Simulation estimation

Fabio Canova
EUI and CEPR
September 2014

Outline

- Simulation estimators

- Simulated method of moments

- Indirect Inference/matching impulse responses.

- Examples.

- Identification problems.

References

Canova, F. and Marrinan, J. (1993) "Profit, Risk and Uncertainty in Exchange rates", *Journal of Monetary Economics*,32, 259-296.

Martin, V.and Pagan, A. (2001) in "Simulation Based estimation of some factor models in econometrics in Mariano, R. Schuermann and M. Weeks (eds.) *Inference Using Simulation Techniques*, Cambridge, Cambridge University Press.

Ingram, B. and Lee, B.S. (1989) "Estimation by Simulation of Time Series Models", *Journal of Econometrics*, 47, 197-207.

Ruge Murcia, F. (2007) "Methods to estimate DSGE models", it Journal of Economic Dynamics and Control, 31, 2599-2636.

Gourieroux, C., Monfort, A. and Renault, E. (1993), Indirect Inference, *Journal of Applied Econometrics*, 8, S85-S118.

Gourieroux, C. and Monfort, A. (1995) "Testing, Encompassing and Simulating Dynamic Econometric Models", *Econometric Theory*, 11, 195-228.

Dridi, R. , Guay, A. and Renault, E. (2007) "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models", *Journal of Econometrics*, 136, 397-430.

Hall, A., Inoue A., Nason, J. and Rossi, B. (2007), Information Criteria for Impulse Response Function Matching Estimation of DSGE Models, manuscript.

Jorda', O. (2005) "Estimation and Inference for Impulse Responses by Local Projections, *American Economic Review*, 95, 161-182.

Smith, A. (1993) "Estimating Nonlinear Time Series Models using Simulated Vector Autoregressions", *Journal of Applied Econometrics*, 8, 63-84.

Linde', J. (2005), "Estimating New Keynesian Phillips curve: A Full Information maximum likelihood", *Journal of Monetary Economics* , 52, 1135-1149.

Canova, F. and Sala, L. (2009) Back to square one: identification issues in DSGE models, *Journal of Monetary Economics* , 56(4), 431-449.

1 Introduction

• Simulation estimators are "distance" estimators similar to GMM.

• They can be used to estimate the structural parameters by simulation.

• Initially conceived for situations where GMM is not applicable. Now they have a very broad application.

Example 1.1 *A model with latent (hidden) variables.*

A social planner maximizes

$$\max_{\{c_t, N_t, K_{t+1}\}_{t=0}^{\infty}} E_0 \sum_t \beta^t u(c_t, N_t, \epsilon_{2t}) \tag{1}$$

$$c_t + K_{t+1} \leq f(K_t, N_t, \zeta_t) + (1-\delta)K_t \tag{2}$$

If $u(c_t, N_t, \epsilon_{2t}) = c_t^{\vartheta}(1-N_t)^{1-\vartheta}\epsilon_{2t}$*, the Euler equation is*

$$g_{\infty} = E_t[\beta \frac{c_{t+1}^{\vartheta-1}(1-N_{t+1})^{1-\vartheta}\epsilon_{2t+1}}{c_t^{\vartheta-1}(1-N_t)^{1-\vartheta}\epsilon_{2t}}[f_K + (1-\delta)] - 1] = 0 \tag{3}$$

where $g_t(y_t, \theta) = \beta \frac{c_{t+1}^{\vartheta-1}(1-N_{t+1})^{1-\vartheta}\epsilon_{2t+1}}{c_t^{\vartheta-1}(1-N_t)^{1-\vartheta}\epsilon_{2t}}[f_K + (1-\delta)]$.

Problem! Can't construct g_T*, since* ϵ_{2t} *is unobservable.*

- We had this problem also when estimating an RBC model (technology shocks are non observable) but we could find a proxy (Solow residuals) which is observable.

General result: if unmeasurable shocks (such as ϵ_{2t}) or unobservable variables (such as capital) enter the orthogonality conditions, GMM and GIV can not be used to estimate structural parameters.

- What to do then? Use simulation estimators.

- Suppose $E_t(g_t(y_t, \theta, \nu_t)) = 0$.

- Suppose ν_t is unobservable, but its distribution is known.

- Draw $\{\nu_t\}^l$ from such distribution, $l = 1, \ldots, L$.

- Construct $g_t^l = g(y_t, \theta, \{\nu_t\}^l)$ for each draw l.

- Under regularity conditions, if draws are iid, by the Law of Large Numbers (LLN) $\frac{1}{L}\sum_{l=1}^{L} g_t^l \xrightarrow{P} g(y_t, \theta, \nu_t)$.

If variables are unobserved but come from a known distribution, simulate them, construct g_t using simulated data, and apply GMM to "simulated" orthogonality conditions.

- What if the distribution of unobservable variables is unknown?

- If L is large, by LLN, it does not matter: we get asymptotic normality.

- If L is short, distribution matters; you need to be careful.

In the previous example with preference shocks:

- Draw $\{\epsilon_{2t}\}^l, l = 1, \ldots, L$ times from a normal distribution.

- Construct $[\beta \frac{c_{t+1}^{\vartheta-1}(1-N_{t+1})^{1-\vartheta}\epsilon_{2t+1}^l}{c_t^{\vartheta-1}(1-N_t)^{1-\vartheta}\epsilon_{2t}^l}[f_K + (1-\delta)] - 1]$ for each draw.

- Use $\frac{1}{T}\sum_t \{\frac{1}{L}\sum_l [\beta \frac{c_{t+1}^{\vartheta-1}(1-N_{t+1})^{1-\vartheta}\epsilon_{2t+1}^l}{c_t^{\vartheta-1}(1-N_t)^{1-\vartheta}\epsilon_{2t}^l}[f_K + (1-\delta)] - 1]\} = 0$ as your orthogonality condition

- Assuming that data for c_t, N_t, K_t are available, estimate β, ϑ, δ, etc.

Main difference between simulation and GMM estimators of orthogonality conditions is in the asymptotic covariance matrix. Now it is $(1 + \frac{1}{L})B^{-1}AB^{-1'} \geq B^{-1}AB^{-1'}$, since there is a simulation error to take into account. For large L, this error is negligible.

Economic models with latent variables

a) CAPM line $R_k = R_f + \beta_k(R_M - R_f)$; R_M unobservable market portfolio, interest is in β_k.

b) Fisher equation: $r_t = i_t - E_t\pi_{t+1}$ (ex-ante vs. ex-post), $E_t\pi_{t+1}$ unobservable, interest is in r_t.

• Simulation estimators are popular in microeconometrics: often there are unobservable reasons for certain choices (preferences) or truncated variables (e.g. some goods can be bought only in positive amounts).

• Simulation estimators can also be used when all variables are observable and with objective functions which are not the difference between orthogonality conditions (they are more general than GMM).

2 Generic Simulation Estimators

- $H_T(x)$ is a $J \times 1$ vector of functions of actual data $\{x_t\}_{t=1}^T$.

- $H_N(y(\theta, \nu))$ is the same $J \times 1$ vector of functions computed using the simulated data $\{y_i\}_{i=1}^N$, once the $k \times 1$ vector of parameters θ and a sequence of shocks ν are chosen.

Assume:

a)x_t and $y_i(\theta, \nu)$ are stationary and ergodic.

b) $H_T(x) \xrightarrow{P} \mu_x$, as $T \to \infty$ and $H_N(y(\theta, \nu)) \xrightarrow{P} \mu_y(\theta)$ as $N \to \infty$ (Consistency of the estimates of H in actual and simulated data).

Technical conditions:

c) Under the null that the model is true, there exists a unique θ^* such that $\mu_x = \mu_y(\theta^*)$ (Identifiability).

d) $H_N(y(\theta, \nu))$ is continuous in the mean.

Then:

$$\theta_{SE} = argmin[H_T(x) - H_N(y(\theta, \nu))]W_{TN}[H_T(x) - H_N(y(\theta, \nu))]'$$

where $W_{NT} \xrightarrow{P} W$ is a $J \times J$ symmetric matrix.

Under a)-d) θ_{SE} is consistent and asymptotically normal.

Intuition for the result:

If $g_T \equiv \frac{1}{T}\sum_{t=1}^{T}[h(x) - \frac{1}{TN}\sum_{i=1}^{TN} h(y_i(\theta))]$, and $W_{TN} = W_T$ we are back into GMM framework, so previous results apply.

Major advantage relative to GMM: g_T is now the difference between continuous functions of actual and simulated data - could be moments, autocorrelation functions, VAR coefficients, etc.

Many estimators are in this class. Two are of interest.

2.1 Simulated Methods of Moments (SMM)

H are moments of the actual and the simulated data. To find θ_{SE}:

i) Choose a θ^0 and a $\{\nu_t\}$, solve and simulate the model and calculate $H_N(y(\theta^0, \nu_t))$.

ii) Find θ^1_{SE} by minimizing: $[H_T(x) - H_N(y(\theta^0_{SE}, \nu))]W_{TN}[H_T(x) - H_N(y(\theta^0_{SE}, \nu))]'$.

iii) Solve and simulate the model and calculate $H_N(y(\theta^1_{SE}, \nu_t))$. Find θ^2_{SE} as in ii). Continue.

iv) If $||[H_T(x)-H_N(y(\theta^i_{SE}, \nu))]W_{TN}[H_T(x)-H_N(y(\theta^i_{SE}, \nu))]'-[H_T(x)-H_N(y(\theta^{i-1}_{SE}, \nu))]W_{TN}[H_T(x)-H_N(y(\theta^{i-1}_{SE}, \nu))]'|| < \iota$, or $||\theta^i_{SE}-\theta^{i-1}_{SE}|| < \iota$, or both, ι small, stop.

IMPORTANT: Must use the same $\{\nu_t\}$ sequence during the iterations; otherwise don't know if objective function changes because parameters change or because shocks change.

• If $W_{NT} = I$, θ_{SE} is consistent but inefficient.

• If you want to use an optimal W, insert between steps ii) and iii) of the algorithm $W_{NT}^i = S_{NT}^i(\omega = 0)$ where $S_{NT}^i(\omega = 0) = \sum_{\tau=-\infty}^{\infty} g_T(\theta_{SE}^i) g_{T-\tau}(\theta_{SE}^i)'$.

• To get standard errors use a Monte Carlo approach, i.e. repeat algorithm for different ν_t sequences, plot the histogram of the resulting θ_{SE} and compute standard errors from this distribution (typically difficult to get meaningful standard errors from the Hessian of the objective function).

• SMM can be used to select parameters for computational experiments. Difference is that we have standard errors for the parameters - and that the model is assumed to be true in the dimensions represented by H only.

Example 2.1 *Equity Premium Puzzle (Merha-Prescott (1985)).*

The interest is in $H_T(x) = [\bar{R}^f, \bar{EP}]$*. Can a RBC model reproduce these data moments? Standard approach: choose* $\theta_2 = (\mu, \sigma, \pi)$ *(parameters of the endowment process) using external information; choose* $\theta_1 = (\beta, \varphi)$ *(parameters of preferences) such that simulated* $H_N(y(\theta)) = [\bar{R}^f(\theta_1, \hat{\theta}_2), \bar{EP}(\bar{\theta}_1, \hat{\theta}_2)]$ *is as close as possible* $H_T(x)$*. A puzzle obtains because for* θ_1 *in a reasonable range* $(H_T(x) - H_N(y(\theta_1, \hat{\theta}_2))$ *is large.*

Can do this exercise formally with SMM:

a) Set $H_T(x) = [\bar{R}^f, \bar{EP}, \bar{PD}, var(R^f), var(EP), var(PD)]$*,* PD *is the price earning ratio. This is what the data gives you.*

b) Set $H_N(y(\theta)) = [\bar{R}^f(\theta), \bar{EP}(\theta), \bar{PD}(\theta), var(R^r(\theta),$ $var(EP(\theta)),$ $var(PD(\theta))]$*. This is what the model gives you, given* θ*.*

c) Choose $W^0_{NT} = I$*.*

d) Iteratively minimize $[H_T(x) - H_N(y(\theta^i_{SE}))]W^i_{TN}[H_T(x) - H_N(y(\theta^i_{SE}))]'$*.*

Recall that if the number of moments is the same as the number of parameters the choice of W does not matter.

2.2 Indirect Inference

Generalization of SMM, where H are continuous function (rather than moments) of the data.

• Data instrumental function: $H(y_t)$. An estimator is $H_T = \frac{1}{T}\sum_t h(y_t)$. Assume consistency: $P \lim H_T = E(h(y_t))$; P is the pdf of y_t.

• Model instrumental function: $H(y_i(\theta))$. An estimator is $H_N = \frac{1}{N}$ $\sum_i h(y_i(\theta))$. Assume consistency: $P_* \lim H_N = E_*(h(y_i(\theta)))$; P_* is the pdf of y_t, given θ.

• Technical conditions:

i) $\theta = [\theta_1, \theta_2]$; θ_2 are nuisance parameters (needed for simulations);

ii) $H_N(\theta_1, \theta_2)$ is a function (unique mapping between θ and H).

iii) There exist a true H^0;

iv) Encompassing: $H^0 = H(\theta_1^0, \bar{\theta}_2)$ for any estimator $\bar{\theta}_2$ of θ_2.

Then an Indirect Inference estimator (IIE) of θ is

$$\theta_{IIE} = arg \min_{\theta_1, \theta_2} [H_T - H_N(\theta_1, \theta_2)]' \Omega_T [H_T - H_N(\theta_1, \theta_2)] \qquad (4)$$

where $P_* \lim \Omega_T = \Omega$.

• Dridi, Guay, Renault (2007) give sufficient conditions and prove consistency and asymptotically normality of this estimator.

Example 2.2 *Suppose you run a regression with data on forward and spot exchange rates of the form*

$$S_{t+1} = a + bF_{t,t+1} + u_t \tag{5}$$

If uncovered interest parity is satisfied we should expect $a = 0, b = 1$*. In practice* $b \neq 1$ *and often negative.*

Suppose you have a model which has something to say about spot and forward rates. Suppose, given some vector of structural parameters θ*, you solve it and simulate data from it. Then you can run the following regression*

$$S^m_{t+1} = a(\theta) + b(\theta)F^m_{t,t+1} + u^m_t \tag{6}$$

where the superscript m *indicates simulated data.*

An indirect inference estimator of θ *is one which makes* $w_1(a - a(\theta)) + w_2(b - b(\theta))$ *as close as possible to zero.*

Special case of interest: $H(y_t)$ are structural impulse responses.

Example 2.3

$$x_t = \frac{h}{1+h} y_{t-1} + \frac{1}{1+h} E_t y_{t+1} + \frac{1}{\varphi}(i_t - E_t \pi_{t+1}) + v_{1t} \quad (7)$$

$$\pi_t = \frac{\omega}{1+\omega\beta} \pi_{t-1} + \frac{\beta}{1+\omega\beta} \pi_{t+1} + \frac{\varphi(1-\zeta\beta)(1-\zeta)}{(1+\omega\beta)\zeta} x_t + v_{2t} \quad (8)$$

$$i_t = \phi_r i_{t-1} + (1-\phi_r)(\phi_\pi \pi_{t-1} + \phi_x x_{t-1}) + v_{3t} \quad (9)$$

$h =$ *degree of habit persistence,* $\varphi =$ *relative risk aversion coefficient,* $\beta =$ *discount factor,* $\omega =$ *degree of indexation of prices,* $\zeta =$ *degree of price stickiness;* ϕ_r, ϕ_π, ϕ_x *are policy parameters;* v_{1t}, v_{2t} *are AR(1) with parameters* ρ_1, ρ_2*,* v_{3t} *is iid. Parameters* $\theta_2 = (\beta, \varphi, \zeta, \phi_r, \phi_\pi, \phi_x, \rho_1, \rho_2, h, \omega)$ *(The variances of the three shocks not identified from scaled impulse response).*

Set $H(y_t) = [IR(x_{t+k}|v_{3t}), IR(\pi_{t+k}|v_{3t}), IR(i_{t+k}|v_{3t})], \;\; k = 1, \ldots, 20.$

- Many arbitrary features: weighting matrix? Max number of IRF considered? Length of VAR?

Hall et al. (2007): criterion to optimally choose the maximum number of IRFs to be used in the exercise (call it p).

Idea: only "relevant" responses should be used, "redundant" ones should be purged (improve efficiency, reduce small sample biases).

• Let $p_2 > p_1$ and V_i be the covariance matrix of the structural parameters where $i = p_1, p_2$ Then $p_1 + 1, \ldots, p_2$ are redundant if $V_{p_2} = V_{p_1}$ (non-redundant if $V_{p_2} - V_{p_1}$ is positive semidefinite).

• p_0 is the horizon associated with the relevant IRF if (i) $p_0 \in (\underline{\text{p}}, \ldots)$ ($\underline{\text{h}}$ is the lower bound of admissible lengths); (ii) $V_{p_1} - V_{p_0}$ is positive semidefinite for $p_1 = p_0 + \Delta p$; (iii) $V_{p_0} = V_{\bar{p}}$ if $p_0 \leq \bar{p}$; ($\bar{p}$ is the upper bound of admissible lengths).

Algorithm 2.1 **1.** *Choose an upper* $\bar{p}$ *and a lower* $\underline{p}$ *and let* $p \in (\underline{p}, \bar{p})$*.*

2. *Estimate impulse responses in the data up to horizon* p*. Collect them into a column vector* $\hat{\gamma}_p$*.*

3. *Calculate theoretical impulse responses up to horizon* p*. Collect them into a column vector* $\gamma_p(\theta)$ *where* θ *are the structural parameters of the model.*

4. *Estimate* θ *using* $\hat{\theta}_p = arg\,\min(\hat{\gamma}_p - \gamma_p(\theta))' W_h (\hat{\gamma}_p - \gamma_p(\theta))$ *where* W_h *is a weighting matrix.*

5. *Compute* $V_p \equiv$ *cov*$(\hat{\theta}) = [\Gamma_p(\theta_0)' W_p \Gamma_p(\theta_0)]^{-1} \, [\Gamma_p(\theta_0)' W_p \Sigma_{\gamma_p} W_p \Gamma_p(\theta_0)]$ $[\Gamma_p(\theta_0)' W_p \Gamma_p(\theta_0)]^{-1}$ *where* $\Gamma_p(\theta) = \frac{\partial \gamma_p(\theta}{\partial \theta}$ *and* Σ_{γ_p} *is the covariance matrix of* $\hat{\gamma}_p$.

6. *Compute* $R(p) = \log(|V_p|) + p\frac{\log(T^{0.5})}{T^{0.5}}$ *if the model has a VAR(q) representation or* $R(p) = \log(|V_p|) + p\frac{\log(T^{0.5}/q)}{T^{0.5}/q}$ *if the model has a VAR(*∞*) representation.*

7. *Choose the* p *that minimizes* $R(p)$.

Aside: Calculation of IRFs by projection methods

Jorda (2005): compute responses using a sequence of VAR(q) models.

$$
\begin{aligned}
y_{t+1} &= B_{0,1} + B_{1,1} y_{t-1} + B_{2,1} y_{t-2} + \ldots + B_{q,1} y_{t-q} + u_{t+1} && (10)\\
y_{t+2} &= B_{0,2} + B_{1,2} y_{t-1} + B_{2,2} y_{t-2} + \ldots + B_{q,2} y_{t-q} + u_{t+2} && (11)\\
\vdots &= \vdots && (12)\\
y_{t+\tau} &= B_{0,\tau} + B_{1,\tau} y_{t-1} + B_{2,\tau} y_{t-2} + \ldots + B_{q,\tau} y_{t-q} + u_{t+\tau} && (13)
\end{aligned}
$$

The non-structural responses are $\hat{B}_{1,k},\ k = 1, \ldots, \tau$ and structural responses are $\hat{B}_{1,k} D,\ k = 1, \ldots, \tau$ where D is an identification matrix.

Call $\tilde{\gamma}_p$ the vector of estimated responses. Estimate θ using $\tilde{\theta}_p = arg \min(\tilde{\gamma}_p - \gamma_p(\theta))' W_p (\tilde{\gamma}_p - \gamma_p(\theta))$ where W_p is a weighting matrix.

Can apply Hall et al. (2007) approach to select optimal p.

3 Comparing estimators a NK Phillips curve

$$\pi_t = \beta E_t \pi_{t+1} + \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p} mc_t \tag{14}$$

where $mc_t = \frac{N_t w_t}{GDP_t}$ are real marginal costs, ζ_p is the probability of not changing prices, π_t is the inflation rate. Assume marginal costs are observable (or proxied by GDP gap).

Alternative way of writing this equation:

$$\pi_{t+1} = \frac{1}{\beta}\pi_t - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p \beta} mc_t + e_{t+1} \tag{15}$$

where $E_t(e_{t+1}) = 0$, i.e. e_t is an expectational error.

- GMM estimates of $\theta = (\beta, \zeta_p)$ are obtained using, for example,

$$\frac{1}{T}\sum_t [\pi_{t+1} - \frac{1}{\beta}\pi_t - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p \beta} mc_t]\pi_t = 0 \quad (16)$$

$$\frac{1}{T}\sum_t [\pi_{t+1} - \frac{1}{\beta}\pi_t - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p \beta} mc_t]\pi_{t-1} = 0 \quad (17)$$

$$\frac{1}{T}\sum_t [\pi_{t+1} - \frac{1}{\beta}\pi_t - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p \beta} mc_t]\pi_{t-2} = 0 \quad (18)$$

That is, by minimizing $(g_T(\theta) z_T) W_T (g_T(\theta) z_T)'$ by choice of θ, given $W_T \xrightarrow{P} W$, where $z_T = (\pi_T, \pi_{T-1}, \pi_{T-2})'$, $g_t = \pi_{t+1} - \frac{1}{\beta}\pi_t - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p \beta} mc_t$.

- SMM estimates (β, ζ_p) are obtained using, for example,

$$\frac{1}{T}\sum_t(\pi_{t+1}\pi_t) = \frac{1}{\beta}\frac{1}{T}\sum_t(\pi_t\pi_t) - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta}\frac{1}{T}\sum_t(mc_t\pi_t) \tag{19}$$

$$\begin{aligned}
\frac{1}{T}\sum_t(\pi_{t+1}\pi_{t-1}) &= \frac{1}{\beta}\frac{1}{T}\sum_t(\pi_t\pi_{t-1}) - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta}\frac{1}{T}\sum_t(mc_t\pi_{t-1}) \\
\frac{1}{T}\sum_t(\pi_{t+1}\pi_{t-2}) &= \frac{1}{\beta}\frac{1}{T}\sum_t(\pi_t\pi_{t-2}) - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta}\frac{1}{T}\sum_t(mc_t\pi_{t-2})
\end{aligned} \tag{20}$$

given that $\frac{1}{T}\sum_t e_{t+1}\pi_{t-\tau} = 0, \forall\tau > 0$. (Here we assume $N = T$.)

If $H_{x_T} = [\frac{1}{T}\sum_t(\pi_{t+1}\pi_t), \frac{1}{T}\sum_t(\pi_{t+1}\pi_{t-1}), \frac{1}{T}\sum_t(\pi_{t+1}\pi_{t-2})']'$ and

$H_{y_T}(\theta) = [\frac{1}{\beta}\frac{1}{T}\sum_t(\pi_t(\theta)\pi_t(\theta)) - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta}\frac{1}{T}\sum_t(mc_t\pi_t(\theta)),$

$\frac{1}{\beta}\frac{1}{T}\sum_t(\pi_t(\theta)\pi_{t-1}(\theta)) - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta}\frac{1}{T}\sum_t(mc_t\pi_{t-1}(\theta)),$

$\frac{1}{\beta}\frac{1}{T}\sum_t(\pi_t(\theta)\pi_{t-2}(\theta)) - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta}\frac{1}{T}\sum_t(mc_t\pi_{t-2}(\theta))]'$

estimates of θ are found minimizing $(H_{x_T} - H_{y_T}(\theta))W_T(H_{x_T} - H_{y_T}(\theta))'$, where again $W_T \xrightarrow{P} W$.

Difference with GMM is that the $\pi_{t-j}, j = 0, 1, 2$ entering $H_N(y(\theta))$ are simulated, given θ. Need to solve the model to be able to simulate the relevant data. Don't need this with GMM.

• Indirect inference estimates θ obtained using, e.g., the reduced form equation

$$\pi_{t+1} = b_{\pi}\pi_t - b_{gap} mc_t + e_{t+1} \tag{21}$$

and the structural equation

$$\pi_{t+1} = \frac{1}{\beta}\pi_t - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta} mc_t + e_{t+1} \tag{22}$$

and minimizing $H_T(\theta)W_T H_T'(\theta)$ by choice of θ where $H_T(\theta) = (b_{\pi} - \frac{1}{\beta}; b_{gap} - \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p\beta})'$ and, again, $W_T \xrightarrow{P} W$.

Need to solve the model to be able to simulate (22).

Since criterion functions are different, the weighting matrices are different, instruments may be different, there is no reason to expect the three procedures will give the same estimates for a given data set.

Table: Estimates of US NK Phillips curve

IV-GMM			SMM		
β	ζ_p	J-Test p-value	β	ζ_p	J-Test p-value
0.907 (10.35)	0.700 (5.08)	$\chi^2(5)$=0.15	0.999 (0.001)	0.999(0.014)	$\chi^2(1) = 0.00$

• GMM estimates obtained with constant and 2 lags of inflation and marginal costs. SMM estimates obtained by matching variance the first three autocovariances of inflation (numerical standard errors reported)

Table: Indirect Inference Estimates of US NK Phillips curve

	b_π	b_{gap}	β	ζ_p	criterion function
Actual	0.993 (0.05)	-0.04 (0.143)			
Simulated (actual gap)	0.996(0.0062)	0.032 (0.001)	0.752	0.481	0.01012
Simulated (simulated gap)	0.997(0.00008)	-0.004 (0.0006)	0.980	0.324	0.02321

• Indirect inference estimates obtained with two specifications one with actual marginal cost(MC); one where a process for the MC is estimated using an AR(2) and a constant on HP filtered data and then simulated together with inflation (standard errors are in parenthesis).

- Model roughly replicates actual b_π.

- Because b_{gap} is poorly estimated, simulations using the actual gap have hard time to produce the correct sign for this coefficient.

- Estimated ζ_p are very low (roughly, prices change every 1-2 quarters), β unreasonably low when the actual gap is used.

- Criterion function still not zero in both cases. Convergence problems? Model incorrect?

4 Identification issues

- Can we identify (and estimate) the parameters of a model?

- Can we get a good fit even though parameter estimates are wrong?

- Can we get wrong policy conclusions because of identification problems?

$$E_t[A(\theta)x_{t+1} + B(\theta)x_t + C(\theta)x_{t-1} + D(\theta)z_{t+1} + F(\theta)z_t] = 0$$

$$z_{t+1} = G(\theta)z_t + e_t$$

Stationary (log-linearized) RE solution:

$$x_t = J(\theta)x_{t-1} + K(\theta)e_t$$

$$z_t = G(\theta)z_{t-1} + e_t$$

Model responses to shock j: $x_{tj}^{M}(\theta) = C(\theta)(L)e_t^j$, $C(\theta)(L) = (I - J(\theta))^{-1}K(\theta)$ and L is the lag operator.

Data responses to shock j: $x_{tj} = W(L)e_t^j$.

$$\theta_{IIE} = \underset{\theta}{argmin}\ g(\theta) = \underset{\theta}{argmin}(x_{tj} - x_{tj}^{M}(\theta))'W(T)(x_{tj} - x_{tj}^{M}(\theta)).$$

• Can we recover the true θs? We need:

- $g(\theta)$ has a unique minimum at $\theta = \theta_0$

- Hessian of $g(\theta)$ is positive definite and has full rank.

- Curvature of $g(\theta)$ is "sufficient".

In DSGE, the distance function is non-linear function of θ; too complicated to work out conditions analytically $\rightarrow$ identifiability of θ could be problematic.

• Different objective functions (different g) may have different "identification power".

Potential Problems

- Observational equivalence: two or more models are consistent with the same empirical impulse responses.

- Under-identification: parameters may not enter impulse responses.

- Partial under-identification: two sets of parameters may enter impulse responses only proportionally.

- Weak identification: the objective function has a unique minimum but it is very flat in the neighborhood of the minimum.

Note: weak identification could be asymmetric. Also problems may emerge because only a subset of the model implications (impulse responses) are considered.

Example 1: Observational equivalence

1) $x_t = \frac{1}{\lambda_2+\lambda_1} E_t x_{t+1} + \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2} x_{t-1} + v_t,$ where: $\lambda_2 \geq 1 \geq \lambda_1 \geq 0$.

The RE (stable) solution is: $x_t = \lambda_1 x_{t-1} + \frac{\lambda_2+\lambda_1}{\lambda_2} v_t$

Given $v_t = 1$, the responses of x_t are $[\frac{\lambda_2+\lambda_1}{\lambda_2}, \lambda_1 \frac{\lambda_2+\lambda_1}{\lambda_2}, \lambda_1^2 \frac{\lambda_2+\lambda_1}{\lambda_2}, ...]$

Using at least two horizons, λ_1 and λ_2 can be estimated.

2) $y_t = \lambda_1 y_{t-1} + w_t$

y_t responses to an impulse in w_t are identical to x_t responses to an impulse in v_t if $\sigma_w = \frac{\lambda_2+\lambda_1}{\lambda_2} \sigma_v$.

3) $y_t = \frac{1}{\lambda_1} E_t y_{t+1}$ where $y_{t+1} = E_t y_{t+1} + w_t$ and w_t iid $(0, \sigma_w^2)$.

The RE (stable) solution is $y_t = \lambda_1 y_{t-1} + w_t$. If $\sigma_w = \frac{\lambda_2 + \lambda_1}{\lambda_2} \sigma_v$, the three processes are indistinguishable from impulse responses.

Beyer and Farmer (2004): models like

$$Ax_t + DE_t x_{t+1} = B_1 x_{t-1} + B_2 E_{t-1} x_t + Cv_t$$

also have a representation as in 3).

Other examples: Kim (2001, JEDC); Ma (2002, EL); Altig, et al. (2005); Ellison (2005).

Example 2: Under-identification

$$
\begin{aligned}
y_t &= a_1 E_t y_{t+1} + a_2(i_t - E_t \pi_{t+1}) + v_{1t} && (23) \\
\pi_t &= a_3 E_t \pi_{t+1} + a_4 y_t + v_{2t} && (24) \\
i_t &= a_5 E_t \pi_{t+1} + v_{3t} && (25)
\end{aligned}
$$

Solution:

$$
\begin{bmatrix} \hat{y}_t \\ \hat{\pi}_t \\ \hat{i}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_2 \\ a_4 & 1 & a_2 a_4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix}
$$

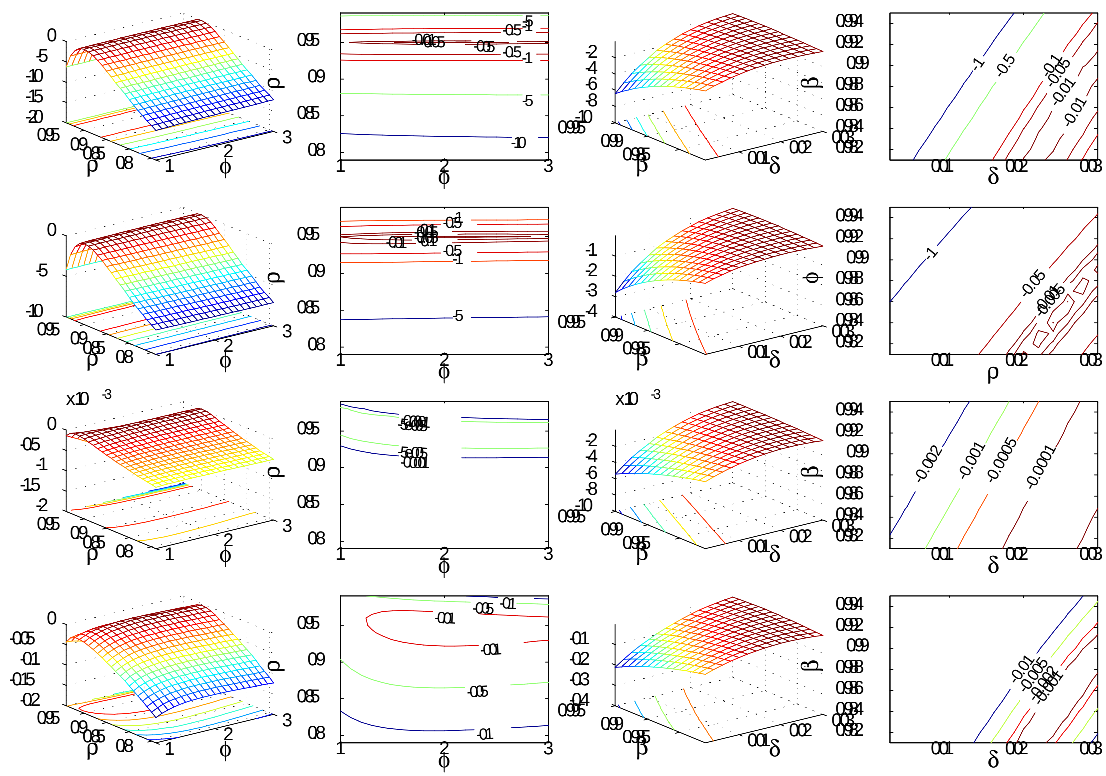
- a_1, a_3, a_5 disappear from the solution.
- Different shocks identify different parameters.
- Different variables identify different parameters.

Example 3: Weak and partial under-identification

$$\max \beta^t \sum_t \frac{c_t^{1-\phi}}{1-\phi}$$

$$c_t + k_{t+1} = k_t^{\eta} z_t + (1-\delta) k_t$$

Select $\beta = 0.985, \phi = 2.0, \rho = 0.95, \eta = 0.36, \delta = 0.025, z^{ss} = 1$. Simulate data. Study out how the population objective function change when two parameters around are changed using responses of capital, real wages, consumption and output to technology shocks.



Distance surface: Basic, Subset, Matching VAR and Weighted

What causes the problems?

Law of motion of capital stock in almost invariant to :

(a) variations of η and ρ (weak identification)

(b) variations of β and δ additive (partial under-identification)

Can we reduce problems by:

(i) Changing $W(T)$? (before $W(T) = I$, long horizon may have little information)

(ii) Matching VAR coefficients?

(iii) Altering the objective function?

Consequences of weak and partial identification:

- Remain stuck at initial conditions if algorithm is poor.

- Estimates could be random.

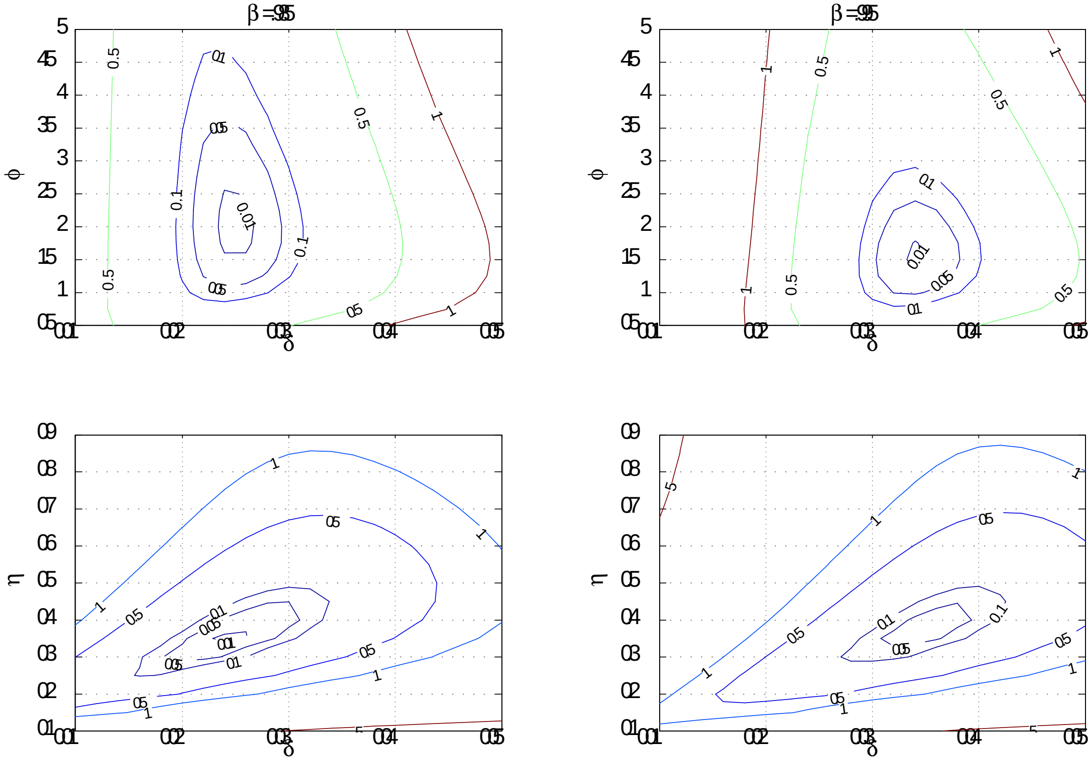
- Parameter estimates inconsistent, asymptotic distribution non-normal, standard t-tests incorrect (Choi and Phillips (1992), Stock and Wright (2003)).

Standard fixups:

- Multiply objective function by 10^{10} (OK for weak identification, does not do it for partial identification).

- Start from different initial conditions; take infimum of minimum (here infimum over all β is $\beta = 0.97$).

- Fix β (problem!).



Fixing beta

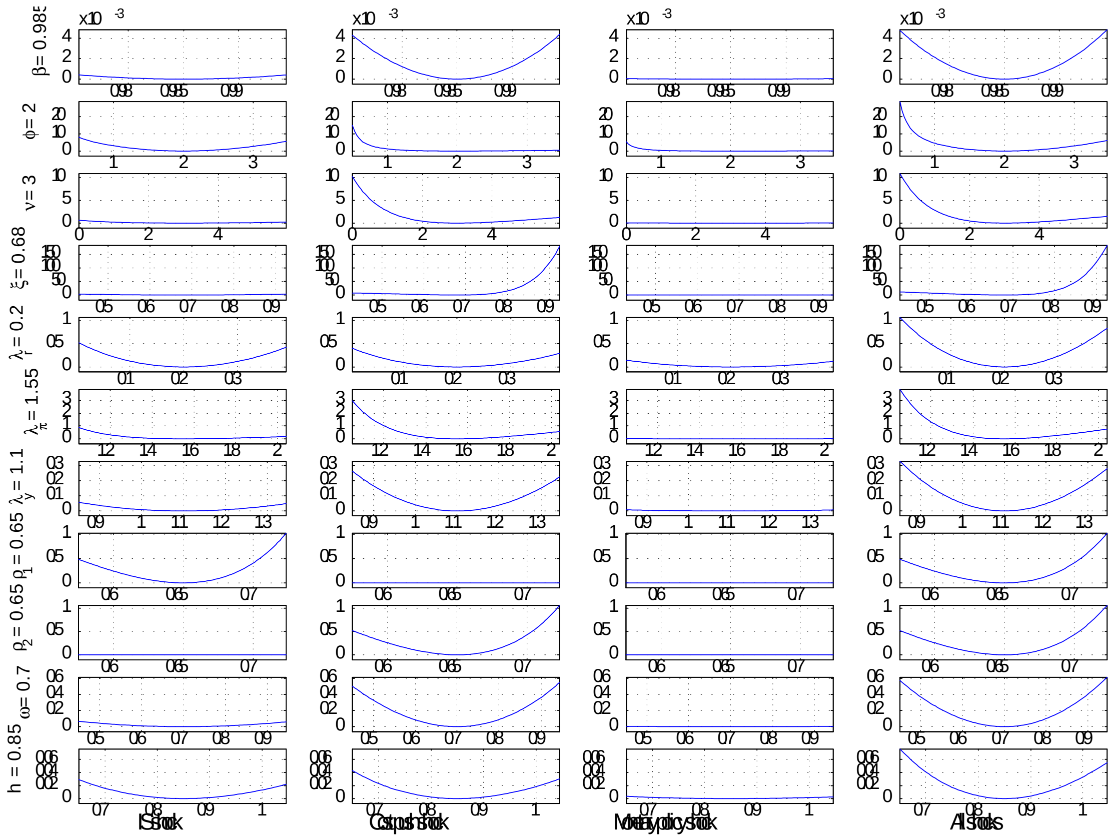
Identification and estimation

$$
\begin{aligned}
y_t &= \frac{h}{1+h} y_{t-1} + \frac{1}{1+h} E_t y_{t+1} + \frac{1}{\phi}(i_t - E_t \pi_{t+1}) + v_{1t} \\
\pi_t &= \frac{\omega}{1+\omega\beta} \pi_{t-1} + \frac{\beta}{1+\omega\beta} \pi_{t+1} + \frac{(\phi+\nu)(1-\zeta\beta)(1-\zeta)}{(1+\omega\beta)\zeta} y_t + v_{2t} \\
i_t &= \lambda_r i_{t-1} + (1-\lambda_r)(\lambda_\pi \pi_{t-1} + \lambda_y y_{t-1}) + v_{3t}
\end{aligned}
$$

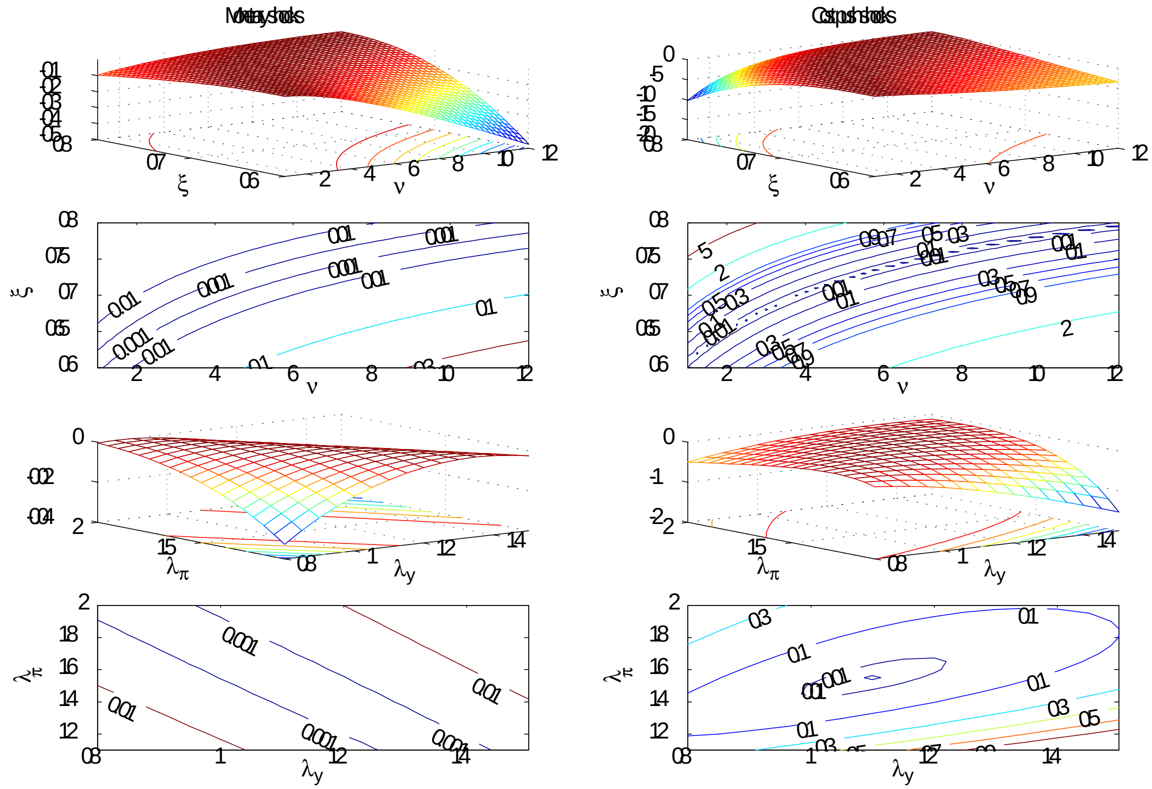
h: degree of habit persistence (.85); ν: inverse elasticity of labor supply (3); ϕ: relative risk aversion (2); β: discount factor (.985); ω: degree of price indexation (.25); ζ: degree of price stickiness (.68);

$\lambda_r, \lambda_\pi, \lambda_y$: policy parameters (.2, 1.55, 1.1);

v_{1t}: AR(ρ_1) (.65); v_{2t}: AR(ρ_2) (.65); v_{3t}: i.i.d.



Distance function



Distance function and contours plots

NK model: Matching monetary policy shocks

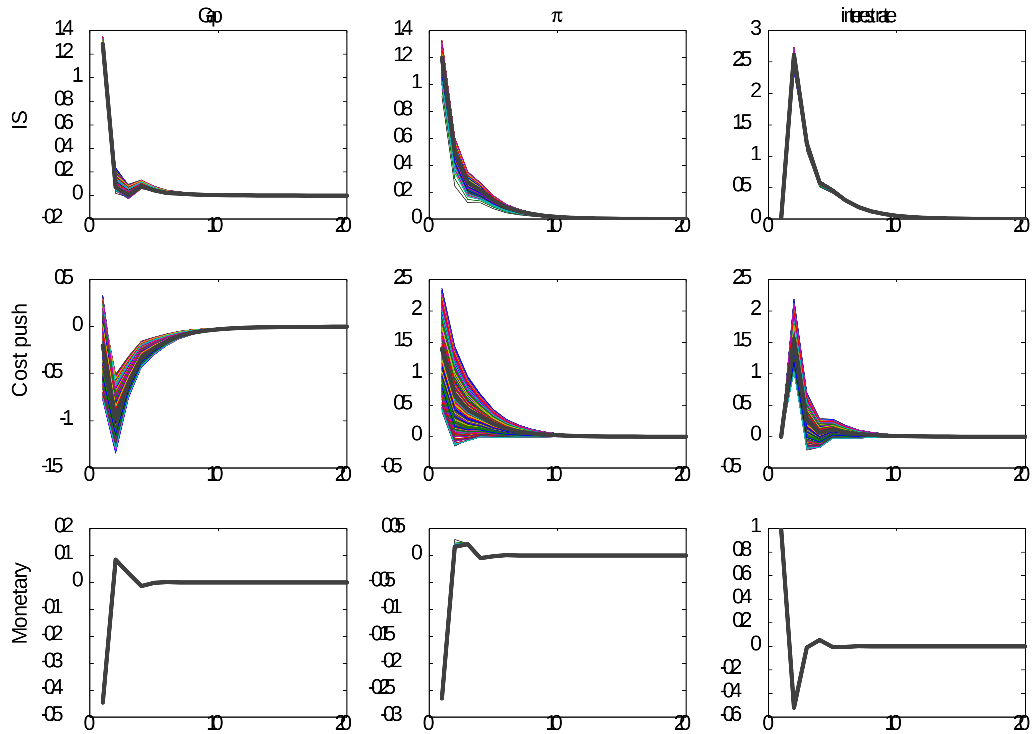
	True values	Population	T = 120	T = 200	T=1000	T=1000 wrong
β	.985	.987 (.003)	.98 (.007)	.98 (.006)	.98 (.007)	.999 (.008)
ϕ	2	2 (.003)	1.49 (2.878)	1.504 (1.906)	1.757 (.823)	10 (.420)
ν	3	4.082 (1.653)	4.184 (1.963)	4.269 (1.763)	4.517 (1.634)	1.421 (2.33)
ζ	.68	.702 (.038)	.644 (.156)	.641 (.112)	.621 (.071)	.998(.072)
λ_r	.2	.247 (.026)	.552 (.272)	.481 (.266)	.352 (.253)	.417 (.099)
λ_π	1.55	1.013 (.337)	1.058 (1.527)	1.107 (1.309)	1.345 (1.186)	3.607 (1.281)
λ_y	1.1	1.683 (.333)	4.304 (2.111)	2.924 (2.126)	1.498 (2.088)	2.59 (1.442)
ρ_1	.65	.5 (.212)	.5 (.209)	.5 (.212)	.5 (.167)	.5 (.188)
ρ_2	.65	.5 (.207)	.5 (.208)	.5 (.213)	.5 (.188)	.5 (.193)
ω	.25	.246 (.006)	1 (.360)	1 (.35)	1 (.306)	0 (.384)
h	.85	.844 (.006)	1 (.379)	1 (.321)	1 (.233)	0 (.166)

Standard errors in parenthesis.

- Population estimates differ from true ones.

- As $T \to \infty$ estimates do not converge to population or true ones.

- Standard errors do not decrease with sample size. They are random.



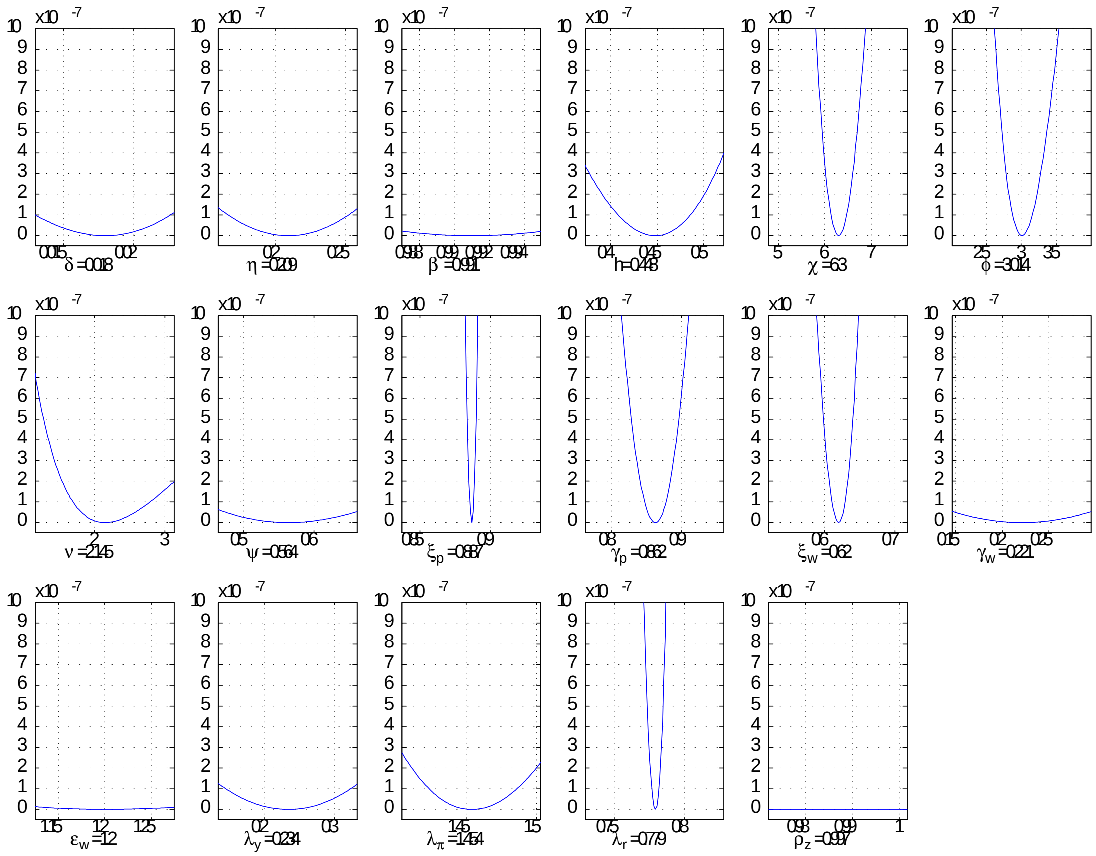
Impulse responses, Monetary Shocks, Population estimates

Think your model is great, but estimates far away from true ones!!

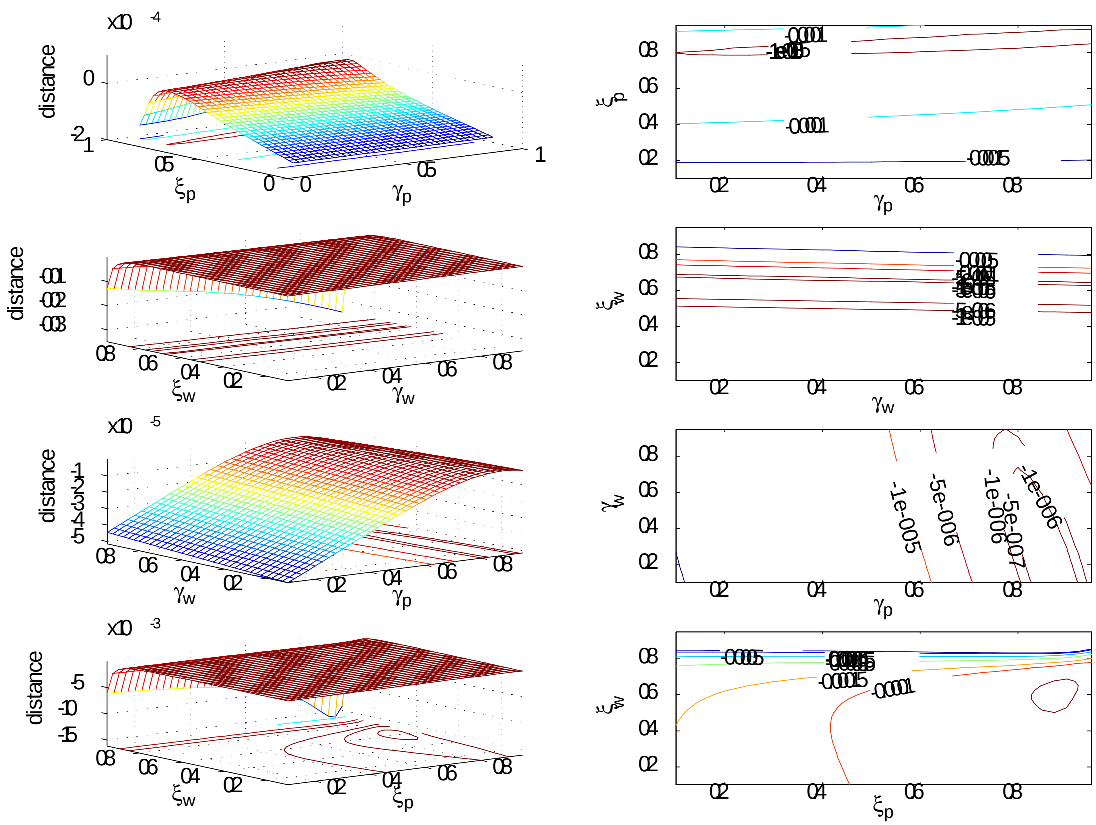
Wrong inference

$$
\begin{aligned}
0 &= -k_{t+1} + (1-\delta)k_t + \delta x_t \\
0 &= -u_t + \psi r_t \\
0 &= \frac{\eta\delta}{\bar{r}} x_t + (1 - \frac{\eta\delta}{\bar{r}})c_t - \eta k_t - (1-\eta)N_t - \eta u_t - ez_t \\
0 &= -R_t + \phi_r R_{t-1} + (1-\phi_r)(\phi_\pi \pi_t + \phi_y y_t) + er_t \\
0 &= -y_t + \eta k_t + (1-\eta)N_t + \eta u_t + ez_t \\
0 &= -N_t + k_t - w_t + (1+\psi)r_t \\
0 &= E_t[\frac{h}{1+h} c_{t+1} - c_t + \frac{h}{1+h} c_{t-1} - \frac{1-h}{(1+h)\varphi}(R_t - \pi_{t+1})] \\
0 &= E_t[\frac{\beta}{1+\beta} x_{t+1} - x_t + \frac{1}{1+\beta} x_{t-1} + \frac{\chi^{-1}}{1+\beta} q_t + \frac{\beta}{1+\beta} ex_{t+1} - \frac{1}{1+\beta} ex_t] \\
0 &= E_t[\pi_{t+1} - R_t - q_t + \beta(1-\delta)q_{t+1} + \beta\bar{r}r_{t+1}] \\
0 &= E_t[\frac{\beta}{1+\beta\gamma_p}\pi_{t+1} - \pi_t + \frac{\gamma_p}{1+\beta\gamma_p}\pi_{t-1} + T_p(\eta r_t + (1-\eta)w_t - ez_t + ep_t)] \\
0 &= E_t[\frac{\beta}{1+\beta\gamma_p} w_{t+1} - w_t + \frac{1}{1+\beta} w_{t-1} + \frac{\beta}{1+\beta}\pi_{t+1} - \\
&\quad \frac{1+\beta\gamma_w}{1+\beta}\pi_t + \frac{\gamma_w}{1+\beta\gamma_w}{}_{t-1}(w_t - \sigma N_t - \frac{\varphi}{1-h}(c_t - hc_{t-1}) - ew_t)]
\end{aligned}
$$

δ	depreciation rate (.0182)	λ_w	wage markup (1.2)
ψ	parameter (.564)	$\bar{\pi}$	steady state π (1.016)
η	share of capital (.209)	h	habit persistence (.448)
φ	risk aversion (3.014)	σ_l	inverse el. of N^s (2.145)
β	discount factor (.991)	χ^{-1}	inv. el. to Tobin's q (.15)
ζ_p	price stickiness (.887)	ζ_w	wage stickiness (.62)
γ_p	price indexation (.862)	γ_w	wage indexation (.221)
ϕ_y	response to y (.234)	ϕ_π	response to π (1.454)
ϕ_r	int. rate smoothing (.779)		
$T_p \equiv$	$\frac{(1-\beta\zeta_p)(1-\zeta_p)}{(1+\beta\gamma_p)\zeta_p}$		
$T_w \equiv$	$\frac{(1-\beta\zeta_w)(1-\zeta_w)}{(1+\beta)(1+(1+\lambda_w)\sigma_l\lambda_w^{-1})\zeta_w}$		



Objective function: monetary and technology shocks



Distance surface and Contours Plots

	ζ_p	γ_p	ζ_w	γ_w	Obj.Fun.
Baseline	**0.887**	**0.862**	**0.62**	**0.221**	
x0 = lb + 1std	0.8944	0.8251	0.615	0	1.8235E-07
x0 = lb + 2std	0.8924	0.7768	0.6095	0.1005	3.75E-07
x0 = ub - 1std	0.882	0.7957	0.6062	0.1316	2.43E-07
x0 = ub - 2std	0.9044	0.7701	0.6301	0	8.72E-07
Case 1	**0**	**0.862**	**0.62**	**0.221**	
x0 = lb + 1std	0.1304	0.0038	0.6401	0.245	2.7278E-08
x0 = lb + 2std	0.1015	0.0853	0.6065	0.1791	4.84E-08
x0 = ub - 1std	0.0701	0.1304	0.6128	0.1979	4.72E-08
x0 = ub - 2std	0.0922	0.0749	0.618	0.215	3.05E-08
Case 2	**0**	**0**	**0.62**	**0.221**	
x0 = lb + 1std	0.1396	0.0072	0.6392	0.2436	3.1902E-08
x0 = lb + 2std	0.0838	0.1193	0.6044	0.1683	4.38E-08
x0 = ub - 1std	0.0539	0.1773	0.6006	0.1575	5.51E-08
x0 = ub - 2std	0.0789	0.0971	0.6114	0.1835	2.61E-08
Case 3	**0**	**0.862**	**0.62**	**0**	
x0 = lb + 1std	0.0248	0	0.6273	0.029	7.437E-09
x0 = lb + 2std	0.4649	0	0.7443	0.4668	2.10E-06
x0 = ub - 1std	0.0652	0.0004	0.6147	0.0447	7.13E-08
x0 = ub - 2std	0.6463	0.2673	0.8222	0.3811	5.56E-06

	ζ_p	γ_p	ζ_w	γ_w	Obj.Fun.
Case 4	**0.887**	**0**	**0.62**	**0.8**	
x0 = lb + 1std	0.9264	0.3701	0.637	0.4919	3.5156E-07
x0 = lb + 2std	0.9076	0.2268	0.6415	0.154	3.51E-07
x0 = ub - 1std	0.9014	0.3945	0.6477	0	6.12E-07
x0 = ub - 2std	0.9263	0.3133	0.6294	0.4252	4.13E-07
Case 5	**0.887**	**0**	**0**	**0.221**	
x0 = lb + 1std	0.9186	0.3536	0.0023	0	4.7877E-07
x0 = lb + 2std	0.8994	0.234	0	0	3.06E-07
x0 = ub - 1std	0.905	0.3494	0.0021	0	4.14E-07
x0 = ub - 2std	0.9343	0.5409	0.0042	0	9.64E-07
Case 6	**0.887**	**0**	**0**	**0.221**	
x0 = lb + 1std	0.877	0.0123	0.0229	0	2.4547E-06
x0 = lb + 2std	0.8919	0.0411	0.0003	0	4.26E-07
x0 = ub - 1std	0.907	0.2056	0.001	0.0001	6.58E-07
x0 = ub - 2std	0.8839	0.0499	0.0189	0	2.46E-06
Case 7	**0.887**	**0**	**0**	**0.221**	
x0 = lb + 1std	0.9056	0.2747	0.0154	0.25	1.60E-06
x0 = lb + 2std	0.9052	0.2805	0	0.25	2.41E-07
x0 = ub - 1std	0.9061	0.3669	0.0003	0.25	4.26E-07
x0 = ub - 2std	0.8985	0.194	0.001	0.25	2.07E-07

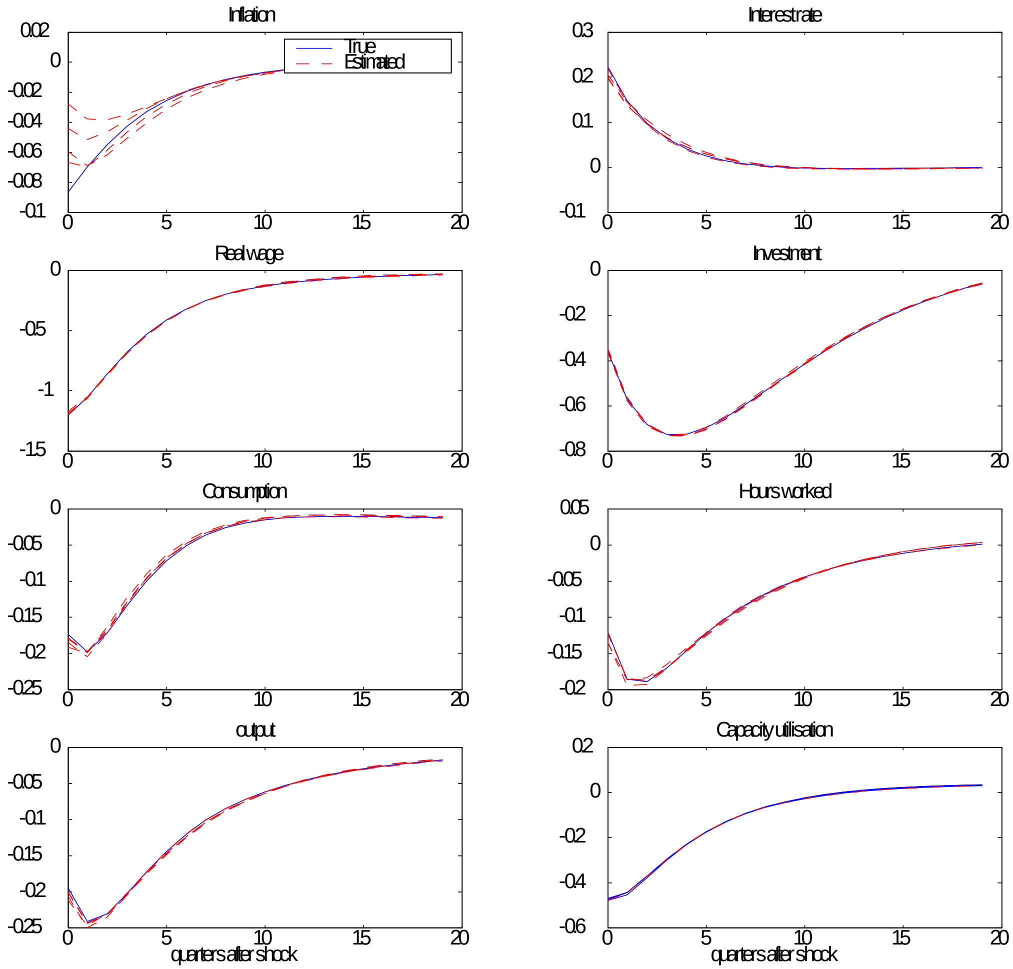


Figure 8: Impulse responses, Case 4.

Detecting identification problems

a) Ex-post diagnostics:

- Erratic parameter estimates as T increases.

- Large or non-computable standard errors. Crazy t-statistics.

b) General Diagnostics:

- Plots objective function (around calibrated values).

- Check the condition number of the Hessian (ratio of the largest to the smallest eigenvalue).

c) Tests:

- Cragg and Donald (1997): Testing rank of Hessian. Under regularity conditions: $(vec(\hat{H}) - vec(H))'\Omega(vec(\hat{H}) - vec(H)) \sim \chi^2((N - L_0)(N - L_0))$ $N = dim(H), L_0 =$rank of H.

- Anderson (1984): Size of characteristic roots of Hessian. Under regularity conditions: $\frac{\sum_{i=1}^{N-m} \hat{\lambda}_i}{\sum_{i=1}^{N} \hat{\lambda}_i} \xrightarrow{D}$ Normal distribution.

Applied to the last model: rank of $H = 6$; sum of 12-13 characteristics roots is smaller than 0.01 of the average root $\rightarrow$ 12-13 dimensions of weak or partial identification problems.

Which are the parameters is causing problems?

$\beta, h, \sigma_l, \delta, \eta, \psi, \gamma_p, \gamma_w, \lambda_w, \phi_\pi, \phi_y, \rho_z.$

Why? Variations of these parameters hardly affect law of motion of states!

Almost a rule: **For identification need states of the model to change substantially when structural parameters are changed**.

5 Exercises

Exercise 1: Consider the CAPM line $R_k = R_f + \beta_k(R_M - R_f)$; where R_k is the return on an a particular asset, R_M is the unobservable market portfolio return and R_f is the return on the risk free rate. Consider US data, use for R_f the ex-post real rate (i.e. $R_f = i - \pi$) and for R_k the return on Dow Jones 30 (DJ30). Estimate the slope of the relationship β_k by simulation.

Exercise 2: Consider the equity premium puzzle. Let $H_T(x) = [\bar{R}^f, \bar{EP}, \bar{PD}, var(R^f),$ $var(EP), var(PD)]$, where R_f is the risk free rate, $EP = R - R_f$, where R is the return on stocks, PD is the price earning ratio, $\bar{x}$ indicates the mean of x and $var(x)$ the variance of x. Using a basic RBC model with labor-leisure choice, utility of the form $\frac{c_t^{1-\phi}}{1-\phi} - \log(1 - N_t)$, AR(1) technology shocks and no adjustment costs to capital, estimate the parameters of the model $\beta, \phi, \rho_z, \sigma_z$ by simulation (Hint: add back steady states to the solution before simulating. Need also to add some steady state parameters to the set of parameters to be estimated).

Exercise 3: In the setup of example 2.2 consider matching 20 responses of output and inflation only to monetary shocks using US data. How different results are from those obtained matching 20 responses of all variables to monetary shocks?