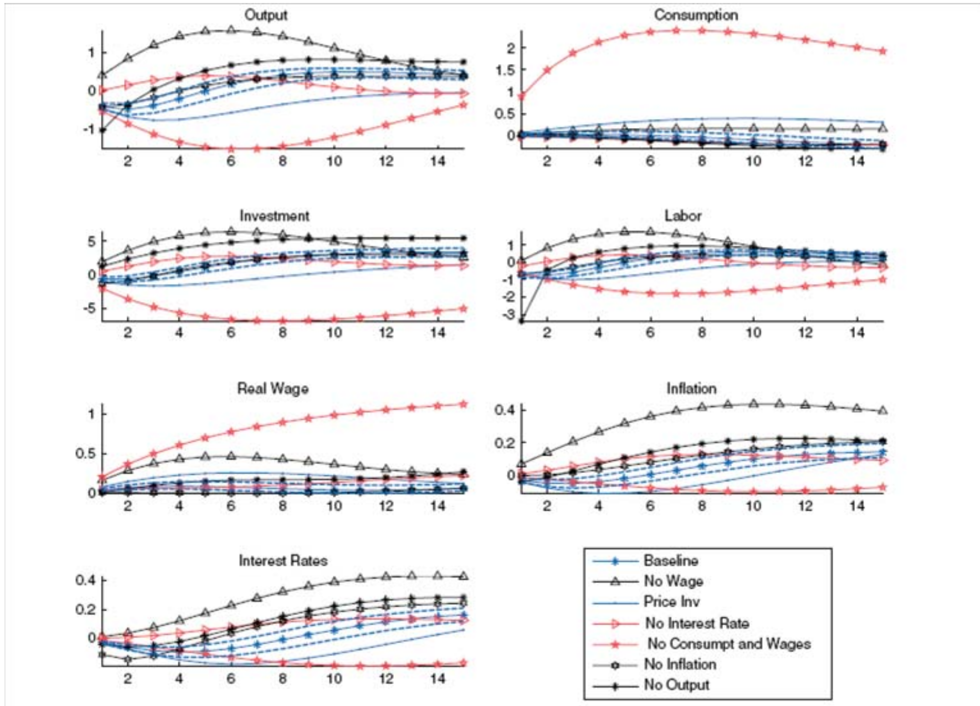


Figure 2. Responses to a positive investment-specific technology shock. This figure is available in color online at www.interscience.wiley.com/journal/jae

Output recession after an investments specific shock and no C and W.

6.3 DSGE and VARs

Del Negro and Schorfheide(2004):

- $f(y|\alpha, \Sigma_u) =$ likelihood of the data conditional on the VAR parameters (represent the data with a VAR)

- $g(\alpha, \Sigma_u|\theta)$ prior for the VAR parameters, conditional on the DSGE model parameters (the hyperparameters)

- $g(\theta)$ the prior distribution for DSGE parameters $\rightarrow g(\alpha, \Sigma_u|\theta)$ is the prior on the reduced form parameters induced by the prior on the structural parameters and the structure of the DGSE model.

Joint posterior of VAR and structural parameters is

$g(\alpha, \Sigma_u, \theta|y) = g(\alpha, \Sigma_u, |\theta, y)g(\theta|y)$.

$g(\alpha, \Sigma_u, |\theta, y)$ is of normal-inverted Wishart form: easy to compute.

Posterior kernel $\breve{g}(\theta|y) = f(y|\theta)g(\theta)$ where $f(y|\theta)$ is given by

$$
\begin{aligned}
f(y|\theta) &= \int f(y|\alpha, \Sigma_u)g(\alpha, \Sigma_u, \theta)d\alpha d\theta \\
&= \frac{f(y|\alpha, \Sigma_u)g(\alpha, \Sigma_u|\theta)}{g(\alpha, \Sigma_u|y)}
\end{aligned}
$$

Given that $g(\alpha, \Sigma_u, |\theta, y) = g(\alpha, \Sigma_u, |y)$. Then

$$f(y|\theta) = \frac{|T_1 x^{s'}(\theta) x^s(\theta) + X'X|^{-0.5M} |(T_1+T)\tilde{\Sigma}_u(\theta)|^{-0.5(T_1+T-k)}}{|\tau x^{s'}(\theta) x^s(\theta)|^{-0.5M} |T_1 \tilde{\Sigma}_u^s(\theta)|^{-0.5(T_1-k)}} \times \frac{(2\pi)^{-0.5MT} 2^{-0.5M(T_1+T-k)} \prod_{i=1}^{M} \Gamma(0.5*(T_1+T-k+1-i))}{2^{-0.5M(T_1-k)} \prod_{i=1}^{M} \Gamma(0.5*(T_1-k+1-i))} \quad (43)$$

T_1 = number of simulated observations, Γ is the Gamma function, X includes all lags of y and the superscript s indicates simulated data.

- Draw θ using an MH algorithm.

- Conditional on θ construct posterior of α (draw α from a Normal-Wishart).

6.4 Practical issues

Log-linear DSGE solution:

$$y_{1t} = \mathcal{A}_{11}(\theta)y_{1t-1} + \mathcal{A}_{13}(\theta)y_{3t} \quad (44)$$
$$y_{2t} = \mathcal{A}_{12}(\theta)y_{1t-1} + \mathcal{A}_{23}(\theta)y_{3t} \quad (45)$$

where y_{2t} are the control, y_{1t} the states (predetermined and exogeneous), y_{3t} the shocks, θ are the structural parameters and $\mathcal{A}_{ij}$ the coefficients of the decision rules.

How to you put DSGE models on the data when:

a) the model implies that the covariance of $y_t = [y_{1t}, y_{2t}]$ is singular.

b) the variables are mismeasured relative to the model quantities.

c) have additional information one would like to use.

For a):

• Choose a selection matrix F_1 such that dim(x_{1t}) = dim $(F_1\ y_t)$ = dim (y_{3t}), i.e. throw away model information. Good strategy to follow if some component of y_t are non-observable.

• Explicitly solve out fraction of variables from the model. Format of the solution is no longer a restricted VAR(1).

• Adds measurement errors to the y_{2t} so that dim(x_{2t}) = dim $(F_2\ y_t)$ = dim (y_{3t})+dim (e_t), where e_t are measurement errors.

- If the model has two shocks and implications for four variables, we could add at least two and up to four measurement errors to the model.

Here (1)-(2) are the state equations and the measurement equation is

$$x_{2t} = F_2 y_t + e_t \tag{46}$$

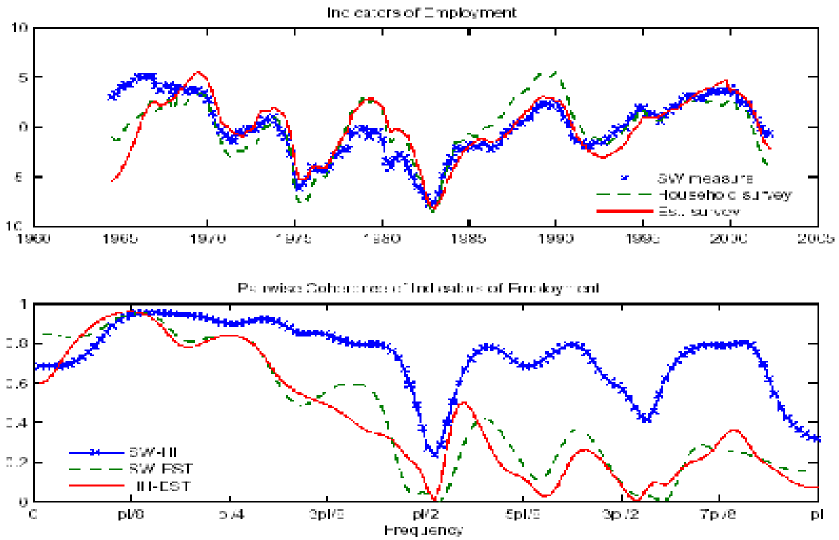
- Need to restrict time series properties of e_t. Otherwise difficult to distinguish dynamics induced by structural shocks and the measurement errors.

i) the measurement error is iid (since θ is identified from the dynamics induced by the reduced form shocks, if measurement error is iid, θ identified by the dynamics due to structural shocks).

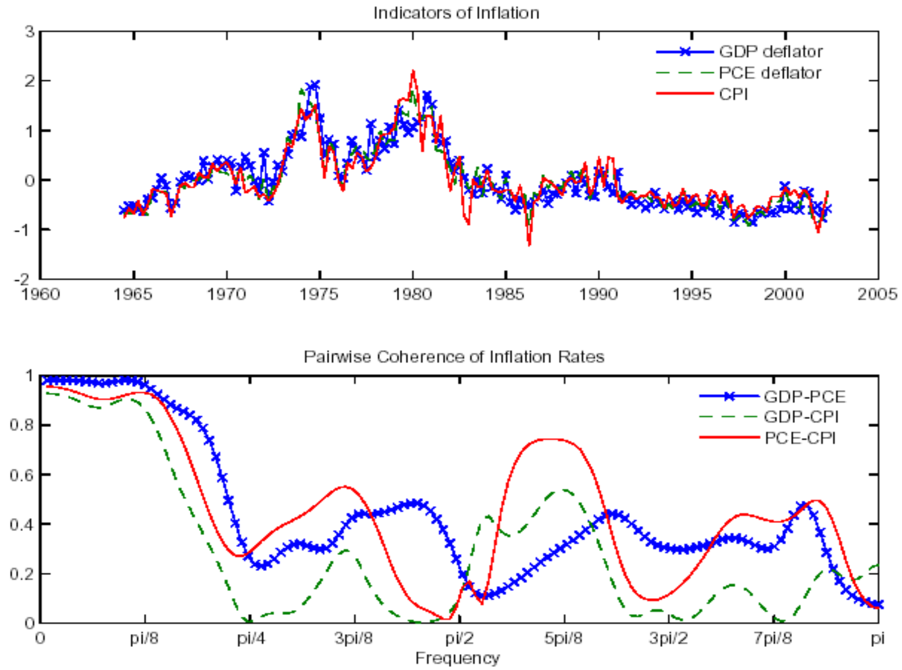
Ireland (2004): VAR(1) process for the measurement error; identification problems! Can be used to verify the quality of the model's approximation to the data - measurement error captures what is missing from the model to fit the data (see also Watson (1993)). Useful device when θ is calibrated. Less useful when θ is estimated.

For b): Recognize that existing measures of theoretical concepts are contaminated.

- How do you measure hours? Use establishment survey series? Household survey series? Employment?



- How do we measure inflation. Do we use CPI, GDP deflator or PCE inflation?



- In principle measures contain information about the true series but not perfectly correlated among each other.

- Use ideas underlying factor models. Let x_{3t} be a $k \times 1$ vector of observable variables and x_{1t} be of dimension $N \times 1$ where dim (N) ¡ dim(k). Then:

$$x_{3t} = \Lambda_3 x_{1t} + u_t \tag{47}$$

where the first row of Λ_3 is normalized to 1. Thus:

$$\begin{aligned} x_{3t} &= \Lambda_3 [F_1 y_{1t}, F_1 \mathcal{A}_{12}(\theta) y_{1t-1} + F_1 \mathcal{A}_{13}(\theta) y_{3t}]' + u_{3t} && (48) \\ &= \Lambda_3 [F_1 y_{1t}, F_1 \mathcal{B}(\theta) y_{1t}]' + u_{3t} && (49) \end{aligned}$$

so that x_{3t} can be used to recover the vector of states y_{1t} and to estimate θ

- What is the advantage of this procedure? If only one component of x_{3t} is used to measure y_{1t}, estimate of θ will probably be noisy.

- By using a vector of information and assuming that the elements of u_t are idiosyncratic:

i) reduce the noise in the estimate of y_{1t} (the estimated variance of y_{1t} will be asymptotically of the order $1/k$ time the variance obtained when only one indicator is used (see Stock and Watson (2002)).

ii) estimates of θ more precise.

- How different is from factor models?. The DSGE model structure is imposed in the specification of the law of motion of the states (states have economic content). In factor models the states are assumed to follow is an assumed unrestricted time series specification, say an AR(1) or a random walk and are uninterpretable.

- How do we separately identify the dynamics induced by the structural shocks and the measurement errors? Since the measurement error is identified from the cross sectional properties of the variables in x_{3t}, possible to have structural disturbances and measurement errors to both be serially correlated of an unknown form.

For c): Sometimes we may have proxy measures for the unobservable states. (commodity prices are often used as proxies for future inflation shocks, stock market shocks are used as proxies for future technology shocks (Beaudry and Portier (2006)).

- Can use these measures to get information about the states. Let x_{4t} a $q \times 1$ vector of variables. Assume

$$x_{4t} = \Lambda_4 y_{1t} + u_{4t} \tag{50}$$

where Λ_4 is unrestricted. Combining all sources of information we have

$$X_t = \Lambda y_{1t} + u_t \tag{51}$$

where $X_t = [x_{3t}, x_{4t}]'$, $u_t = [u_{3t}, u_{4t}]$ and $\Lambda = [\Lambda_3 F, \Lambda_3 F \mathcal{B}(\theta), \Lambda_4]'$.

- The fact that we are using the DSGE structure ($\mathcal{B}$ depends on θ) imposes restrictions on the way the data behaves. (interpret data information through the lenses of the DSGE model).

- Can still jointly estimate the structural parameters and the unobservable states of the economy.

6.5 An example

Use a simple three equation New-keynesian model:

$$x_t = E_t(x_{t+1}) - \frac{1}{\phi}(i_t - E_t\pi_{t+1}) + e_{1t} \quad (52)$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t + e_{2t} \quad (53)$$

$$i_t = \psi_r i_{t-1} + (1-\psi_r)(\psi_\pi \pi_t + \psi_x x_t) + e_{3t} \quad (54)$$

where β is the discount factor, ϕ the relative risk aversion coefficient, κ the slope of Phillips curve, $(\psi_r, \psi_\pi, \psi_x)$ policy parameters. Here x_t is the output gap, π_t the inflation rate and i_t the nominal interest rate. Assume

$$e_{1t} = \rho_1 e_{1t-1} + v_{1t} \quad (55)$$

$$e_{2t} = \rho_2 e_{2t-1} + v_{2t} \quad (56)$$

$$e_{3t} = v_{3t} \quad (57)$$

where $\rho_1, \rho_2 < 1$, $v_{jt} \sim (0, \sigma_j^2), j = 1, 2, 3$.

6.5.1 Contaminated data

- Ambiguities in linking the output gap, the inflation rate and the nominal interest rate to empirical counterparts. e.g. for the nominal interest rate: should we use a short term measure or a long term one? for the output gap, should we use a statistical based measure or a theory based measure? In the last case, what is the flexible price equilibrium?

The solution of the model can be written as

$$w_t = RR(\theta)w_{t-1} + SS(\theta)v_t \tag{58}$$

where w_t is a 8×1 vector including x_t, π_t, i_t, the three shocks and the expectations of x_t and π_t and $\theta = (\phi, \kappa, \psi_r, \psi_y, \psi_\pi, \rho_1, \rho_2, \sigma_1, \sigma_2, \sigma_3)$.

Let $x_t^j, j = 1, \ldots N_x$ be observable indicators for x_t, let $\pi_t^j, j = 1, \ldots N_\pi$ observable indicators for π_t, and $i_t^j, j = 1, \ldots N_i$ observable indicators for i_t. Let $W_t = [x_t^1, \ldots, x_t^{N_x}, \pi_t^1, \ldots, \pi_t^{N_\pi}, i_t^1, \ldots i_t^{N_i}]'$ be a $N_x + N_\pi + N_i \times 1$ vector.

Assume that (58) is the state equation of the system and that the measurement equation is

$$W_t = \Lambda w_t + e_t \tag{59}$$

where λ is $N_x + N_\pi + N_i \times 3$ matrix with at most one element different from zero in each row.

- Once we normalize the nonzero element of the first row of Λ to be one, we can estimate (58)-(59) with standard methods. The routines give us estimates of λ, RR, SS and of w_t which are consistent with the data.

6.5.2 Conjunctoral information

- Suppose we have available measures of future inflation (from surveys, from forecasting models) or data which may have some information about future inflation, for example, oil prices, housing prices, etc.

- Want to predict inflation h periods ahead, $h = 1, 2, \ldots$.

Let $\pi_t^j, j = 1, \ldots N_\pi$ be the observable indicators for π_t and let $W_t = [x_t, i_t, \pi_t^1, \ldots, \pi_t^{N_\pi}]'$ be a $2 + N_\pi \times 1$ vector.

The measurement equation is:

$$W_t = \Lambda w_t + e_t \tag{60}$$

where Λ is $2 + N_\pi \times 3$ matrix, $\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_1 \\ \ldots & \ldots & \ldots \\ 0 & 0 & \lambda_{N_\pi} \end{bmatrix}$.

- Estimates of the unobservable w_t can be obtained with the Kalman filter. Using estimates of $RR(\theta)$ and $SS(\theta)$ from the state equation we can unconditionally predict w_t h-steps ahead or predict its path conditional on a path for $v_{l,t+h}$.

- Forecast will incorporate information from the model, information from conjunctural data and from standard data and information about the path of the shocks. This information will be optimally mixed depending on their relative precision.

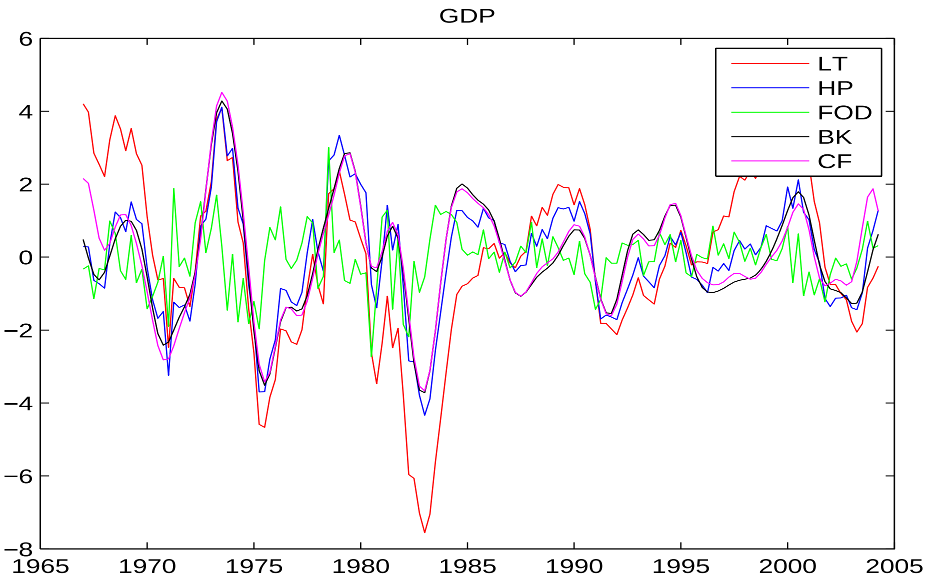
6.6 Dealing with trends

- Most of models available for policy are stationary and cyclical.

- Data is close to non-stationary, has trends and displays breaks.

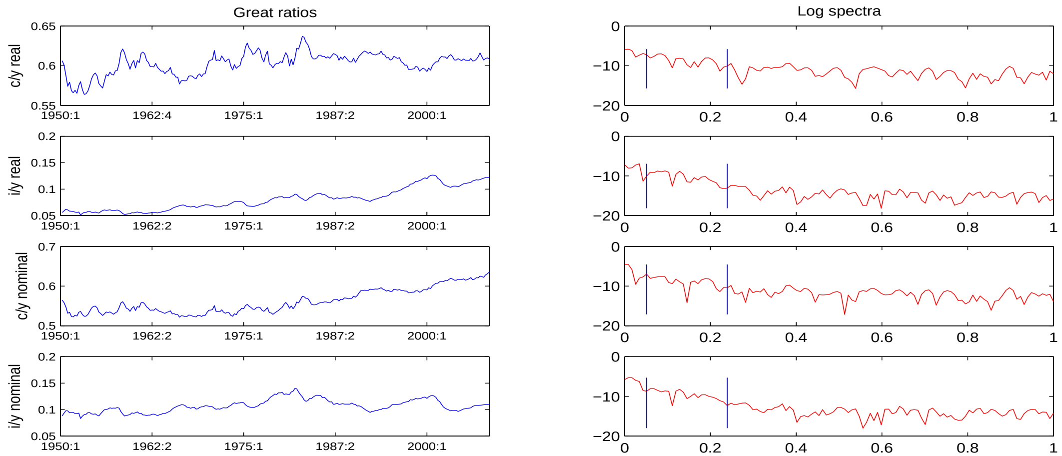
- How to we match models to the data?

a) Detrend actual data. Model is a representation for detrended data standard approach. Problem: which detrended data is the model representing?



GDP
6
4
2
0
−2
−4
−6
−8
1965
1970
1975
1980
1985
1990
1995
2000
2005
LT
HP
FOD
BK
CF

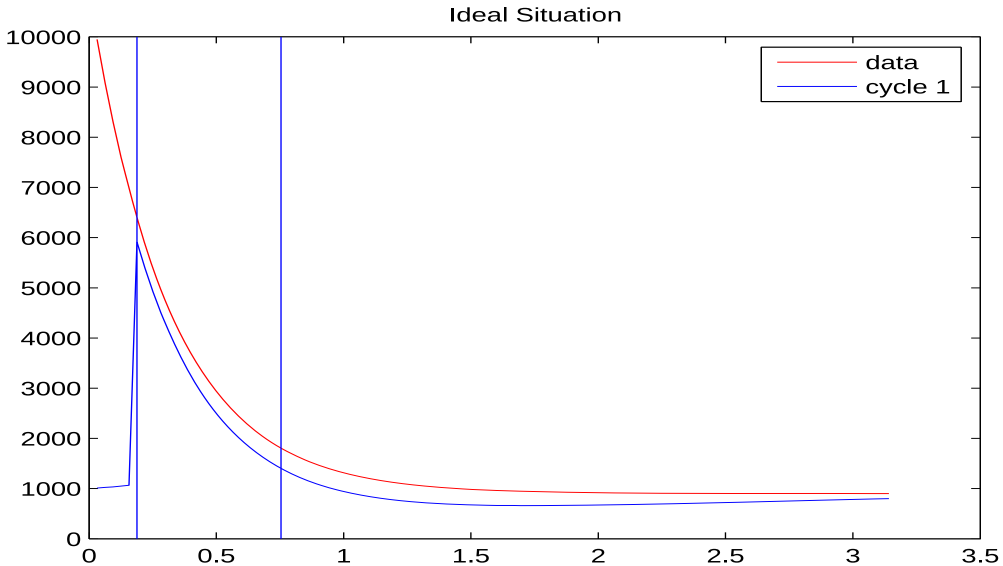
b) Build-in a trend into the model. Detrend the data with model-based-trend. Problem: data does not seem to satisfy balanced growth.



Real and nominal Great ratios in US, 1950-2008.

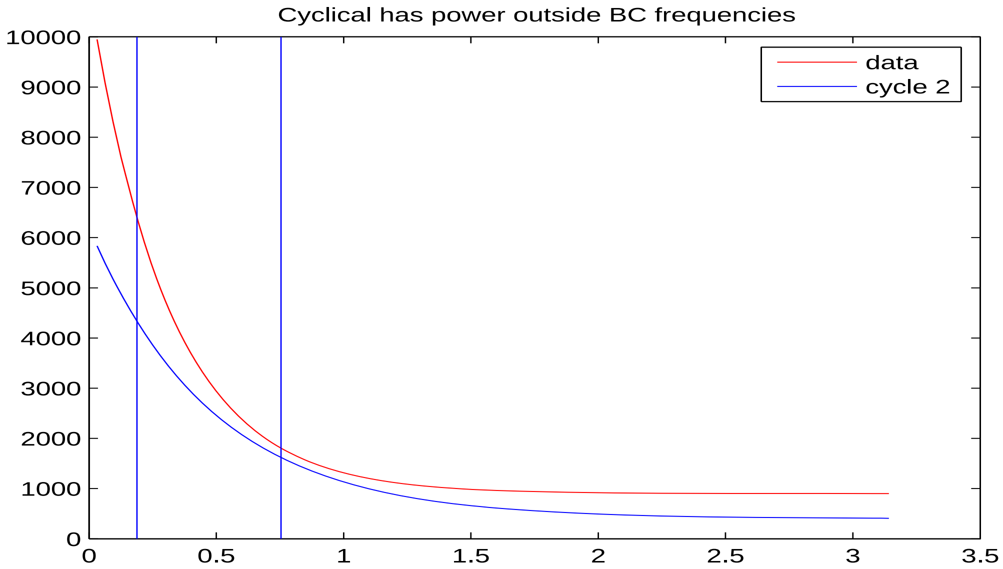
c) Use transformation of the data which allow you to estimate jointly cycle and the parameters trend (see e.g. growth rates in Smets and Wouter 2007). Problem: hard to fit models to quarterly growth rates

- General problem: statistical definition of cycles different than economic definition. All statistical approaches are biased even in large samples.

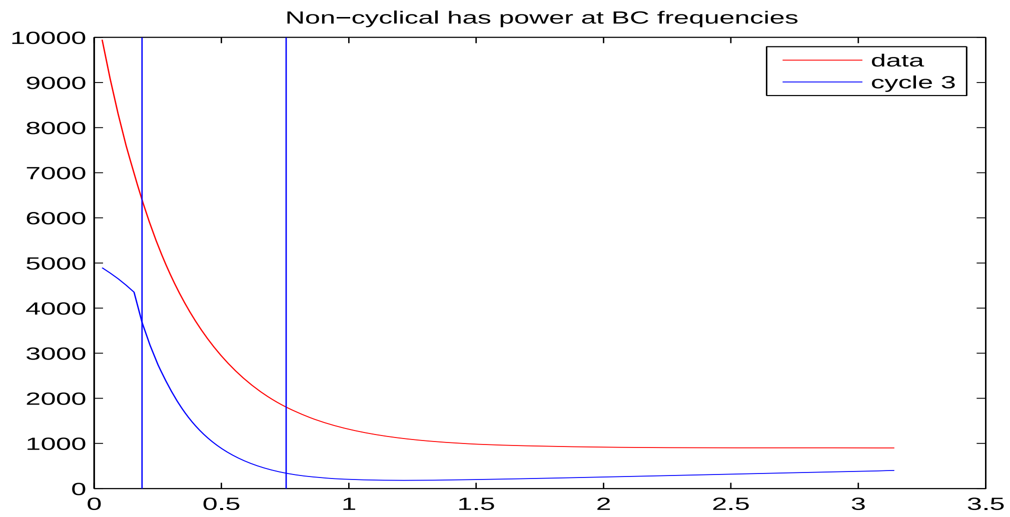


Ideal Situation
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
0.5
1
1.5
2
2.5
3
3.5
data
cycle 1

Ideal case



Realistic case



General case

- In developing countries most of cyclical fluctuations driven by trends (Aguiar and Gopinath (2007)).

Two potential approaches:

1) Data-rich envirornment (Canova and Ferroni (2011)). Let y_t^i be the actual data filtered with method $i = 1, 2, ..., I$ and $y_t^d = [y_t^1, y_t^2, \ldots]$. Assume:

$$y_t^d = \lambda_0 + \lambda_1 y_t(\theta) + u_t \tag{61}$$

where $\lambda_j, j = 0, 1$ are matrices of parameters, measuring bias and correlation between data and model based quantities, u_t are measurement errors and θ the structural parameters.

- Factor model setup a-la Boivin and Giannoni (2005).

- Can jointly estimate θ and λ's. Can obtain a more precise estimate of the unobserved $y_t(\theta)$ if measurement error is uncorrelated across methods.

- Same interpretation as GMM with many instruments.

2) Bridge cyclical model and the data with a flexible specification for the trend (Canova, 2010)).

$$y_t^d = c + y_t^T + y_t^m(\theta) + u_t \qquad (62)$$

where $y_t^d \equiv \tilde{y}_t^d - E(\tilde{y}_t^d)$ the log demeaned vector of observables, $c = \bar{y} - E(\tilde{y}_t^d)$, y_t^T is the non-cyclical component, $y_t^m(\theta) \equiv S[y_t, x_t]'$, S is a selection matrix, is the model based- cyclical component, u_t is a iid $(0, \Sigma_u)$ (measurement) noise, $y_t^T, y_t^m(\theta)$ and u_t are mutually orthogonal.

- Model (linearized) solution: cyclical component

$$\begin{aligned} y_t &= RR(\theta)x_{t-1} + SS(\theta)z_t & (63)\\ x_t &= PP(\theta)x_{t-1} + QQ(\theta)z_t & (64)\\ z_{t+1} &= NN(\theta)z_t + \epsilon_{t+1} & (65) \end{aligned}$$

$PP(\theta), QQ(\theta), RR(\theta), SS(\theta)$ functions of the structural parameters $\theta = (\theta_1, \ldots, \theta_k)$, $x_t = \tilde{x}_t - \bar{x}$; $y_t = \tilde{y}_t - \bar{y}$; and z_t are the disturbances, $\bar{y}, \bar{x}$ are the steady states of $\tilde{y}_t$ and $\tilde{x}_t$.

- Non cyclical component

$$y_t^T = y_{t-1}^T + \bar{y}_{t-1} + e_t \qquad e_t \sim \ iid\ (0, \Sigma_e^2) \tag{66}$$
$$\bar{y}_t = \bar{y}_{t-1} + v_t \qquad v_t \sim \ iid\ (0, \Sigma_v^2) \tag{67}$$

$\Sigma_v^2 > 0$ and $\Sigma_e^2 = 0$, y_t^T is a vector of I(2) processes.

$\Sigma_v^2 = 0$, and $\Sigma_e^2 > 0$, y_t^T is a vector of I(1) processes.

$\Sigma_v^2 = \Sigma_e^2 = 0$, y_t^T is deterministic.

$\Sigma_v^2 > 0$ and $\Sigma_e^2 > 0$ and $\sigma_v^2 \sigma_e^2$ is large, y_t^t is "smooth" and nonlinear (as in HP).

- Jointly estimate structural θ and non-structural parameters.

Example 6.3 *The log linearized equilibrium conditions of basic NK model are:*

$$\lambda_t = \chi_t - \frac{\sigma_c}{1-h}(y_t - hy_{t-1}) \tag{68}$$

$$y_t = z_t + (1-\alpha)n_t \tag{69}$$

$$w_t = -\lambda_t + \sigma_n n_t \tag{70}$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)(\rho_\pi \pi_t + \rho_y y_t) + v_t \tag{71}$$

$$\lambda_t = E_t(\lambda_{t+1} + r_t - \pi_{t+1}) \tag{72}$$

$$\pi_t = k_p(w_t + n_t - y_t + \mu_t) + \beta E_t \pi_{t+1} \tag{73}$$

$$z_t = \rho_z z_{t-1} + \iota_t^z \tag{74}$$

where $k_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}\frac{1-\alpha}{1-\alpha+\varepsilon\alpha}$, λ *is the Lagrangian on the consumer budget constraint,* z_t *is a technology shock,* χ_t *a preference shock,* v_t *is an iid monetary policy shock and* ϵ_t *an iid markup shock.*

Filter		LT	HP	FOD	BP	Flexible
Parameter	True	Median (s.d.)	Median (s.d.)	Median (s.d.)	Median(s.d.)	Median(s.d.)
σ_c	3.00	2.08 (0.11)	2.08 (0.14)	1.89 (0.14)	2.13 (0.12)	3.68(0.40)
σ_n	0.70	1.72 (0.09)	1.36 (0.07)	1.24 (0.06)	1.58 (0.08)	0.54(0.14)
h	0.70	0.67 (0.02)	0.58 (0.03)	0.36 (0.03)	0.66 (0.02)	0.55(0.04)
α	0.60	0.28 (0.03)	0.15 (0.02)	0.14 (0.02)	0.17 (0.02)	0.19(0.03)
ϵ	7.00	3.19 (0.11)	5.13 (0.19)	3.76 (0.18)	3.80 (0.13)	6.19(0.07)
ρ_r	0.20	0.54 (0.03)	0.77 (0.03)	0.72 (0.04)	0.53 (0.03)	0.16(0.04)
ρ_π	1.20	1.69 (0.08)	1.65 (0.06)	1.65 (0.07)	1.63 (0.10)	0.30(0.04)
ρ_y	0.05	-0.14 (0.04)	0.45 (0.04)	0.63 (0.06)	0.40 (0.04)	0.07(0.03)
ζ_p	0.80	0.85 (0.03)	0.91 (0.03)	0.93 (0.03)	0.90 (0.03)	0.78(0.04)
ρ_χ	0.50	1.00 (0.03)	0.96 (0.03)	0.96 (0.03)	0.95 (0.03)	0.53(0.02)
ρ_z	0.80	0.84 (0.03)	0.96 (0.03)	0.97 (0.03)	0.96 (0.03)	0.71(0.03)
σ_χ	1.12	0.11 (0.02)	0.17 (0.02)	0.21 (0.03)	0.14 (0.02)	1.29(0.01)
σ_z	0.51	0.07 (0.01)	0.09 (0.01)	0.09 (0.01)	0.07 (0.01)	0.72(0.02)
σ_{mp}	0.10	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.22(0.004)
σ_μ	20.60	6.30 (0.50)	16.75 (0.62)	22.75 (0.83)	14.40 (0.58)	15.88(0.06)
σ_χ^{nc}	3.21					

σ_χ^{nc} is the standard deviation of the non-cyclical component. Parameters Estimates using different filters, small variance of non-cyclical shock

Filter		LT	HP	FOD	BP	Flexible
Parameter	True	Median (s.d.)	Median (s.d.)	Median (s.d.)	Median(s.d.)	Median(s.d.)
σ_c	3.00	1.89 (0.07)	1.89 (0.07)	1.87 (0.07)	2.03 (0.09)	3.26 (0.29)
σ_n	0.70	2.13 (0.08)	2.11 (0.08)	2.15 (0.08)	1.90 (0.08)	0.80 (0.13)
h	0.70	0.58 (0.02)	0.60 (0.02)	0.56 (0.02)	0.69 (0.02)	0.77 (0.04)
α	0.60	0.47 (0.02)	0.46 (0.02)	0.49 (0.02)	0.24 (0.03)	0.41 (0.04)
ϵ	7.00	3.85 (0.13)	3.92 (0.13)	3.46 (0.11)	4.16 (0.13)	6.95 (0.09)
ρ_r	0.20	0.68 (0.03)	0.59 (0.03)	0.43 (0.04)	0.50 (0.03)	0.31 (0.04)
ρ_π	1.20	1.14 (0.04)	1.25 (0.04)	1.25 (0.04)	1.23 (0.04)	1.25 (0.03)
ρ_y	0.05	-0.07 (0.00)	-0.01 (0.01)	-0.05 (0.02)	0.23 (0.01)	0.08 (0.10)
ζ_p	0.80	0.81 (0.03)	0.78 (0.03)	0.76 (0.03)	0.89 (0.03)	0.72 (0.02)
ρ_χ	0.50	1.00 (0.03)	1.00 (0.03)	1.00 (0.03)	0.97 (0.03)	0.69 (0.05)
ρ_z	0.80	0.90 (0.03)	0.92 (0.03)	0.91 (0.03)	0.98 (0.03)	0.90 (0.03)
σ_χ	1.12	0.09 (0.01)	0.31 (0.05)	0.61 (0.15)	1.87 (0.14)	1.28 (0.03)
σ_z	0.51	0.61 (0.07)	0.30 (0.04)	0.40 (0.05)	0.10 (0.01)	0.69 (0.01)
σ_{mp}	0.10	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)	0.24 (0.004)
σ_μ	20.60	18.00 (0.74)	18.04 (0.61)	15.89 (0.83)	17.55 (0.57)	12.73 (0.04)
σ_χ^{nc}	23.21					

Parameters Estimates using different filters; σ_χ^{nc} is the standard deviation of the non-cyclical component.

Why are estimates distorted?

- *Posterior proportional to likelihood times prior.*

- *Log-likelihood of the parameters (see Hansen and Sargent (1993))*

$$L(\theta|y_t) = A_1(\theta) + A_2(\theta) + A_3(\theta)$$

$$A_1(\theta) = \frac{1}{\pi} \sum_{\omega_j} \log \det G_\theta(\omega_j)$$

$$A_2(\theta) = \frac{1}{\pi} \sum_{\omega_j} \text{ trace } [G_\theta(\omega_j)]^{-1} F(\omega_j)$$

$$A_3(\theta) = (E(y) - \mu(\theta)) G_\theta(\omega_0)^{-1} (E(y) - \mu(\theta))$$

where $\omega_j = \frac{\pi j}{T}, j = 0, 1, \ldots, T-1,$, $G_\theta(\omega_j)$ *is the model based spectral density matrix of* y_t, $\mu(\theta)$ *the model based mean of* y_t, $F(\omega_j)$ *is the data based spectral density of* y_t *and* $E(y)$ *the unconditional mean of the data.*

- first term: sum of the one-step ahead forecast error matrix across frequencies;

- the second a penalty function, emphasizing deviations of the model-based from the data-based spectral density at various frequencies.

- the third another penalty function, weighting deviations of model-based from data-based means, with the spectral density matrix of the model at frequency zero.

- Suppose that the actual data is filtered so that frequency zero is eliminated and low frequencies deemphasized. Then

$$L(\theta|y_t) = A_1(\theta) + A_2(\theta)^*$$

$$A_2(\theta)^* = \frac{1}{\pi}\sum_{\omega_j} \text{ trace } [G_\theta(\omega_j)]^{-1}F(\omega_j)^*$$

where $F(\omega_j)^* = F(\omega_j)I_\omega$ *and* I_ω *is an indicator function.*

Suppose that $I_\omega = I_{[\omega_1,\omega_2]}$*, an indicator function for the business cycle frequencies, as in an ideal BP filter.*

The penalty $A_2(\theta)^*$ *matters only at these frequencies.*

Since $A_2(\theta)^*$ and $A_1(\theta)$ enter additively in the log-likelihood function, there are two types of biases in $\hat{\theta}$.

- estimates $F_\theta(\omega_j)^*$ only approximately capture the features of $F(\omega_j)^*$ at the required frequencies - the sample version of $A_2(\theta)^*$ has a smaller values at business cycle frequencies and a nonzero value at non-business cycle ones.

- To reduce the contribution of the penalty function to the log-likelihood, parameters are adjusted to make $[G_\theta(\omega_j)]$ close to $F(\omega_j)^*$ at those frequencies where $F(\omega_j)^*$ is not zero. This is done by allowing fitting errors in $A_1(\theta)$ large at frequencies $F(\omega_j)^*$ is zero - in particular the low frequencies.

Conclusions:

1) The volatility of the structural shocks will be overestimated - this makes $[G_\theta(\omega_j)]$ *close to* $F(\omega_j)^*$ *at the relevant frequencies.*

2) Their persistence underestimated - this makes $G_\theta(\omega_j)$ *small and the fitting error large at low frequencies.*

Estimated economy very different from the true one: agents' decision rules are altered.

- Higher perceived volatility implies distortions in the aversion to risk and a reduction in the internal amplification features of the model.

- Lower persistence implies that perceived substitution and income effects are distorted with the latter typically underestimated relative to the former.

- Distortions disappear if:

i) the non-cyclical component has low power at the business cycle frequencies. Need for this that the volatility of the non-cyclical component is considerably smaller than the volatility of the cyclical one.

ii) The prior eliminates the distortions induced by the penalty functions.

Question: What if we fit the filtered version of the model to the filtered data? (CKM (2008))

- *Log-likelihood*$=A_1(\theta)^* = \frac{1}{\pi}\sum_{\omega_j} \log\det G_\theta(\omega_j) I_\omega + A_2(\theta)$. *Suppose that* $I_\omega = I_{[\omega_1,\omega_2]}$.

- $A_1(\theta)^*$ *matters only at business cycle frequencies while the penalty function is present at all frequencies.*

- *If the penalty is more important in the low frequencies (typical case) parameters adjusted to make* $[G_\theta(\omega_j)]$ *close to* $F(\omega_j)$ *at these frequencies.*

-Procedure implies that the model is fitted to the low frequencies components of the data!!!

i) Volatility of the shocks will be generally underestimated.

ii) Persistence overestimated.

iii) Since less noise is perceived, decision rules will imply a higher degree of predictability of simulated time series.

iv) Perceived substitution and income effects are distorted with the latter overestimated.

How can we avoid distortions?

- Build models with non-cyclical components (difficult).

- Use filters which flexibly adapt, see Gorodnichenko and Ng (2007) and Eklund, et al. (2008).

- ?

Advantages of suggested approach:

• No need to take a stand on the properties of the non-cyclical component and on the choice of filter to tone down its importance - specification errors and biases limited.

• Estimated cyclical component not localized at particular frequencies of the spectrum.

- Cyclical, non-cyclical and measurement error fluctuations driven by different and orthogonal shocks. But model is observationally equivalent to one where cyclical and non-cyclical are correlated.

An experiment, again

- Simulate data from the model, assuming that the preference shock has two components: a nonstationary one and a stationary one (the properties of the other three shocks are unchanged).

- Variance of the non-cyclical shock is large or small relative to the variance of the other shocks.

- Use same Bayesian approach, same prior for structural parameters and gamma priors with large variance for non-structural ones.

- Compute the model-based cyclical component; calculate the autocorrelation function and the log spectrum of output after passing it through LT, HP, FOD, BP.

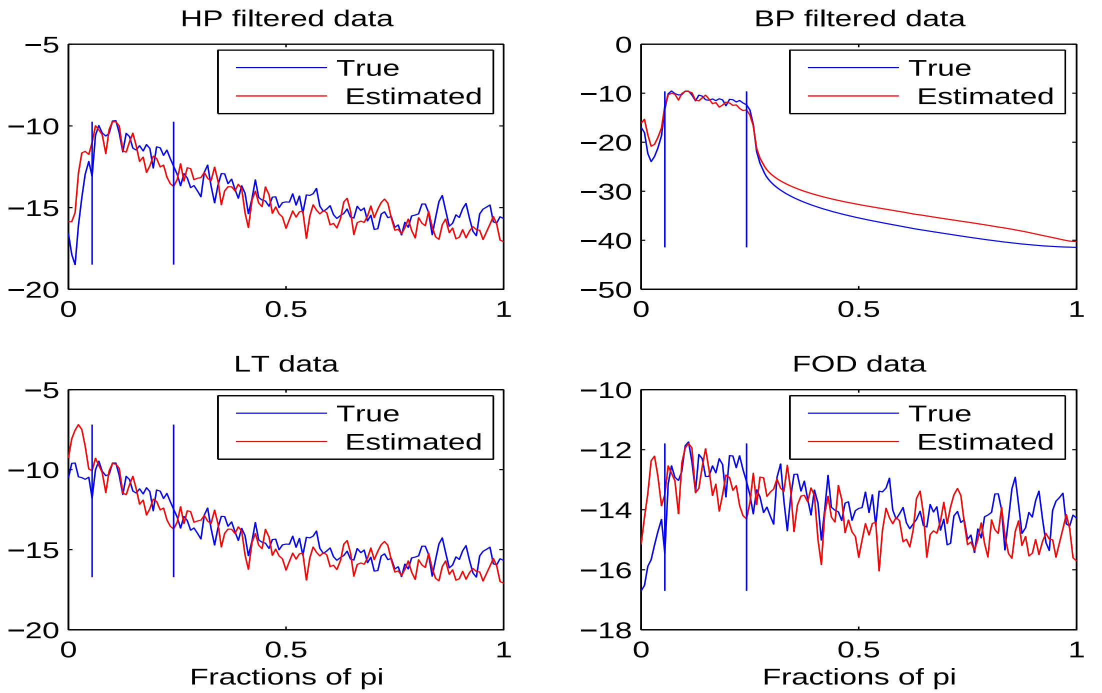
		Small variance			Large variance	
	True	Median	(s.e)	True	Median	(s.e)
σ_c	3.00	3.68	(0.40)	3.00	3.26	(0.29)
σ_n	0.70	0.54	(0.14)	0.70	0.80	(0.13)
h	0.70	0.55	(0.04)	0.70	0.77	(0.04)
α	0.60	0.19	(0.03)	0.60	0.41	(0.04)
ϵ	7.00	6.19	(0.07)	7.00	6.95	(0.09)
ρ_r	0.20	0.16	(0.04)	0.24	0.31	(0.04)
ρ_π	1.30	1.30	(0.04)	1.30	1.25	(0.03)
ρ_y	0.05	0.07	(0.03)	0.05	0.08	(0.10)
ζ_p	0.80	0.78	(0.04)	0.80	0.72	(0.02)
ρ_χ	0.50	0.53	(0.04)	0.50	0.69	(0.05)
ρ_z	0.80	0.71	(0.03)	0.80	0.90	(0.03)
σ_χ	0.011	0.012	(0.0003)	0.011	0.012	(0.0003)
σ_z	0.005	0.006	(0.0001)	0.005	0.007	(0.0001)
σ_{mp}	0.001	0.002	(0.0004)	0.001	0.002	(0.0004)
σ_μ	0.206	0.158	(0.0006)	0.206	0.1273	(0.0004)
σ_χ^{nc}	0.02			0.23		

Parameters estimates using flexible specification. σ_χ^{nc} is the standard error of the shock to the non-cyclical component.

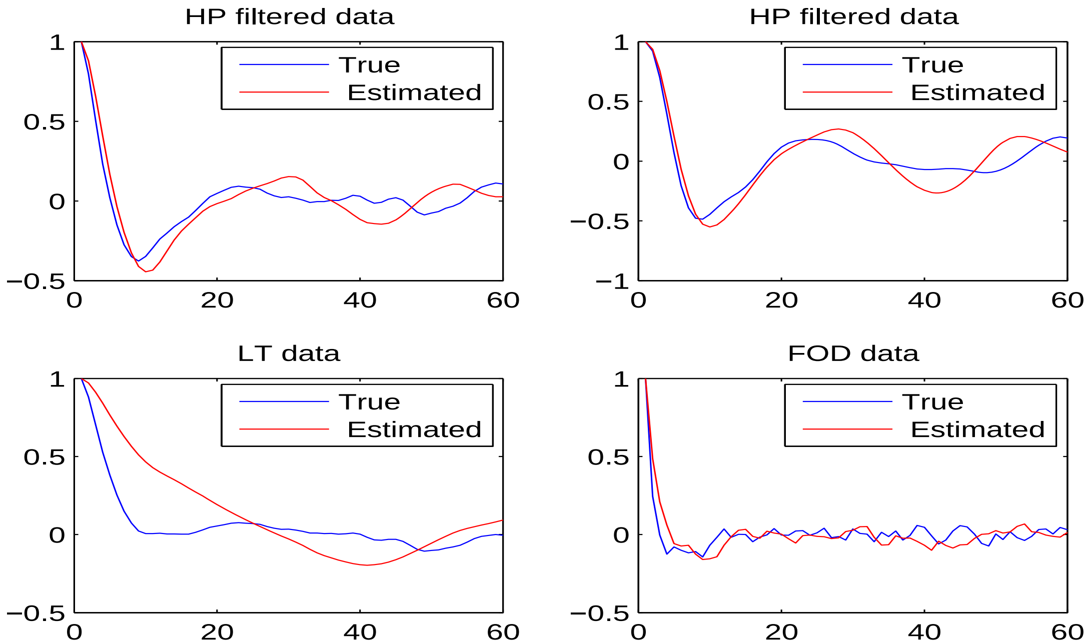
- *Estimates of the structural parameters are roughly unchanged in two specifications.*

- *Estimates are precise but the median is not the true value (problem bigger for* α *or* σ_n *which are only weakly identified).*

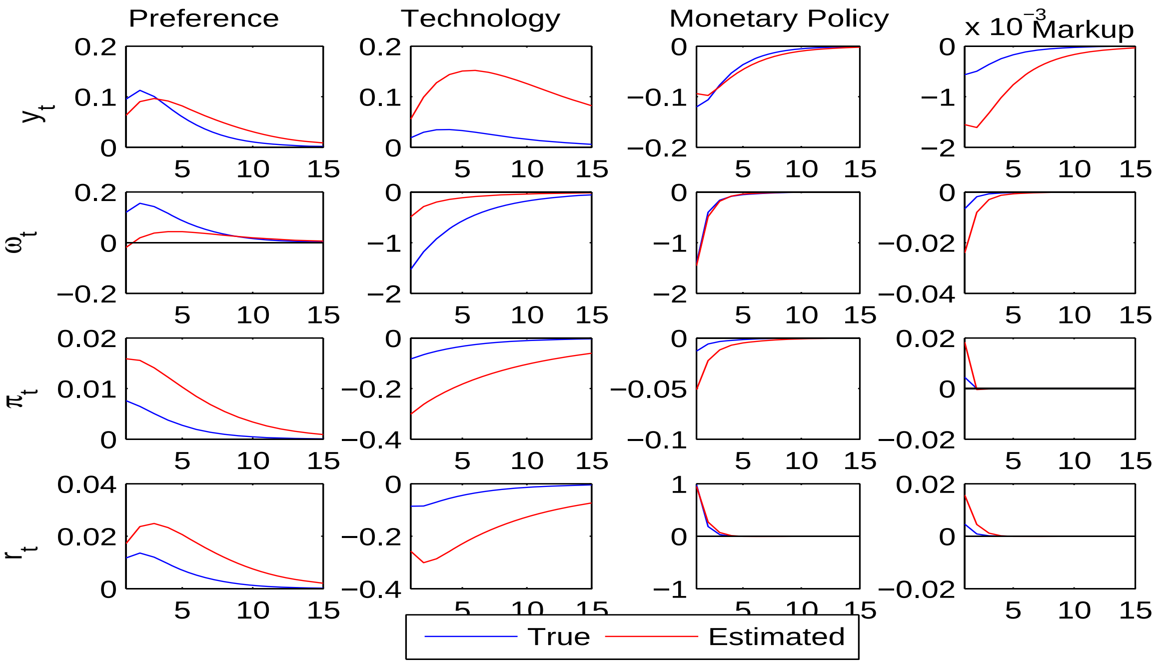
- *The relative magnitude of the various shocks and their persistence is well estimated. Hence, true and estimated decision rules are similar.*



Model based cyclical output spectra, true and estimated, different filtering. Vertical bars indicate the frequencies where cycles with 8-32 quarters periodicities are located



Autocorrelation function of filtered cyclical component, true and estimated



Model based IRF, true and estimated.

- The true and estimated log spectrum and the autocorrelation function of the model-based cyclical component close, regardless of the filter.

- Both true and estimate cyclical components have power at all frequencies of the spectrum.

Actual data: do we get a different story?

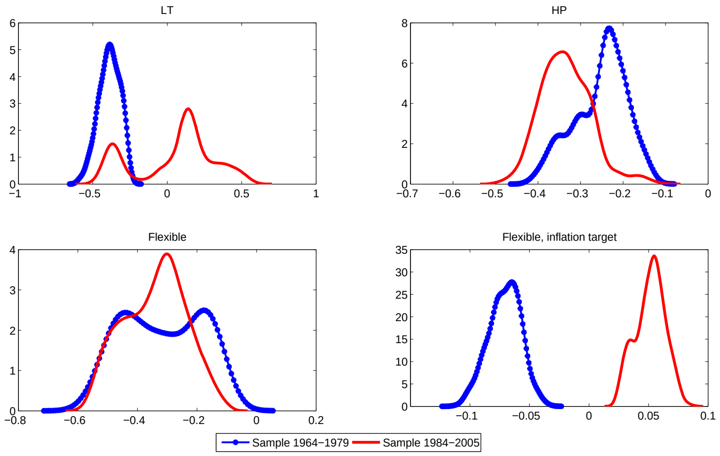


Figure 5: Posterior distributions of the policy activism parameter, samples 1964:1-1979:4 and 1984:1-2007:4. LT refers to linearly detrended data, HP to Hodrick and Prescott filtered data and Flexible to the approach the paper suggests

	LT		FOD		Flexible	
	Output	Inflation	Output	Inflation	Output	Inflation
TFP shocks	0.01	0.04	0.00	0.01	0.01	0.19
Gov. expenditure shocks	0.00	0.00	0.00	0.00	0.00	0.02
Investment shocks	0.08	0.00	0.00	0.00	0.00	0.05
Monetary policy shocks	0.01	0.00	0.00	0.00	0.00	0.01
Price markup shocks	0.75(*)	0.88(*)	0.91(*)	0.90(*)	0.00	0.21
Wage markup shocks	0.00	0.01	0.08	0.08	0.03	0.49(*)
Preference shocks	0.11	0.04	0.00	0.00	0.94(*)	0.00

Variance decomposition at the 5 years horizon. Estimates are obtained using the median of the posterior of the parameters. A (*) indicates that the 68 percent highest credible set is entirely above 0.10. The model and the data set are the same as in Smets Wouters (2007). LT refers to linearly detrended data, FOD to growth rates and Flexible to the approach this paper suggests.

Non linear DSGE models

$$y_{2t+1} = h_1(y_{2t}, \epsilon_{1t}, \theta) \quad (75)$$
$$y_{1t} = h_2(y_{2t}, \epsilon_{2t}, \theta) \quad (76)$$

ϵ_{2t} = *measurement errors,* ϵ_{1t} = *structural shocks,* θ = *vector of structural parameters,* y_{2t} = *vector of states,* y_{1t} = *vector of controls. Let* $y_t = (y_{1t}, y_{2t})$*,* $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})$*,* $y^{t-1} = (y_0, \ldots, y_{t-1})$ *and* $\epsilon^t = (\epsilon_1, \ldots, \epsilon_t)$*.*

• *Likelihood is* $\mathcal{L}(y^T, \theta y_{20}) = \prod_{t=1}^T f(y_t|y^{t-1}, \theta) f(y_{20}, \theta)$*. Integrating the initial conditions* y_{20} *and the shocks out, we have:*

$$\mathcal{L}(y^T, \theta) = \int [\prod_{t=1}^{T} \int f(y_t|\epsilon^t, y^{t-1}, y_{20}, \theta) f(\epsilon^t|y^{t-1}, y_{20}, \theta) d\epsilon^t] f(y_{20}, \theta) dy_{20} \quad (77)$$

(77) is intractable.

- *If we have* L *draws for* y_{20} *from* $f(y_{20}, \theta)$ *and* L *draws for* $\epsilon^{t|t-1,l}$*,* $l = 1, \ldots, L, \;\; t = 1, \ldots, T$*, from* $f(\epsilon^t | y^{t-1}, y_{20}, \theta)$ *approximate (77) with*

$$\mathcal{L}(y^T, \theta) = \frac{1}{L}[\prod_{t=1}^{T} \frac{1}{L} \sum_{l} f(y_t | \epsilon^{t|t-1,l}, y^{t-1}, y_{20}^{l}, \theta)] \tag{78}$$

Drawing from $f(y_{20}, \theta)$ *is simple; drawing from* $f(\epsilon^t | y^{t-1}, y_{20}, \theta)$ *complicated. Fernandez-Villaverde and Rubio-Ramirez (2004):*
use $f(\epsilon^{t-1} | y^{t-1}, y_{20}, \theta)$ *as importance sampling for* $f(\epsilon^t | y^{t-1}, y_{20}, \theta)$*:*

- Draw y_{20}^l *from* $f(y_{20},\theta)$*. Draw* $\epsilon^{t|t-1,l}$ L *times from* $f(\epsilon^t|y^{t-1},y_{20}^l,\theta) = f(\epsilon^{t-1}|y^{t-1},y_{20}^l,\theta)f(\epsilon_t|\theta)$*.*

- Construct $IR_t^l = \frac{f(y_t|\epsilon^{t|t-1,l},y^{t-1},y_{20}^l,\theta)}{\sum_{l=1}^L f(y_t|\epsilon^{t|t-1,l},y^{t-1},y_{20}^l,\theta)}$ *and assign it to each draw* $\epsilon^{t|t-1,l}$*.*

- Resample from $\{\epsilon^{t|t-1,l}\}_{l=1}^L$ *with probabilities equal to* IR_t^l*.*

- Repeat above steps for every $t = 1,2,\ldots,T$*.*

Step 3) is crucial, if omitted, only one particle will asymptotically remain and the integral in (77) diverges as $T \to \infty$*.*

• *Algorithm is computationally demanding. You need a MC within a MC. Fernandez-Villaverde and Rubio-Ramirez (2004): some improvements over linear specifications.*