Topics in Bayesian estimation of DSGE models

Fabio Canova
EUI and CEPR
February 2014

Outline

- DSGE-VAR.

- Data selection.

- Data rich DSGE (proxies, multiple data, conjunctural information, indicators of future variables).

- Dealing with trends and non-balanced growth

- Prior elicitation.

- Non-linear DSGE.

References

Aguiar, M. and G. Gopinath, (2007) Emerging market business cycles: The cycle is the trend, Journal of Political Economy, 115, 69–102.

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

Bi, X. and Traum, N. (2013) Estimating Fiscal limits: the case of Greece, forthcoming, Journal of Applied Econometrics.

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 raw data, manuscript.

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

Canova, F., F. Ferroni, C. Matthes (2013). Choosing the variables to estimated singular DSGE models, Journal of Applied Econometrics, forthcoming.

Canova, F. and Pappa, E. (2007) Price dispersion in monetary unions. The role of fiscal shocks, Economics Journal, 117, 713-737.

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

Chari, V., Kehoe, P. and McGrattan, E. (2008) Are structural VARs with long run restrictions useful in developing business cycle theory, Journal of Monetary Economics, 55, 1137-1355.

Del Negro, M. and F. Schorfheide (2004), " 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.

Eklund, J., R.Harrison, G. Kapetanios, and A. Scott, (2008). Breaks in DSGE models, manuscript.

Faust, J. and Gupta, A. (2012) Posterior Predictive Analysis for Evaluating DSGE Models, NBER working paper 17906.

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

Gorodnichenko Y. and S. Ng, 2010 Estimation of DSGE models when the data are persistent, Journal of Monetary Economics, 57, 325–340.

Hansen, L. and T. Sargent, 1993. Seasonality and approximation errors in rational expectations models, Journal of Econometrics, 55, 21–55.

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.

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

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

Lombardi, M and Nicoletti, G. (2011) Bayesian prior elicitation in DSGE models, ECB working paper 1289.

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.

Smets, F. and Wouters, R. (2007). Shocks and frictions in the US Business Cycle: A Bayesian DSGE approach, American Economic Review, 97, 586–606.

1 Combining DSGE and VARs

Recall:

- Log linearized solution of a DSGE model is

$$y_{2t} = \mathcal{A}_{22}(\theta) y_{2t-1} + \mathcal{A}_{21}(\theta) y_{3t} \quad (1)$$

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

- $y_{2t} =$ states and the driving forces, $y_{1t} =$ controls, y_{3t} shocks.

- $\mathcal{A}_{ij}(\theta), i, j = 1, 2$ are time invariant (reduced form) matrices which depend on θ, the structural parameters of preferences, technologies, policies, etc.

- So far we have used the likelihood $f(y|\theta)$ and a prior $g(\theta)$ to construct a posterior $g(\theta|y)$, where the likelihood is built using the DSGE model.

- Now we take an intermediate step. We specify $g(\theta)$, we use the model to derive $g(\alpha, \Sigma_u|\theta)$ and build the likelihood $f(y|\alpha, \Sigma_u)$.

Thus: if

- $g(\theta)$ is the prior distribution for DSGE parameters

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

- $f(y|\alpha, \Sigma_u)$ is likelihood of the data conditional on the reduced form parameters (this the VAR represention of the data)

Del Negro and Schorfheide(2004): The joint posterior of VAR and structural parameters is

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

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

$$
\begin{aligned}
f(y|\theta) &= \frac{|T_1 x^{s'}(\theta) x^s(\theta) + X'X|^{-0.5M} |(T_1 + T)\tilde{\Sigma}_u(\theta)|^{-0.5(T_1+T-k)}}{|\tau x^{s'}(\theta) x^s(\theta)|^{-0.5M} |T_1 \tilde{\Sigma}_u^s(\theta)|^{-0.5(T_1-k)}} \\
&\times \frac{(2\pi)^{-0.5MT} 2^{-0.5M(T_1+T-k)} \prod_{i=1}^{M} \Gamma(0.5 * (T_1 + T - k + 1 - i))}{2^{-0.5M(T_1-k)} \prod_{i=1}^{M} \Gamma(0.5 * (T_1 - k + 1 - i))} \qquad (3)
\end{aligned}
$$

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

- Since $g(\theta|y)$ is non-standard draw θ using a MH algorithm.

- Dynare has now an option to jointly estimate a DSGE model and the VAR which is consistent with the (log-) linear decision rules of the model.

- This is an application of Hierachical Bayes models (see Canova, ch.9).

- Advantage of the procedure do not need to choose between estimating a VAR or a DSGE. Can do both.

- First, construct a draw for θ. Then, given θ, construct posterior of α (draw α from a Normal-Wishart, conditional on θ).

Estimation algorithm: Set $T_1 = \bar{T}_1$.

1) Draw a candidate θ. Use MCMC to decide if accept or reject.

2) With the draw compute the model induced prior for the VAR parameters.

3) Compute the posterior for the VAR parameters (analytically if you have a conjugate structure or via the Gibbs sampler if you do not have one). Draw from this posterior

4) Repeat steps 1)-3) $NL + \bar{L}$ times. Check convergence and compute the Marginal likelihood.

5) Repeat 1)-4) for different T_1. Choose the T_1 that maximizes the marginal likelihood.

Example 1.1 *In a basic sticky price-sticky wage economy, fix* $\eta = 0.66, \pi^{ss} = 1.005, N^{ss} = 0.33, \frac{c}{gdp} = 0.8, \beta = 0.99, \zeta_p = \zeta_w = 0.75, a_0 = 0, a_1 = 0.5, a_2 = -1.0, a_3 = 0.1$. *Run a VAR with output, interest rates, money and inflation using actual quarterly data from 1973:1 to 1993:4 and data simulated from the model conditional on these parameters. Overall, only a modest amount of simulated data (roughly, 20 data points) should be used to set up a prior.*

Marginal Likelihood, Sticky price sticky wage model.

$\kappa = 0$	$\kappa = 0.1$	$\kappa = 0.25$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$
-1228.08	-828.51	-693.49	-709.13	-913.51	-1424.61

2 Choice of data and estimation

- DSGE models typically singular. Does it matter which variables are used to estimate the parameters? Yes.

i) Omitting relevant variables may lead to distortions in parameter estimates.

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 2.1

$$\begin{aligned} y_t &= a_1 E_t y_{t+1} + a_2(i_t - E_t \pi_{t+1}) + v_{1t} && (4) \\ \pi_t &= a_3 E_t \pi_{t+1} + a_4 y_t + v_{2t} && (5) \\ i_t &= a_5 E_t \pi_{t+1} + v_{3t} && (6) \end{aligned}$$

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 *identifies no parameter !!)*

iv) Likelihood function may change shape depending on the variables used. 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 Villaverde and Rubio Ramirez (2008, NBER macro annual) find habit in consumption is 0.88. Why? They use different data to estimate the same model!

Can we say something systematic about the choice of variables?

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

- Internal persistence of the model changes 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.

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)

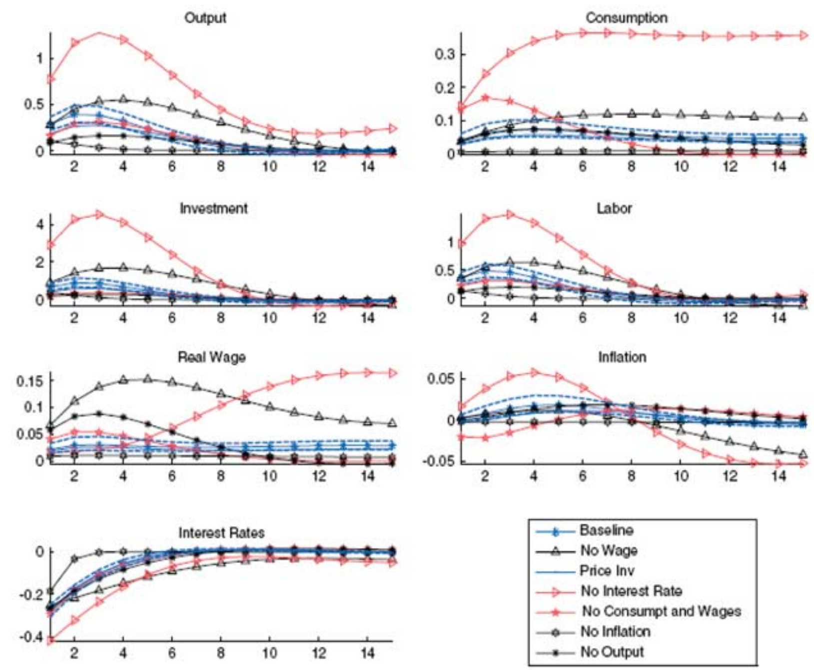


Figure 1. Responses to an expansionary monetary shock. This figure is available in color online at www.interscience.wiley.com/journal/jae

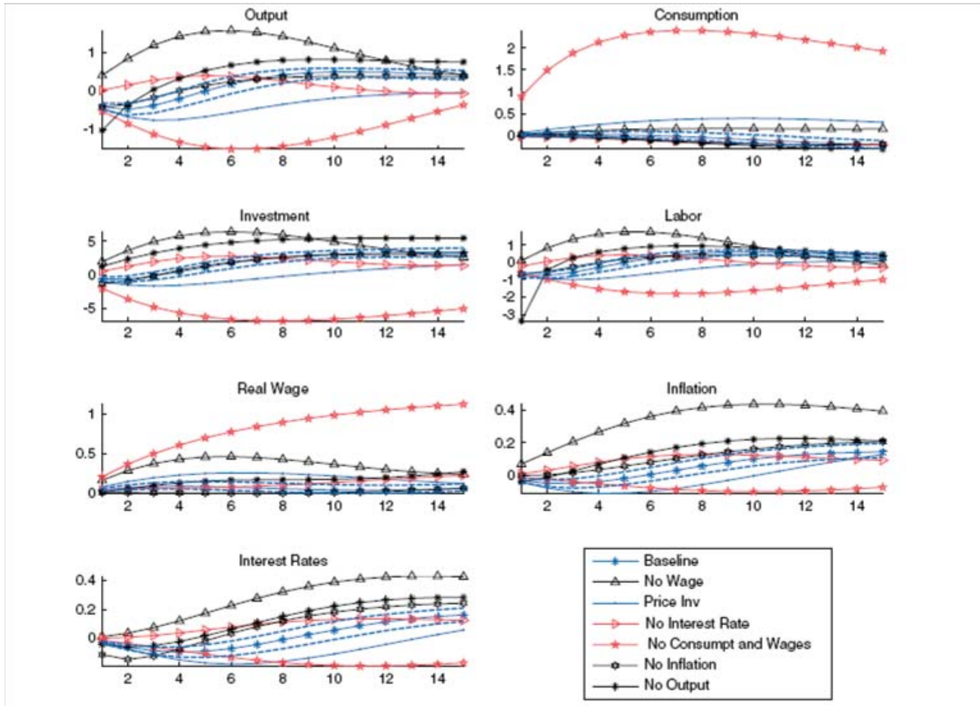


Figure 2. Responses to a positive investment-specific technology shock. This figure is available in color online at www.interscience.wiley.com/journal/jae

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

Canova, Ferroni and Matthes (2013)

- Use statistical criteria to select variables to be used in estimation

1) Choose vector that maximize the identificability of relevant parameters.

Compute the rank of the derivative of the spectral density of the model solution with respect to the parameters, see Komunjer and Ng (2011)

Choose the combination of observables which gives you a rank as close as possible to the ideal.

2) Compare the curvature of the convoluted likelihood in the singular and the non-singular systems in the dimensions of interest to eliminate ties.

3) Choose vector that minimize the information loss going from the larger scale to the smaller scale system. Information loss is measured by

$$p_t^j(\theta, e^{t-1}, u_t) = \frac{\mathcal{L}(W_{jt}|\theta, e^{t-1}, u_t)}{\mathcal{L}(Z_t|\theta, e^{t-1}, u_t)} \tag{7}$$

where $\mathcal{L}(.|\theta, y_{1t})$ is the likelihood of Z_t, W_{jt} defined by

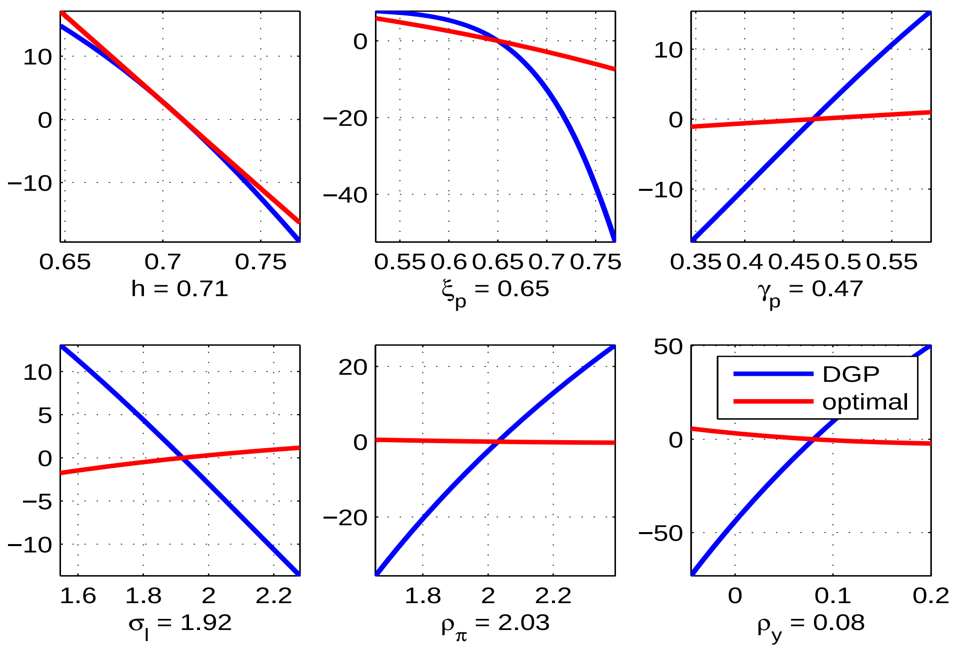
$$\begin{aligned} Z_t &= y_t + u_t \quad (8) \\ W_{jt} &= Sy_{jt} + u_t \quad (9) \end{aligned}$$

u_t is an iid convolution error, y_t the original set of variables and y_{jt} the j-th subset of the variables producing a non-singular system.

• Apply procedures to SW model driven with 4 shocks and 7 potential observables.

Vector	Unrest Rank(Δ)	SW Restr Rank(Δ)	SW Restr and Sixth Restr
y, c, i, w	186	188	ψ
y, c, i, π	185	188	ψ
y, c, r, h	185	188	ψ
y, i, w, r	185	188	ψ
c, i, w, h	185	188	ψ, σ_c, ρ_i
c, i, π, h	185	188	ψ
c, i, r, h	185	188	$\zeta_\omega, \zeta_p, i_\omega$
y, c, i, r	185	187	
...			
c, w, π, r	183	187	
c, w, π, h	183	187	
i, w, π, r	183	187	
w, π, r, h	183	187	
c, i, π, r	183	186	
Ideal	189	189	

Rank conditions for all combinations of variables in the unrestricted SW model (columns 2) and in the restricted SW model (column 3), where $\delta = 0.025$, $\varepsilon_p = \varepsilon_w = 10$, $\lambda_w = 1.5$ and $c/g = 0.18$. The fourth columns reports the extra parameter restriction needed to achieve identification; a blank space means that there are no parameters able to guarantee identification.



Likelihood curvature

	Basic		T=1500		$\Sigma_u = 0.01 * I$	
Order	Vector	Relative Info	Vector	Relative info	Vector	Relative Info
1	(y, c, i, h)	1	(y, c, i, h)	1	(y, c, i, h)	1
2	(y, c, i, w)	0.89	(y, c, i, w)	0.87	(y, c, i, w)	0.86
3	(y, c, i, r)	0.52	(y, c, i, r)	0.51	(y, c, i, r)	0.51
4	(y, c, i, π)	0.5	(y, c, i, π)	0.5	(y, c, i, π)	0.5

Ranking based on the information statistic. The first two column present the results for the basic setup, the next six columns the results obtained altering some nuisance parameters. Relative information is the ratio of the $p(\theta)$ statistic relative to the best combination.

• How different are good and bad combinations?

- Simulate 200 data points from the model with four shocks and estimate structural parameters using

(1) Model A: 4 shocks and (y, c, i, w) as observables (best rank analysis).

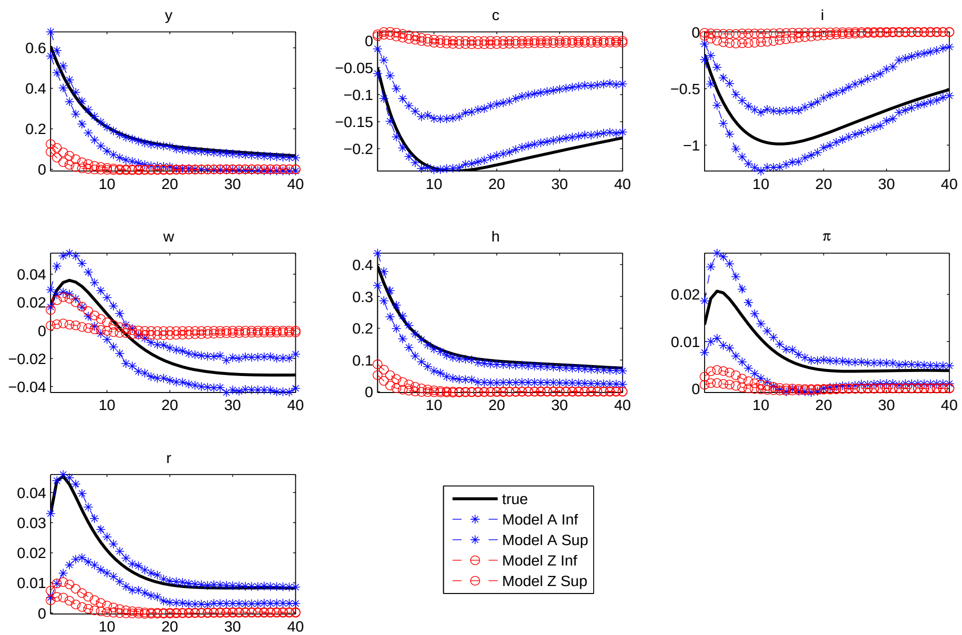
(2) Model B: 4 shocks and (y, c, i, w) as observables (best information analysis).

(3) Model Z: 4 shocks and (c, i, π, r) as observables(worst rank analysis).

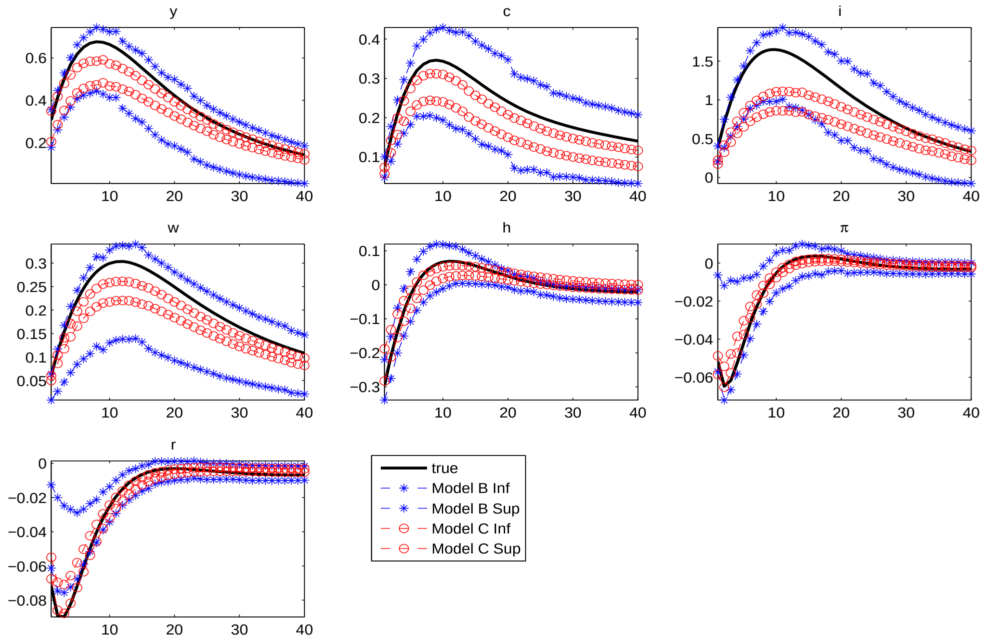
(4) Model C: 4 structural shocks, three measurement errors and $(y_t, c_t, i_t, w_t, \pi, r_t, h_t)$ as observables.

(5) Model D: 7 structural shocks (add price and wage markup and preference shocks) and $(y_t, c_t, i_t, w_t, \pi, r_t, h_t)$ as observables.

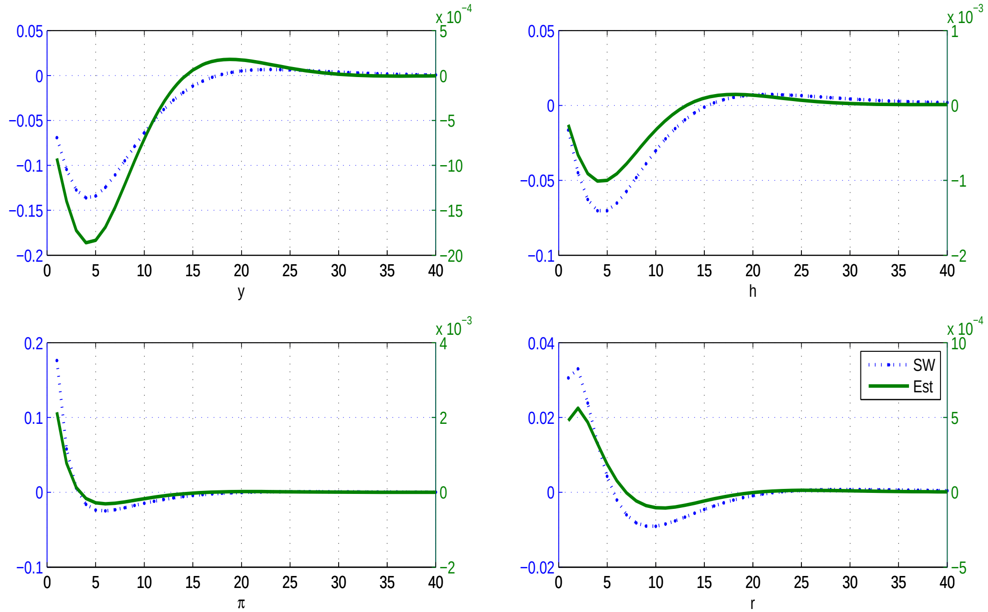
	True	Model A	Model B	Model Z	Model C	Model D
ρ_a	0.95	(0.920 , 0.975)	(0.905 , 0.966)	(0.946 , 0.958)	(0.951 , 0.952)	(0.939 , 0.943)*
ρ_g	0.97	(0.930 , 0.969)	(0.930 , 0.972)	(0.601 , 0.856)*	(0.970 , 0.971)	(0.970 , 0.972)
ρ_i	0.71	(0.621 , 0.743)	(0.616 , 0.788)	(0.733 , 0.844)*	(0.681 , 0.684)*	(0.655 , 0.669)*
ρ_{ga}	0.51	(0.303 , 0.668)	(0.323 , 0.684)	(0.010 ,0.237)*	(0.453 , 0.780)	(0.114 , 0.885)*
σ_n	1.92	(1.750 , 2.209)	(1.040 , 2.738)	(0.942 , 2.133)	(1.913 , 1.934)	(1.793 , 1.864)*
σ_c	1.39	(1.152 , 1.546)	(1.071 , 1.581)	(1.367 , 1.563)	(1.468 , 1.496)*	(1.417 , 1.444)*
h	0.71	(0.593 , 0.720)	(0.591 , 0.780)	(0.716 , 0.743)	(0.699 , 0.701)*	(0.732 , 0.746)*
ζ_ω	0.73	(0.402 , 0.756)	(0.242, 0.721)*	(0.211 ,0.656)*		(0.806 , 0.839)*
ζ_p	0.65	(0.313 , 0.617)*	(0.251 , 0.713)	(0.512 , 0.616)*	(0.317 , 0.322)*	(0.509 , 0.514)*
i_ω	0.59	(0.694 , 0.745)	(0.663 , 0.892)*	(0.532 ,0.732)	(0.728 , 0.729)*	(0.683 , 0.690)*
i_p	0.47	(0.571 , 0.680)*	(0.564 , 0.847)*	(0.613 , 0.768)*	(0.625 , 0.628)*	(0.606 , 0.611)*
ϕ_p	1.61	(1.523 , 1.810)	(1.495 , 1.850)	(1.371 , 1.894)	(1.624 , 1.631)*	(1.654 , 1.661)*
φ	0.26	(0.145 , 0.301)	(0.153 , 0.343)	(0.255 , 0.373)	(0.279 , 0.295)*	(0.281 , 0.306)*
ψ	5.48	(3.289 , 7.955)	(3.253 , 7.623)	(2.932 , 7.530)	(11.376 , 13.897)*	(4.332 , 5.371)*
α	0.2	(0.189 , 0.331)	(0.167 , 0.314)	(0.136 , 0.266)	(0.177 , 0.198)*	(0.174 , 0.199)*
ρ_π	2.03	(1.309 , 2.547)	(1.277 , 2.642)	(1.718 , 2.573)	(1.868 , 1.980)*	(2.119 , 2.188)*
ρ_y	0.08	(0.001 , 0.143)	(0.001 , 0.169)	(0.012 , 0.173)	(0.124 , 0.162)*	
ρ_R	0.87	(0.776 , 0.928)	(0.813 , 0.963)	(0.868 , 0.916)	(0.881 , 0.886)*	
$\rho_{\Delta y}$	0.22	(0.001 , 0.167)*	(0.010, 0.192)*	(0.130 ,0.215)*	(0.235 , 0.244)*	
σ_a	0.46	(0.261 , 0.575)	(0.382 , 0.460)	(0.420 ,0.677)	(0.357 , 0.422)*	(0.386 , 0.455)*
σ_g	0.61	(0.551 , 0.655)	(0.551 , 0.657)	(0.071 ,0.113)	(0.536 , 0.629)	(0.585 , 0.688)*
σ_i	0.6	(0.569 , 0.771)	(0.532 , 0.756)	(0.503 ,0.663)	(0.561 , 0.660)	(0.693 , 0.819)*
σ_r	0.25	(0.100 , 0.259)	(0.078 , 0.286)	(0.225 ,0.267)	(0.226 , 0.265)	(0.222 , 0.261)



Responses to a goverment spending shock



Responses to a technology shock



Responses to an price markup shock

Alternatives:

- Solve out variables from the FOC before you compute the solution until the number of observables is the same as the number of shocks. Which variables do we solve out?

- Good strategy to follow if some component of y_t are non-observable.

- But format of the solution is no longer a restricted VAR(1) (it is a VARMA).

- Add measurement errors until the combined number of structural shocks and measurement errors equal the number of observables. Thus, 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. Can add up to four. How many should we use?

Here the model represents the state equations (all are non-observables) and the measurement equation is

$$x_{2t} = F_1 y_t + e_t \tag{10}$$

- 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).

ii) 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 (see also Watson (1993)). Useful device when θ is calibrated. Less useful when θ is estimated.

iii) Canova (2010): measurement error has a complex structure (see later).

3 Practical issues

Log-linear DSGE solution:

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

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

How to you estimate DSGE models on the data when:

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

b) there are multiple observables that correspond to model quantities?

c) have additional information one would like to use, but it is not included in the model.

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

- GDP is revised for up to three years; savings in the model do not correspond to the savings computed in the national statistics. 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?

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

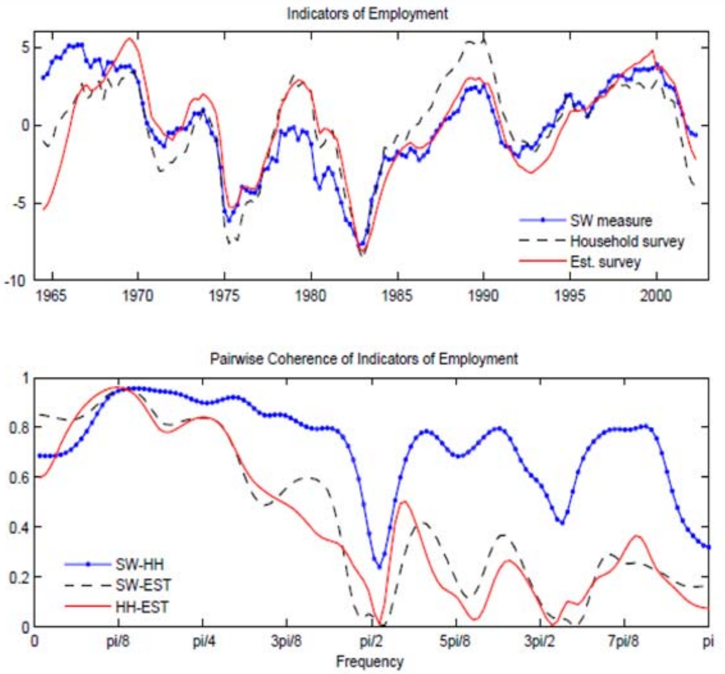


Figure 1: Indicators of employment (de-trended) and pairwise coherence.

- Do we use CPI inflation, GDP deflator or PCE inflation?

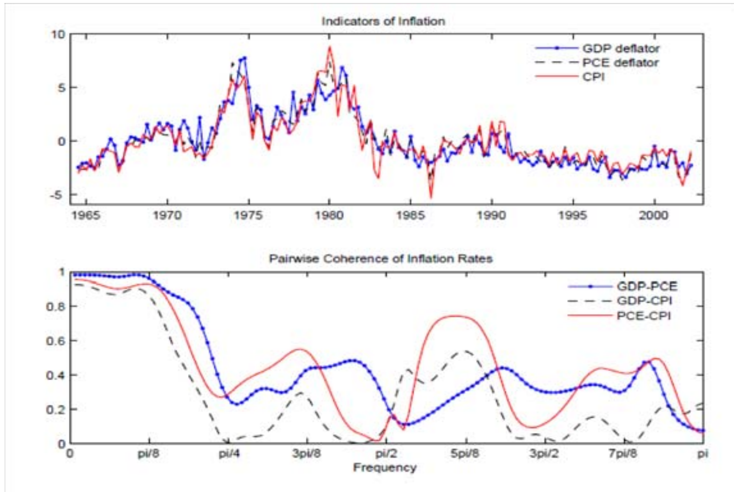


Figure 2: Indicators of quarterly inflation rates (de-meaned) and pairwise coherence.

- Different measures contain (noisy) information about the true series. Not perfectly correlated among each other.

Case 1: Measurement error is present.

Observables x_t. Model based quantities $x_t^m(\theta) = S[y_{1t}, y_{2t}]$, F is a selection matrix.

$$x_t = x_t^m(\theta) + u_t$$

where u_t is iid measurement error.

- In all other cases use ideas underlying factor models

- For b) let x_{1t} be a $k \times 1$ vector of observable variables and $x_t^m(\theta)$ be of dimension $N \times 1$ where dim dim (N) <dim(k). Then:

$$x_{1t} = \Lambda_3 x_t(\theta)^m + u_{1t} \tag{13}$$

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

$$\begin{aligned} x_{1t} &= \Lambda_3[S_1 y_{1t}, S_2 \mathcal{A}_{12}(\theta) y_{1t-1} + F_1 \mathcal{A}_{13}(\theta) y_{3t}]' + u_{3t} && (14) \\ &= \Lambda_3[S_1 y_{1t}, S_1 \mathcal{B}(\theta) y_{1t}]' + u_{3t} && (15) \end{aligned}$$

where u_t is iid measurement error.

- x_{1t} 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_t is used to measure y_{1t}, estimate of θ will probably be noisy.

- 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, see Justiniano et al. (2012).

- How different is the specification 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.

Many cases fit in c):

1) 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, see Beaudry and Portier (2006).

2) Sometimes we have survey data to proxy for unbosreved states (e.g business cycles).

3) Sometimes we have conjunctoral information.

- Can use these measures to get information about the states. Let x_t a $q \times 1$ vector of variables. Assume

$$x_{2t} = \Lambda_4 y_t + u_{2t} \tag{16}$$

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

$$X_t = \Lambda y_{1t} + u_t \tag{17}$$

where $X_t = [x_{1t}, x_{2t}]'$, $u_t = [u_{1t}, u_{2t}]$ 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.

- Thus, we interpret data information through the lenses of the DSGE model.

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

3.1 An example

Consider a three equation New-keynesian model:

$$x_t = E_t(x_{t+1}) - \frac{1}{\phi}(i_t - E_t\pi_{t+1}) + e_{1t} \quad (18)$$

$$\pi_t = \beta E_t\pi_{t+1} + \kappa x_t + e_{2t} \quad (19)$$

$$i_t = \psi_r i_{t-1} + (1-\psi_r)(\psi_\pi \pi_t + \psi_x x_t) + e_{3t} \quad (20)$$

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

$$e_{1t} = \rho_1 e_{1t-1} + v_{1t} \quad (21)$$

$$e_{2t} = \rho_2 e_{2t-1} + v_{2t} \quad (22)$$

$$e_{3t} = v_{3t} \quad (23)$$

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

- There are ambiguities in linking the output gap, the inflation rate and the nominal interest rate to empirical counterparts. Which the nominal interest rate should we use? How do we measure the gap?

Write the solution of the model as

$$x_t^m = RR(\theta)x_{t-1}^m + SS(\theta)v_t \tag{24}$$

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

Assume that (24) is the state equation of the system and that the measurement equation is

$$W_t = \Lambda w_t + e_t \tag{25}$$

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 (24)-(25) with standard methods. The routines give us estimates of Λ, RR, SS and of w_t which are consistent with the data.

Conjunctoral information

- Can use conjunctoral information in the same way as any other data that can give us information about the states.

- 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.

- Suppose 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:

$$W_t = \Lambda w_t + e_t \tag{26}$$

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

- Estimates of 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 and regular data and information about the path of the shocks. Information is optimally mixed depending on their relative precision.

Using Mixed frequency data

- High frequency data very useful to understand the state of the economy (e.g. tapering of US expansionary onetary policy).

- Macro data available at much lower frequencies. How do we combine high and low freqnecy information?

- Suppose use have monthly data in addition to standard quartely macro data. Let x_{jt} the quartely version of the monthly data, obtained using data from the j-month of the quarter. Set $X_t = [x_{1t}, x_{2t}.x_{3t}]'$. The model is

$$X_t = \Lambda x_t^m(\theta) + u_t \tag{27}$$

See Foroni and Marcellino (2013).

4 Dealing with trends and non-balanced growth paths

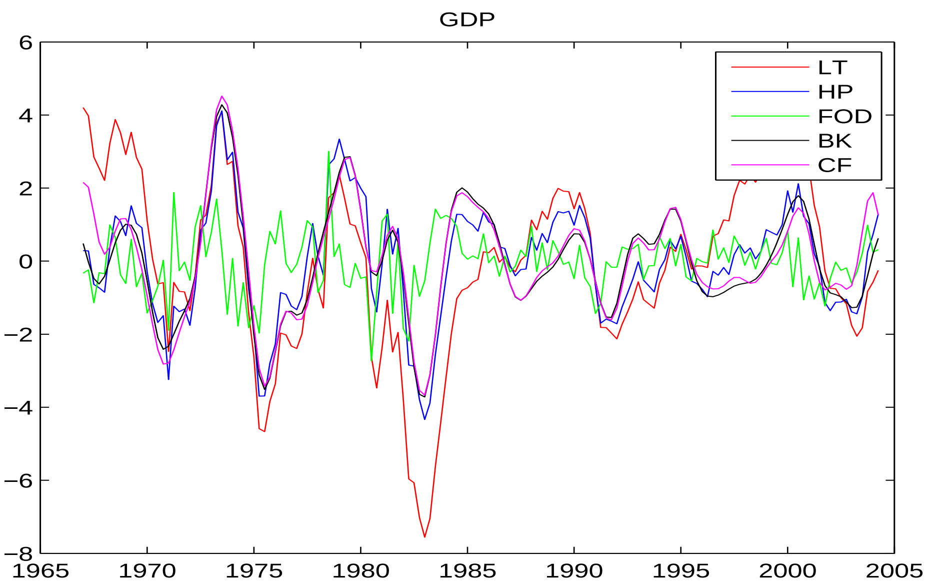
- Most of models available for policy are stationary and cyclical.

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

- How to we match models to the data?

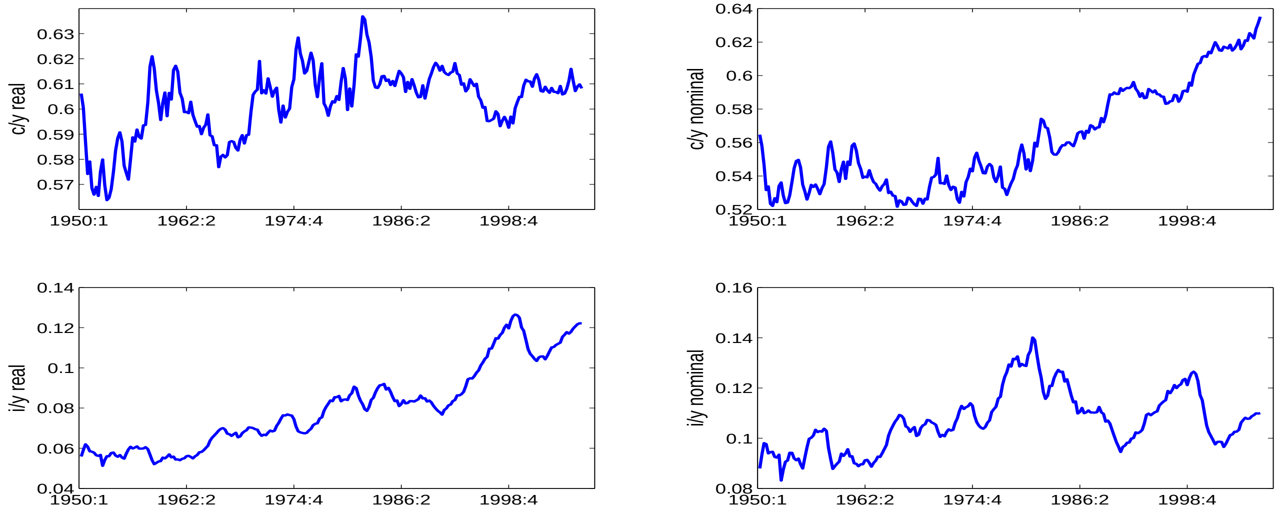
a) Detrend actual data: the model is a representation for detrended data.

Problem: which detrended data is the model representing?



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

b) Take ratios in the dta and in the model - will get rid of trends if variables in the ratio are cointegrated. Problem: data does not seem to satisfy balanced growth (the variables in the ratios are not cointegrated)



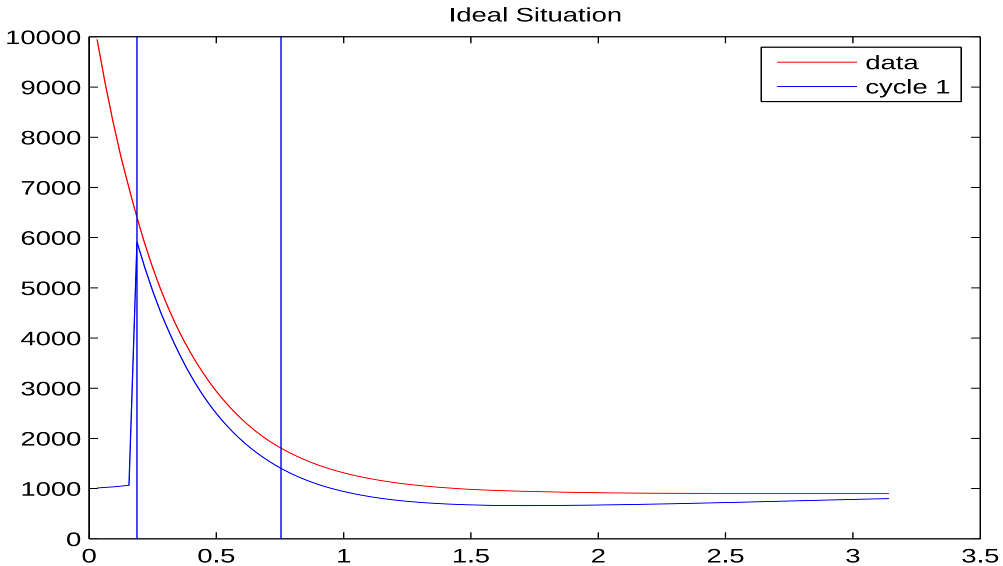
Real and nominal Great ratios in US, 1950-2008.

c) Build-in a trend into the model. Detrend the data with model-based trend. Problems

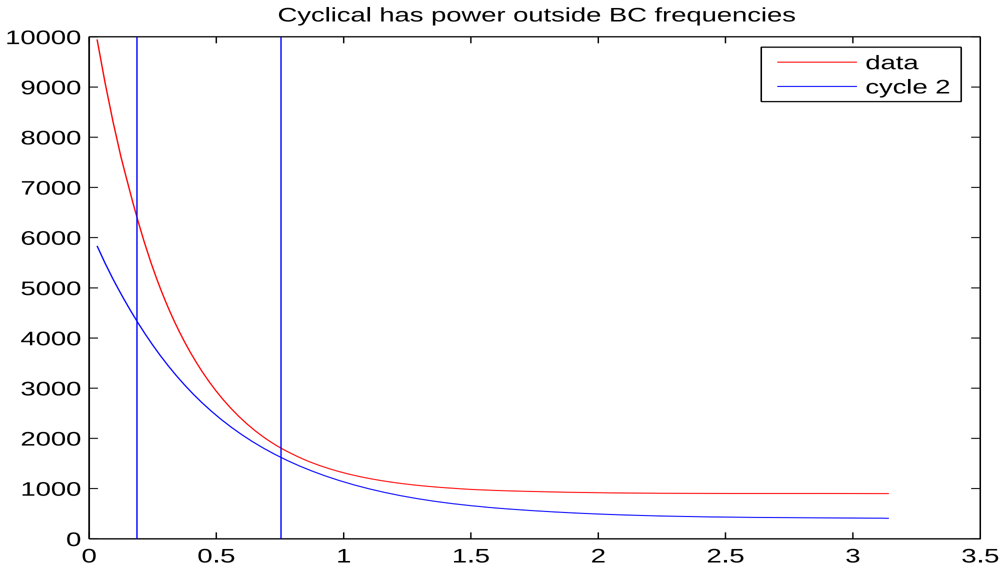
1) Specification of the trend is arbitary (deterministic? stochastic?).

2) Where you put the trend (TFP? preference?) matters for estimation and inference.

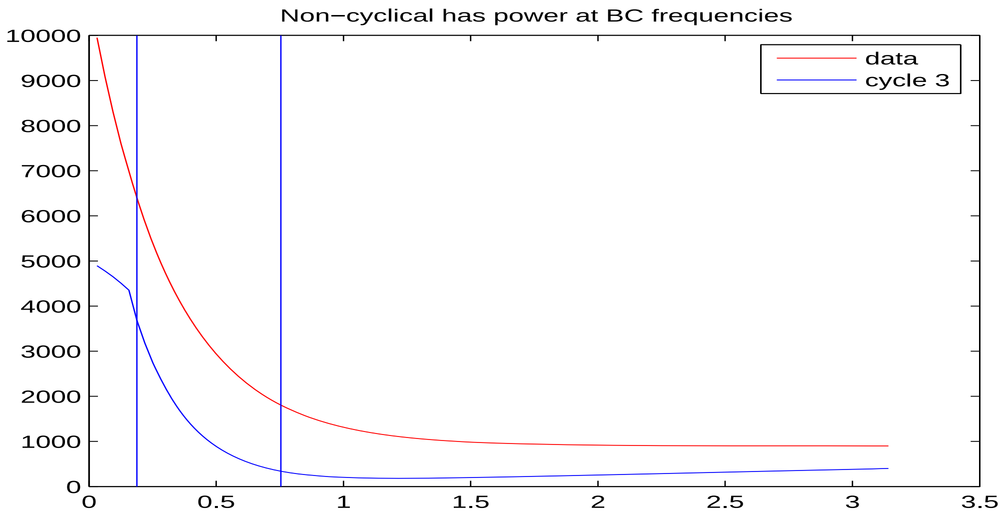
• General problem: statistical definition of a cycle is different from the economic definition. All statistical approaches are biased, even in large samples.


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

Ideal case



Realistic case



General case

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

Two approaches to deal with the problem:

1) Data-rich environment, see 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:

$$y_t^d = \lambda_0 + \lambda_1' y_t(\theta) + u_t \tag{28}$$

where $\lambda_j, j = 0, 1$ are matrices of parameters, measuring the bias and correlation between the filter data y_t^d and model based quantities $y_t(\theta)$; u_t are measurement errors and θ the structural parameters.

- Factor model setup a-la Boivin and Giannoni (2005); model based quantities are non-observable.

- Jointly estimate θ and λ's. Can obtain a more precise estimates 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 (Canova, 2014)).

$$y_t^d = c + y_t^T + y_t^m(\theta) + u_t \tag{29}$$

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

$$y_t = RR(\theta)x_{t-1} + SS(\theta)z_t \tag{30}$$

$$x_t = PP(\theta)x_{t-1} + QQ(\theta)z_t \tag{31}$$

$$z_{t+1} = NN(\theta)z_t + \epsilon_{t+1} \tag{32}$$

$PP(\theta), QQ(\theta), RR(\theta), SS(\theta)$ 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

$$y_t^T = \rho_1 y_{t-1}^T + \bar{y}_{t-1} + e_t \qquad e_t \sim \ iid \ (0, \Sigma_e^2) \tag{33}$$

$$\bar{y}_t = \rho_2 \bar{y}_{t-1} + v_t \qquad v_t \sim \ iid \ (0, \Sigma_v^2) \tag{34}$$

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

$\rho_1 = \rho_2 = I, \Sigma_v^2 = 0$, and $\Sigma_e^2 > 0$, y_t^T is a vector of I(1) processes.

$\rho_1 = \rho_2 = I, \Sigma_v^2 = \Sigma_e^2 = 0$, y_t^T is deterministic.

$\rho_1 = \rho_2 = I, \Sigma_v^2 > 0$ and $\Sigma_e^2 > 0$ and $\frac{\sigma_v^2}{\sigma_e^2}$ is large, y_t^t is "smooth"(as in HP).

$\rho_1 \neq I, \rho_2 \neq I$ or both, nonmodel based component has power at particular freqnecies

- Jointly estimate structural θ and non-structural parameters $(\rho_1, \rho_2, \Sigma_e, \Sigma_u)$.

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.

Example 4.1 *The log linearized equilibrium conditions of basic NK model are:*

$$\lambda_t = \chi_t - \frac{\sigma_c}{1-h}(y_t - hy_{t-1}) \quad (35)$$

$$y_t = z_t + (1-\alpha)n_t \quad (36)$$

$$w_t = -\lambda_t + \sigma_n n_t \quad (37)$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)(\rho_\pi \pi_t + \rho_y y_t) + v_t \quad (38)$$

$$\lambda_t = E_t(\lambda_{t+1} + r_t - \pi_{t+1}) \quad (39)$$

$$\pi_t = k_p(w_t + n_t - y_t + \mu_t) + \beta E_t \pi_{t+1} \quad (40)$$

$$z_t = \rho_z z_{t-1} + \iota_t^z \quad (41)$$

where $k_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}\frac{1-\alpha}{1-\alpha+\varepsilon\alpha}$, λ_t *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.*

Estimate this model with a number of detrending transformations. Do we get different estimates?

	Prior	LT	HP	FOD	BP	Ratio 1	Ratio2
		Median (s.e.)	Median (s.e.)	Median (s.e.)	Median (s.e.)	Median(s.e.)	Median (s.e.)
σ_c	$\Gamma(20, 0.1)$	1.90 (0.25)	1.41 (0.21)	0.04 (0.01)	0.96 (0.11)	2.33 (0.27)	0.81 (0.15)
σ_n	$\Gamma(20, 0.1)$	1.75 (0.16)	1.37 (0.13)	5.23 (0.08)	1.19 (0.09)	3.02 (0.24)	2.68 (0.19)
h	$B(6, 8)$	0.83 (0.02)	0.88 (0.02)	0.45 (0.01)	0.96 (0.01)	0.72 (0.05)	0.88 (0.02)
α	$B(3, 8)$	0.07 (0.04)	0.09 (0.05)	0.42 (0.01)	0.07 (0.03)	0.05 (0.04)	0.03 (0.01)
ρ_r	$B(6, 6)$	0.19 (0.05)	0.11 (0.04)	0.62 (0.01)	0.09 (0.02)	0.38 (0.06)	0.28 (0.04)
ρ_π	$N(1.5, 0.1)$	1.33 (0.08)	1.37 (0.05)	1.53 (0.02)	1.51(0.06)	1.92 (0.06)	1.80 (0.05)
ρ_y	$N(0.4, 0.1)$	-0.16 (0.03)	-0.18 (0.03)	0.06 (0.00)	-0.22 (0.03)	0.16 (0.02)	-0.03 (0.02)
ζ_p	$B(6, 6)$	0.82 (0.02)	0.80 (0.03)	0.63 (0.01)	0.86 (0.01)	0.82 (0.02)	0.80 (0.02)
ρ_χ	$B(18, 8)$	0.69 (0.04)	0.40 (0.05)	0.52 (0.01)	0.70(0.02)	0.67 (0.03)	0.66 (0.02)
ρ_z	$B(18, 8)$	0.96 (0.02)	0.95 (0.02)	0.99 (0.01)	0.97(0.01)	0.97 (0.01)	0.96 (0.01)
σ_χ	$\Gamma^{-1}(10, 20)$	0.53 (0.19)	0.47 (0.11)	4.96(0.13)	0.23 (0.05)	3.41 (0.74)	0.97 (0.13)
σ_z	$\Gamma^{-1}(10, 20)$	0.20 (0.04)	0.23 (0.04)	2.00 (0.22)	0.19 (0.03)	0.06 (0.01)	0.06 (0.01)
σ_r	$\Gamma^{-1}(10, 20)$	0.11 (0.01)	0.08 (0.01)	2.30(0.23)	0.07 (0.01)	0.10 (0.01)	0.11 (0.18)
σ_μ	$\Gamma^{-1}(10, 20)$	25.06 (0.97)	14.25 (0.93)	7.17 (0.13)	18.19 (0.66)	22.89 (1.91)	15.94 (0.49)

	Prior	Ratio 3	TFP	Preferences	TFP FD		
		Median (s.e.)	Median (s.e.)	Median (s.e.)	Median (s.e.)		
σ_c	$\Gamma(20, 0.1)$	0.12 (0.03)	1.0	1.0	1.0		
σ_n	$\Gamma(20, 0.1)$	2.09 (0.14)	2.24 (0.26)	2.43 (0.20)	0.50 (0.28)		
h	$B(6, 8)$	0.10 (0.03)	0.08 (0.04)	0.78 (0.03)	0.54 (0.29)		
α	$B(3, 8)$	0.03 (0.02)	0.17 (0.03)	1.0	0.49 (0.29)		
ρ_r	$B(6, 6)$	0.20 (0.06)	0.30 (0.04)	0.61 (0.02)	0.49 (0.28)		
ρ_π	$N(1.5, 0.1)$	1.51 (0.07)	1.74 (0.06)	1.69 (0.05)	1.69 (2.13)		
ρ_y	$N(0.4, 0.1)$	0.77 (0.04)	0.49 (0.03)	0.38 (0.07)	0.25 (1.97)		
ζ_p	$B(6, 6)$	0.81 (0.01)	0.41 (0.03)	0.84 (0.01)	0.47 (0.29)		
ρ_χ	$B(18, 8)$	0.75 (0.03)	0.63 (0.03)		0.49 (0.28)		
ρ_z	$B(18, 8)$	0.62 (0.03)		0.59 (0.02)			
σ_χ	$\Gamma^{-1}(10, 20)$	0.26 (0.04)	0.21 (0.03)	0.06 (0.008)	3.49(0.48)		
σ_z	$\Gamma^{-1}(10, 20)$	0.08 (0.01)	0.05 (0.006)	0.15 (0.02)	2.09 (0.89)		
σ_r	$\Gamma^{-1}(10, 20)$	2.68 (0.27)	0.10 (0.01)	0.07 (0.007)	0.79(0.55)		
σ_μ	$\Gamma^{-1}(10, 20)$	15.98 (1.09)	0.25 (0.04)	36.68 (1.42)	8.34(0.44)		

Table 2: Posterior estimates. LT refers to linearly detrended data, HP to Hodrick and Prescott filtered data, FOD to demeaned growth rates, BP to band pass filtered data. For Ratio 1 the observables are $\log(y_t/n_t), \log(w_t), \pi_t, r_t$, all demeaned, for Ratio 2 they are $\log(y_t/w_t), \log(n_t), \pi_t, r_t$, all demeaned, For Ratio 3, the observables are $\log((w_t n_t)/y_t), \log(w_t/y_t), \pi_t, r_t$, all demeaned. For TFP trending, the observable are linearly detrending output and real wages and demeaned inflation and interest rates. For Preference trending, the observable are demeaned growth rate of output, demeaned log real wages, demeaned inflation and demeaned interest rates. When frequency domain estimation is used, only information in the band $(\frac{\pi}{32}, \frac{\pi}{8})$ is employed. The sample is 1980:1-2007:4.

- Simulate data from a model where trend is unimportant and where trend is important.

- What happens to parameter estimates obtained with standard methods?

- Does the new method recover the DGP better in both cases?

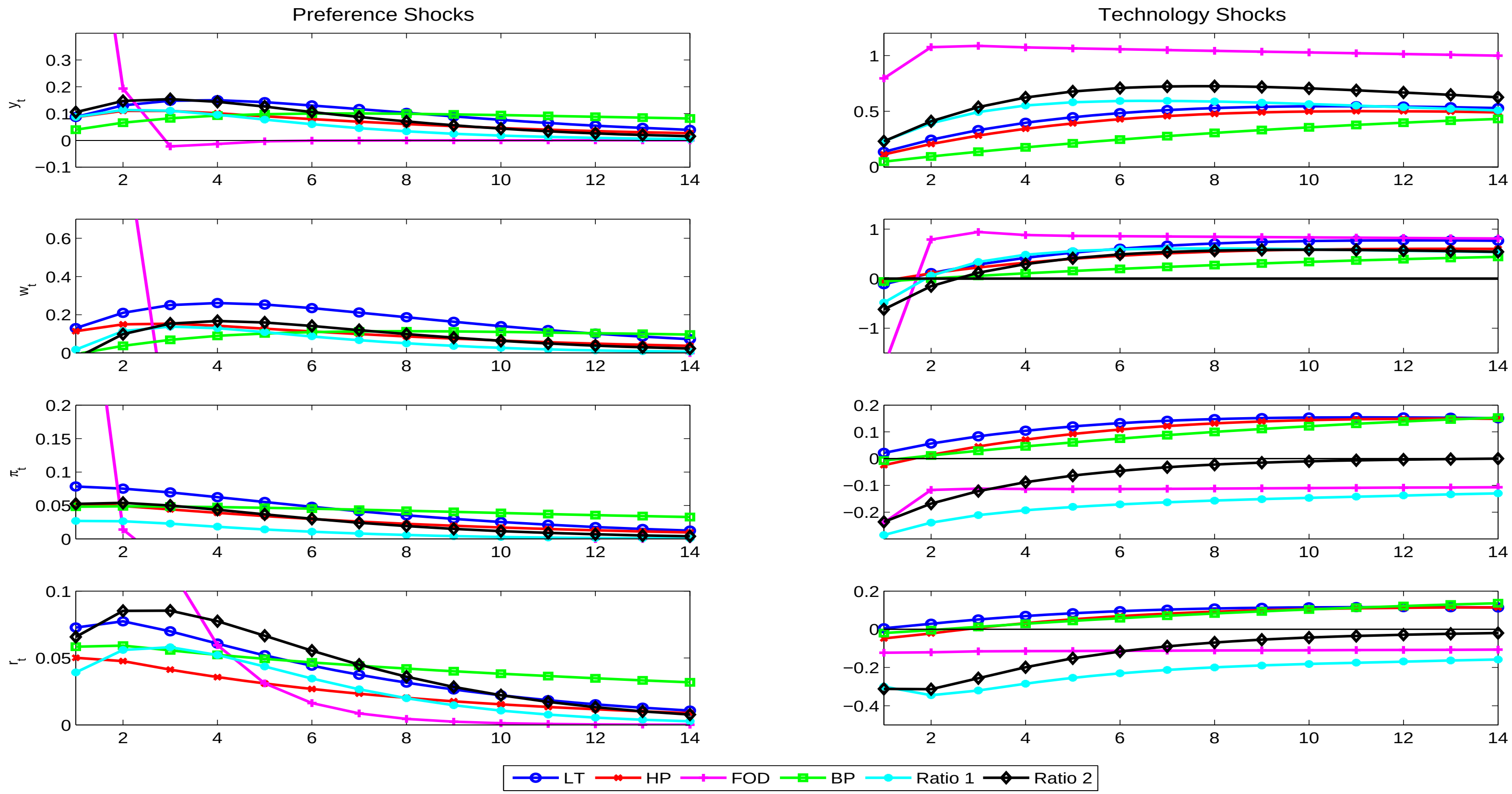
- What kind of parameters are distorted?

DGP1							
	True value	LT	HP	FOD	BP	Ratio1	Flexible
σ_n	0.50	0.04	0.08	0.00	0.11	0.05	0.04
h	0.70	0.00	0.00	0.00	0.01	0.07	0.10
α	0.30	0.00	0.04	0.00	0.06	0.04	0.06
ρ_r	0.70	0.05	0.05	0.01	0.06	0.13	0.01
ρ_π	1.50	0.00	0.00	0.00	0.01	0.02	0.00
ρ_y	0.40	0.17	0.20	0.17	0.19	0.15	0.00
ζ_p	0.75	0.03	0.04	0.03	0.03	0.02	0.03
ρ_χ	0.50	0.00	0.04	0.00	0.00	0.00	0.07
ρ_z	0.80	0.03	0.05	0.00	0.05	0.00	0.05
σ_χ	1.12	1.60	0.45	3.89	0.64	8.79	1.00
σ_z	0.50	1.47	0.01	3.18	0.03	0.02	0.16
σ_r	0.10	1.37	0.03	3.75	0.03	0.00	0.00
σ_μ	1.60	13.14	18.81	17.68	38.52	38.36	1.94
Total1		0.30	0.40	0.21	0.48	0.49	0.24
Total2		17.91	19.79	28.71	39.75	47.66	3.45

MSE. In DPG1 there is a unit root component to the preference shock and $\frac{\sigma_\chi^{nc}}{\sigma_\chi^T} = [1.1, 1.9]$.

DGP2							
	True value	LT	HP	FOD	BP	Ratio1	Flexible
σ_n	0.50	0.04	0.11	0.17	0.12	0.12	0.06
h	0.70	0.01	0.00	0.00	0.03	0.08	0.17
α	0.30	0.00	0.05	0.00	0.06	0.02	0.07
ρ_r	0.70	0.05	0.05	0.04	0.05	0.13	0.02
ρ_π	1.50	0.00	0.00	0.00	0.00	0.01	0.00
ρ_y	0.40	0.16	0.21	0.08	0.19	0.15	0.00
ζ_p	0.75	0.03	0.04	0.02	0.05	0.04	0.03
ρ_χ	0.50	0.00	0.04	0.00	0.00	0.01	0.08
ρ_z	0.80	0.04	0.05	0.03	0.03	0.00	0.06
σ_χ	1.12	10.41	0.87	2.80	0.69	9.43	0.97
σ_z	0.50	9.15	0.06	1.91	0.06	0.01	0.17
σ_r	0.10	9.35	0.00	1.05	0.03	0.00	0.00
σ_μ	1.60	10.41	20.72	20.33	57.03	40.17	1.90
Total1		0.29	0.46	0.32	0.51	0.55	0.35
Total2		39.65	22.20	26.44	58.34	50.17	3.54

MSE. In DGP2 all shocks are stationary but there is measurement error and $\frac{\sigma_u}{\sigma_\chi^T} = [0.09, 0.11]$ The MSE is computed using 50 replications.



Estimated impulse responses.

Why are estimates distorted with standard filtering?

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

- the third term: a 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? as suggested by Chari, Kehoe and McGrattan (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 (2010) and Eklund, et al. (2008).

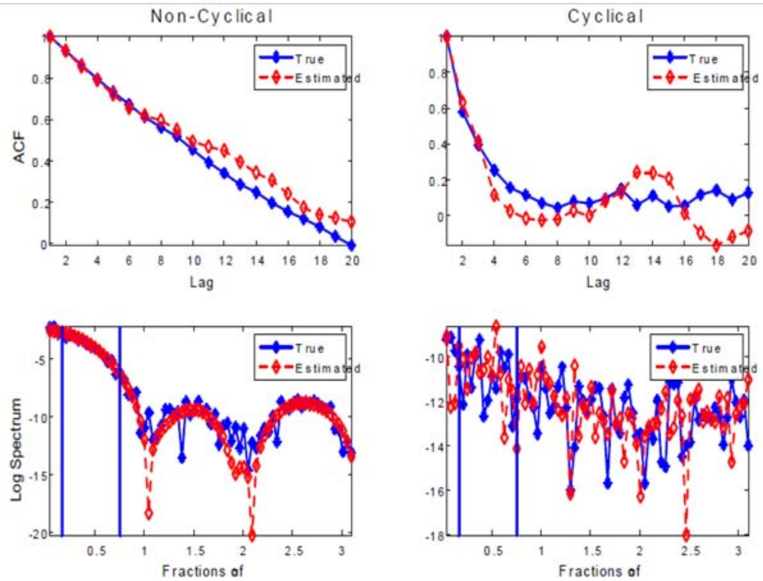
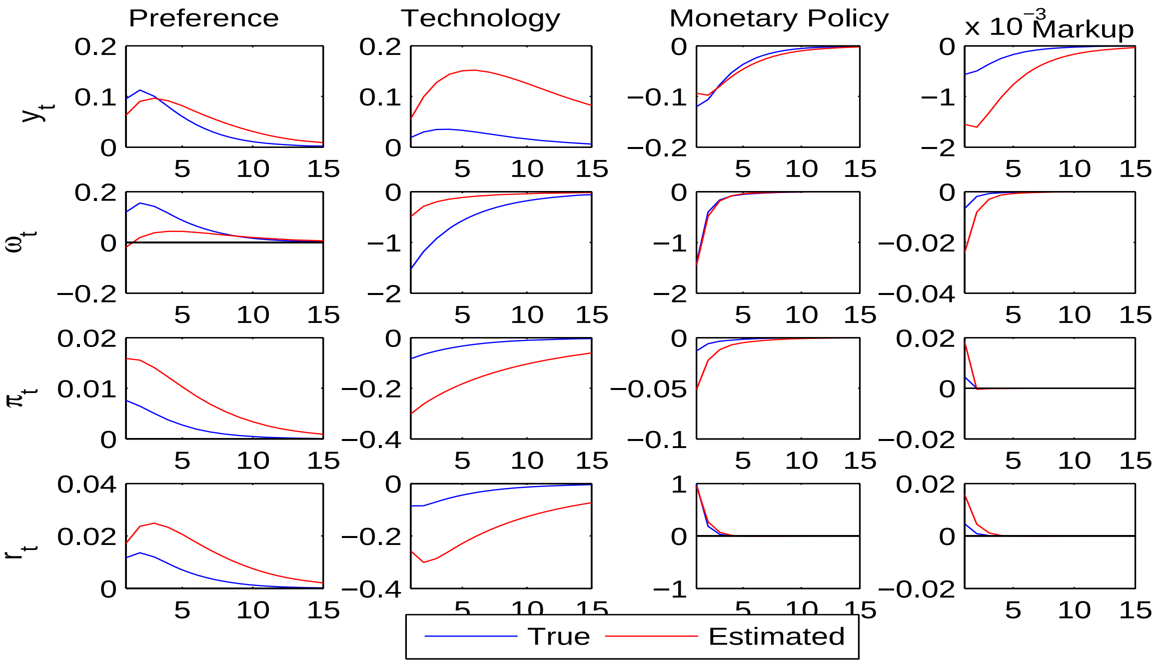


Figure 3: Output decomposition, true and estimated with a flexible approach. Vertical bars indicate business cycle frequencies

- The true and estimated log spectrum and ACF close.

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



Model based IRF, true and estimated.

Actual data: do we get a different story?

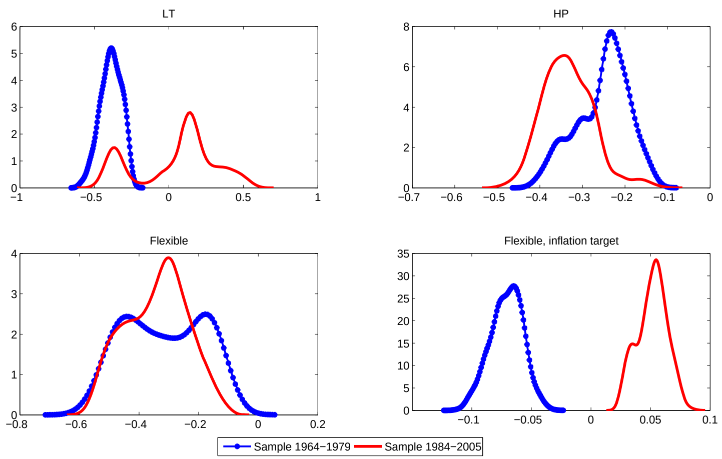


Figure 5: Posterior distributions of the policy activism parameter, samples 1964:1-1979:4 and 1984:1-2007:4. LT refers to linearly detrended data, HP to Hodrick and Prescott filtered data and Flexible to the approach the paper suggests

	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, SW model. 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.

5 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.

Example 5.1 *Let* $y_t = \theta_1 y_{t-1} + \theta_2 + u_t,\ u_t \sim N(0,1)$.. *Suppose* θ_1 *and* θ_2 *are independent and* $p(\theta_1) \sim U(0, 1-\epsilon), \epsilon > 0$; $p(\theta_2|\theta_1) \sim N(\bar{\mu}, \lambda)$.

Since the mean of y_t *is* $\mu = \frac{\theta_2}{1-\theta_1}$, *the prior for* θ_1 *and* θ_2 *imply that* $\mu|\theta_1 \sim N(\bar{\mu}, \frac{\lambda}{(1-\theta_1)^2})$. *Hence, the prior mean of* y_t *has a variance which is increasing in the persistence parameter* θ_1*! Why? Reasonable ?*

Alternative: state a prior for μ, *derive the prior for* θ_1 *and* θ_2 *(change of variables). For example, if* $\mu \sim N(\bar{\mu}, \lambda^2)$ *then* $p(\theta_1) = U(0, 1-\epsilon),\ p(\theta_2|\theta_1) = N(\bar{\mu}(1-\theta_1), \lambda^2(1-\theta_1)^2)$. *Note here that the priors for* θ_1 *and* θ_2 *are correlated.*

Suppose you want to compare the model with $y_t = \theta + u_t,\ u_t \sim N(0,1)$. *If* $p(\theta) = N(\bar{\mu}, \lambda^2)$ *the two models are immediately comparable. If, instead, we had assumed independent priors for* $p(\theta_1)$ *and* $p(\theta_2)$, *the two models would not be comparable (standard prior has weird predictions for the prior of the mean of* y_t*).*

- Del Negro and Schorfheide (2008): elicit priors consistent with some distribution of statistics of actual data (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

$$S_T = S_N(\theta) + \eta \quad \eta \sim (0, \Sigma_{TN}) \tag{42}$$

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 generic 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 (42).

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

$$p(\theta|S_T) \propto L(S_N(\theta)|S_T)\pi(\theta) \tag{43}$$

- $dim(S_T) \geq dim(\theta)$: 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, t.

- What are the S_T? Could be steady states, autocorrelation functions, etc. What S_T is depends on where the parameters enters.

Example 5.2

$$
\begin{aligned}
\max_{(c_t, K_{t+1}, N_t)} \quad & E_0 \sum_t \beta^t \frac{(c_t^{\vartheta}(1-N_t)^{1-\vartheta})^{1-\varphi}}{1-\varphi} && (44) \\
G_t + c_t + K_{t+1} &= GDP_t + (1-\delta)K_t && (45) \\
\ln \zeta_t &= \bar{\zeta} + \rho_z \ln \zeta_{t-1} + \epsilon_{1t} \quad \epsilon_{1t} \sim (0, \sigma_z^2) && (46) \\
\ln G_t &= \bar{G} + \rho_g \ln G_{t-1} + \epsilon_{4t} \quad \epsilon_{4t} \sim (0, \sigma_g^2) && (47) \\
GDP_t &= \zeta_t K_t^{1-\eta} N_t^{\eta} && (48)
\end{aligned}
$$

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 (45).

The FOC are ((52) and (53) equate factor prices and marginal products)

$$\lambda_t = \vartheta c_t^{\vartheta(1-\varphi)-1}(1-N_t)^{(1-\vartheta)(1-\varphi)} \tag{49}$$

$$\lambda_t \eta \zeta_t k_t^{1-\eta} N_t^{\eta-1} = -(1-\vartheta) c_t^{\vartheta(1-\varphi)}(1-N_t)^{(1-\vartheta)(1-\varphi)-1} \tag{50}$$

$$\lambda_t = E_t \beta \lambda_{t+1}[(1-\eta)\zeta_{t+1} K_{t+1}^{-\eta} N_{t+1}^{\eta} + (1-\delta)] \tag{51}$$

$$w_t = \eta \frac{GDP_t}{N_t} \tag{52}$$

$$r_t = (1-\eta)\frac{GDP_t}{K_t} \tag{53}$$

Using (49)-(50) we have:

$$-\frac{1-\vartheta}{\vartheta}\frac{c_t}{1-N_t} = \eta \frac{GDP_t}{N_t} \tag{54}$$

Log linearizing the equilibrium conditions

$$
\begin{aligned}
\hat{\lambda}_t - (\vartheta(1-\varphi)-1)\hat{c}_t + (1-\vartheta)(1-\varphi)\frac{N^{ss}}{1-N^{ss}}\hat{N}_t &= 0 \quad (55)\\
\hat{\lambda}_{t+1} + \frac{(1-\eta)(GDP/K)^{ss}}{(1-\eta)(GDP/K)^{ss}+(1-\delta))}(\widehat{GDP}_{t+1} - \hat{K}_{t+1}) &= \hat{\lambda}_t \quad (56)\\
\frac{1}{1-N^{ss}}\hat{N}_t + \hat{c}_t - \widehat{gdp}_t &= 0 \quad (57)\\
\hat{w}_t - \widehat{GDP}_t + \hat{n}_t &= 0 \quad (58)\\
\hat{r}_t - \widehat{GDP}_t + \hat{k}_t &= 0 \quad (59)\\
\widehat{GDP}_t - \hat{\zeta}_t - (1-\eta)\hat{K}_t - \eta\hat{N}_t &= 0 \quad (60)\\
(\frac{g}{GDP})^{ss}\hat{g}_t + (\frac{c}{GDP})^{ss}\hat{c}_t + (\frac{K}{GDP})^{ss}(\hat{K}_{t+1} - (1-\delta)\hat{K}_t) - \widehat{GDP}_t &= 0 \quad (61)
\end{aligned}
$$

(60) and (61) 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 (51)-(54)-(45)) is:

$$\frac{1-\vartheta}{\vartheta}\left(\frac{c}{GDP}\right)^{ss} = \eta\frac{1-N^{ss}}{N^{ss}} \quad (62)$$

$$\beta\left[(1-\eta)\left(\frac{GDP}{K}\right)^{ss}+(1-\delta)\right] = 1 \quad (63)$$

$$\left(\frac{g}{GDP}\right)^{ss}+\left(\frac{c}{GDP}\right)^{ss}+\delta\left(\frac{K}{GDP}\right)^{ss} = 1 \quad (64)$$

$$\frac{GDP}{wc} = \eta \quad (65)$$

$$\frac{K}{i} = \delta \quad (66)$$

Five equations in 8 parameters!! Need to choose.

For example: (62)-(66) 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 (62)-(66) 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 a prior for θ_2? Use additional information (statistics)!

- $(\frac{g}{GDP})^{ss}$ *could be centered at the average G/Y in the data with standard error covering the existing range of variations*

- $\beta = (1+r)^{-1}$ *and typically* $r^{ss} = [0.0075, 0.0150]$ *per quarter. Choose a prior centered at around those values and e.g. uniformly distributed.*

- ϑ *is related to Frish elasticity of labor supply: use estimates of labor supply elasticity to obtain histograms and to select a prior shape.*

Note: uncertainty in this case 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.*

Prior standard errors should reflect variations in the data of these parameters.

- 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 $var(c_t)$ and use

$$var(c_t) = var(c_t(\varphi)|\theta_1, \theta_2, \theta_3) + \eta \tag{67}$$

to back out a prior for φ.

For some parameters (call them θ_5) we have no moments to match but 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

$$
\begin{aligned}
p(\theta) \; = \; & p(\theta_1|S_{1T})p(\theta_2|S_{2T})p(\theta_3|S_{3T})p(\theta_4|S_{4T}) \\
& \pi(\theta_1)\pi(\theta_2)\pi(\theta_3)\pi(\theta_4)\Pi(\theta_5)
\end{aligned}
\tag{68}
$$

- 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 these priors, need a normalizing constant ((43 is not necessarily a density). 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.

Extension: Lombardi and Nicoletti (2011)

• Employ user-supplied impulse response to get a joint prior for the parameters

- γ^* user supplied vector of IRF; $\gamma(\theta)$ model based IRF.

- Distance function $d(\theta|\gamma^*) = vec(\gamma(\theta) - \gamma^*)W(\gamma(\theta) - \gamma^*)'$, W weighting matrix.

- Prior kernel: $w(\theta|\gamma^*, K) = \frac{exp(-d(\theta|\gamma^*))}{K(1+exp(-d(\theta|\gamma^*))}$.

- Prior: $p(\theta|\gamma^*) = \frac{w(\theta|\gamma^*,K)}{\int w(\theta|\gamma^*,K)d\theta}$.

$\theta = [\theta_1, \theta_2]$. Two special cases:

1) Prior kernel $w(\theta_1|\gamma^*, K, \bar{\theta_2})$ (some parameters do not enter impulse responses and are calibrated).

2) Prior kernel $w(\theta_1|\gamma^*, K, \theta_2)g(\theta_2|\bar{\theta}_2, \Sigma_{\theta_2})$ (prior for some parameters obtained from sources other than IRF).

6 Non linear DSGE models

$$y_{2t+1} = h_1(y_{2t}, \epsilon_{1t}, \theta) \quad (69)$$
$$y_{1t} = h_2(y_{2t}, \epsilon_{2t}, \theta) \quad (70)$$

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:

$$\mathcal{L}(y^T, \theta) = \int [\prod_{t=1}^{T} \int f(y_t|\epsilon^t, y^{t-1}, y_{20}, \theta) f(\epsilon^t|y^{t-1}, y_{20}, \theta) d\epsilon^t] f(y_{20}, \theta) dy_{20} \quad (71)$$

(71) is intractable.

• If we have L draws for y_{20} from $f(y_{20}, \theta)$ and L draws for $\epsilon^{t|t-1,l}$, $l = 1, \ldots, L, \quad t = 1, \ldots, T$, from $f(\epsilon^t|y^{t-1}, y_{20}, \theta)$ approximate (71) with

$$\mathcal{L}(y^T, \theta) = \frac{1}{L}[\prod_{t=1}^{T} \frac{1}{L} \sum_{l} f(y_t|\epsilon^{t|t-1,l}, y^{t-1}, y_{20}^{l}, \theta)] \tag{72}$$

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 (71) 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.