Structural Macroeconometrics

Chapter VI
Identification

Johannes Pfeifer

Kobe University

Summer 2018

Outline

1. The Problem
2. Local Identification
3. Identification Strength

Motivation

- We estimated parameters θ by evaluating the posterior $\mathcal{L}(\theta|y^T)$
- But we never formally verified whether all elements of θ are properly identified

Simple Example

- Consider simple monetary model without price rigidities and real shocks
- Real interest r is fixed at steady-state level
- Inflation π_t is determined via Fisher equation and central bank reaction function that determines nominal interest rate i_t:

$$i_t = r + E_t\pi_{t+1} \tag{1}$$
$$i_t = r + \phi\pi_t + \nu_t \tag{2}$$

- The two equations imply

$$E_t\pi_{t+1} = \phi\pi_t + \nu_t \tag{3}$$

- We need $\phi > 1$ (**Taylor-principle**) for Blanchard-Kahn conditions to be satisfied
- Otherwise: indeterminacy

Solution and Estimation

- Assume that the error term is serially correlated (interest smoothing, policy mistakes, etc.)
$$\nu_t = \rho\nu_{t-1} + \varepsilon_t \tag{4}$$
- Iterate equation (3) forward:
$$\begin{aligned}\pi_t &= -\sum\nolimits_{j=0}^{\infty}\left(\frac{1}{\phi}\right)^{j+1} E_t\nu_{t+j} = -\frac{1}{\phi-\rho}\nu_t \\ &= -\frac{1}{\phi-\rho}\left(\rho\nu_{t-1}+\varepsilon_t\right) = \rho\pi_{t-1} - \frac{\varepsilon_t}{\phi-\rho} = \rho\pi_{t-1} + w_t\end{aligned} \tag{5}$$
where w_t is the new error term (ε_t and ν_t not observed)
- In equilibrium, the nominal interest rate is given by
$$i_t = r + E_t\pi_{t+1} = r + \rho\pi_t \tag{6}$$
- Regression of (2) will not estimate ϕ, but autocorrelation ρ
- Reason: inflation and policy shock ν_t are perfectly correlated to assure unique stable solution
- Central bank reaction describes off-equilibrium behavior, it cannot be identified from equilibrium observations!

The Likelihood Function

- Remember: Likelihood function is a **sufficient statistic**
- The likelihood function in this case is given by

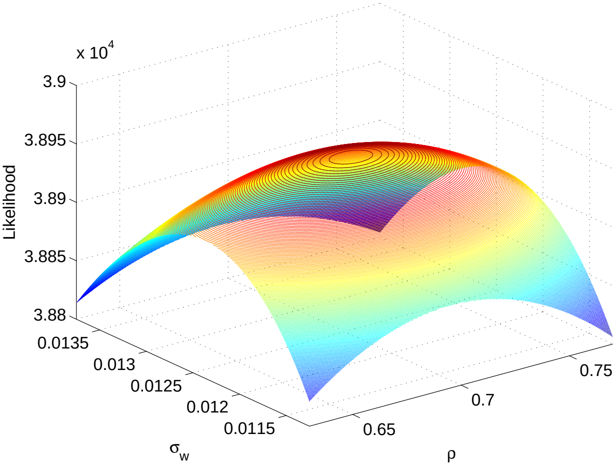
$$\log\left(f\left(\pi_t|\pi^{t-1}\right)\right) = -\frac{1}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left(\sigma_{w_t}^2\right) - \frac{1}{2}\frac{\left(\pi_t - \rho\pi_{t-1}\right)^2}{\sigma_{w_t}^2} \quad (7)$$

where

$$\sigma_{w_t}^2 = \left(\frac{\sigma_\varepsilon}{\phi - \rho}\right)^2 \quad (8)$$

- ρ and σ_{w_t} enter the likelihood function independently from each other and can be estimated

Likelihood Plot for ρ and σ_{w_t}



The likelihood function shows a clear peak in ρ and σ_{w_t}

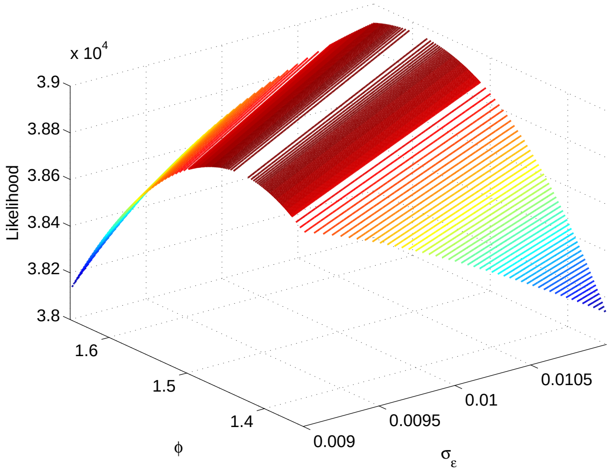
Estimating the Components of σ_{w_t}

- But: σ_ε and ϕ enter multiplicatively and cannot be disentangled

$$
\begin{aligned}
\frac{\partial \log\left(f\left(\pi_t|\pi^{t-1}\right)\right)}{\partial \sigma_\varepsilon} &= \frac{\partial \log\left(f\left(\pi_t|\pi^{t-1}\right)\right)}{\partial \sigma_{w_t}} \frac{\partial \sigma_{w_t}}{\partial \sigma_\varepsilon} \\
&= \frac{\partial \log\left(f\left(\pi_t|\pi^{t-1}\right)\right)}{\partial \sigma_{w_t}} \frac{1}{\phi - \rho} \qquad (9) \\
\frac{\partial \log\left(f\left(\pi_t|\pi^{t-1}\right)\right)}{\partial \phi} &= \frac{\partial \log\left(f\left(\pi_t|\pi^{t-1}\right)\right)}{\partial \sigma_{w_t}} \frac{\partial \sigma_{w_t}}{\partial \phi} \\
&= \frac{\partial \log\left(f\left(\pi_t|\pi^{t-1}\right)\right)}{\partial \sigma_{w_t}} \frac{\sigma_\varepsilon}{(\phi - \rho)^2} (-1) \\
&\overset{(9)}{=} -\frac{\sigma_\varepsilon}{\phi - \rho} \frac{\partial \log\left(f\left(\pi_t|\pi^{t-1}\right)\right)}{\partial \sigma_\varepsilon} \qquad (10)
\end{aligned}
$$

- The partial derivatives are linearly dependent and the Jacobian is thus singular

Likelihood Plot for σ_ε and ϕ



The likelihood function has no unique maximum in σ_ε and ϕ, only a range of maxima

Identification Problems

1. **Observational Equivalence**: mapping between structural parameters and objective function has no unique maximum

$\Rightarrow$ structural models with potentially different economic interpretations may be indistinguishable

2. **Under-identification**: objective function is independent of certain structural parameters, e.g. because they disappear from rational expectations solution

$\Rightarrow$ **partial identification** with two structural parameters entering objective function only proportionally, making them separately unrecoverable, is a special case

3. **weak identification**: parameter theoretically identified, but curvature may be small in certain regions of the parameter space

State Space System: Transition Equation

- Is there a systematic way to detect such issues for elements of θ?
- Consider generic log-linearized model in state-space form where $\hat{x}_t = \log(x_t) - \log(x)$

$$\hat{x}_{t+1} = F(\theta)\hat{x}_t + w_{t+1},\ w_{t+1} \overset{iid}{\sim} \mathcal{N}\left(0, Q(\theta)\right) \tag{11}$$

with

$$Q = CC'$$

- Define the vector

$$\tau = \left[\tau_x', \tau_F', \tau_\Omega'\right] \tag{12}$$

as the collection of non-constant elements of $\log(x), vec(F)$, and $vech(Q)$, respectively, that depend on θ

- Drops, e.g., steady states that are always zero or calibrated values
- The unconditional first and second moments are given by:

$$E\left(\hat{x}_t\right) = 0$$
$$E\left(\hat{x}_t \hat{x}_t'\right) \equiv \Sigma_x(0) = F\Sigma_x F' + Q$$

State Space System: Observation Equation

- The observables are given by

$$y_t = H(\theta)x + H(\theta)\hat{x}_t + \nu_t,\ \nu_t \overset{iid}{\sim} N\left(0, R(\theta)\right) \tag{13}$$

- In general, only observing y_t would be insufficient to fully characterize the distribution of x_t
- Fortunately, our model implies restrictions through θ

The Likelihood Function

- The unconditional first and second moments of the observables y_t are given by:

$$E(y_t) \equiv \mu_Y = H(\theta)\, x$$
$$\operatorname{cov}(y_{t+i}, y_t) \equiv \Sigma_Y(i) = \begin{cases} C\Sigma_x(0)\, C' \text{ if } i = 0 \\ CF^i\Sigma_x(0)\, C' \text{ if } i > 0 \end{cases}$$

- Stack the unique elements of $\Sigma_Y(i)$, $i = \{0, \ldots, T-1\}$ into a vector:

$$\sigma_Y = [vech(\Sigma_Y(0)), vec(\Sigma_Y(1)), \ldots, vec(\Sigma_Y(T-1))] \tag{14}$$

and define the vector of first and second moments as

$$m_T = [\mu_Y; \sigma_Y] \tag{15}$$

Definitions

Definition 1 (Global Identification)

Let $\theta \subseteq \Theta \subset \mathcal{R}^k$ be the parameter vector of interest and suppose inference about it is made based on T observations of a random vector y with known joint probability density function $f(Y^T|\theta)$. A point $\theta_0 \in \Theta$ is **globally identified** if

$$f\left(Y^T|\tilde{\theta}\right) = f\left(Y^T|\theta_0\right) \text{ with probability 1 } \Rightarrow \tilde{\theta} = \theta_0 \qquad (16)$$

for any $\tilde{\theta} \in \Theta$

Definition 2 (Local Identification)

If Equation (16) only holds in an open neighborhood of θ_0, then θ_0 is **locally identified**

Which Moments?

- Typically, the distribution is unknown or assumed to be normal
- People often use the first two moments
- Higher order moments might provide additional information (see e.g. Fernández-Villaverde and Rubio-Ramírez 2007)
- Thus, identification based on mean and (co)-variances is sufficient, but not necessary

Identification Condition

Theorem 3 (Identification Condition)

*Suppose our model in state-space form with θ_0 is the data generating process for Y^T. Then θ_0 is **globally identified** if*

$$m_T\left(\tilde{\theta}\right) = m_T\left(\theta_0\right) \Leftrightarrow \tilde{\theta} = \theta_0 \tag{17}$$

for any $\tilde{\theta} \in \Theta$ (and correspondingly for local identification).

- Requires mapping from population moments m_T to θ to be unique
- If not satisfied, there exists parameter values that imply the same joint distribution

$\Rightarrow$ Even with infinitely many observations, the true value could not be identified

- Checking global identification is hard $\Rightarrow$ check local identification for relevant parameter range

Identification Condition

Theorem 4 (Identification Condition)

Suppose m_T is continuously differentiable function of θ. Then θ_0 is locally identifiable if the Jacobian matrix

$$J(q) \equiv \frac{\partial m_q}{\partial \theta'} \tag{18}$$

has full column rank at θ_0 for $q \leq T$

Proof.

Follows from implicit function theorem. □

- We need at least as many moments as parameters
- For sufficient and necessary conditions see the paper
- One would have to check this condition everywhere $\Rightarrow$ impossible
- Nevertheless helpful in detecting observational equivalence (column of zeros) and underidentification (linear dependence of columns)

Computing the Jacobian

- Numerical differentiation and checking for singularity of the numerical Jacobian is prone to numerical errors when functions are nonlinear

$\Rightarrow$ Use analytical derivatives

- Using the chain rule, the Jacobian in equation (18) can be written as

$$J(T) \equiv \underbrace{\frac{\partial m_T}{\partial \tau'}}_{J_1(T)} \underbrace{\frac{\partial \tau}{\partial \theta'}}_{J_2(T)} \tag{19}$$

- Iskrev (2010) details how the two parts of the Jacobian can be computed, but we take Dynare's implementation
- Full rank is often achieved for $q \ll T$: Dynare starts at $q = 5$ and increases it if necessary

Computing the Jacobian

$$J(T) \equiv \underbrace{\frac{\partial m_T}{\partial \tau'}}_{J_1(T)} \underbrace{\frac{\partial \tau}{\partial \theta'}}_{J_2(T)} \tag{19}$$

- $J_2(T)$ shows how the parameters θ affect the non-constant model solution parts τ
- $J_1(T)$ shows how the model solution maps into the observed data moments

Corollary 5

The point θ_0 is locally identifiable only if $J_2(T)$ has full rank

- Necessary condition as parameters only affect distribution of observables through their effect on model solution
- It is not sufficient unless all states are observed
- $J_1(T)$ provides information on whether a theoretically identifiable parameter can be identified using the observables contained in y

Dynare Output for the Cochrane Toy Model

```
==== Identification analysis ====

Testing ML Starting value
Evaluating simulated moment uncertainty ... please wait
Doing 100 replicas of length 300 periods.
Simulated moment uncertainty ... done!

All parameters are identified in the model (rank of H).


WARNING !!!
The rank of J (moments) is deficient!


    [sig_nu,phi] are PAIRWISE collinear (with tol = 1.e-10) !
```

- All parameters are identified in the model
- However, without observing ν_t, the elements of σ_w cannot be disentangled

Dynare Output for Our RBC Model

```
==== Identification analysis ====

Testing prior mean

All parameters are identified in the model (rank of H).


All parameters are identified by J moments (rank of J)
```

- All parameters are identified in the model and using our observables

The Information Matrix Revisited

- The precision of the parameter estimates at the mode was given (asymptotically) by the inverse of the **Fisher Information Matrix**

$$I(\theta) = E\left[\left(\frac{\partial \log(p(Y|\theta))}{\partial \theta'}\right)'\left(\frac{\partial \log(p(Y|\theta))}{\partial \theta'}\right)\right] \tag{20}$$

- Typically, non-singularity of this matrix is sufficient for local identification (for details see Iskrev 2011)
- Can be decomposed into a variance matrix Δ and a correlation matrix $\tilde{I}(\theta)$

$$I(\theta) = \Delta^{\frac{1}{2}}\tilde{I}(\theta)\Delta^{\frac{1}{2}} \tag{21}$$

where

$$\Delta = \text{diag}(I(\theta))$$

Identification Strength

$$I(\theta) = \Delta^{\frac{1}{2}} \tilde{I}(\theta) \Delta^{\frac{1}{2}} \tag{21}$$

- Here, we can again see our two reasons for non-identifiability:
  1. The likelihood does not change when parameter i changes, i.e.

$$\Delta_i = 0 \tag{22}$$

  2. The effect on the likelihood is offset due to perfect correlation, i.e.

$$\rho_i \equiv corr\left(\frac{\partial \log(p(Y|\theta))}{\partial \theta_i}, \frac{\partial \log(p(Y|\theta))}{\partial \theta_{-i}}\right) = 1 \tag{23}$$

- This suggests that we can use this curvature information also for **weak identification** as in that case $\Delta_i \approx 0$ and/or $\rho_i \approx 1$

Identification Strength

- Identification strength can be measured using

$$s_i(\theta) = \sqrt{\Delta_i \left(1 - \rho_i^2\right)} \tag{24}$$

- If $I(\theta)$ is not singular, $1/s_i(\theta)$ is equal to the square root of the ith diagonal element of $(I(\theta))^{-1}$, i.e.

$$s_i(\theta) = \sqrt{\frac{1}{\left((I(\theta))^{-1}\right)_{(ii)}}} \tag{25}$$

- As this measure only uses the population objective function (expected log-likelihood), it is an a-priori measure
- At the same time, it can be interpreted as the lower bound on the estimation uncertainty in an unbiased finite sample estimator

Identification Strength

- Problem: measure is percentage change in likelihood for unit change in parameter $\rightarrow$ not unit free
- Solution: multiply with value of θ_i to obtain elasticity measure:

$$s_i^r(\theta) = \sqrt{\theta_i \Delta_i} \times \sqrt{(1 - \rho_i^2)} \tag{26}$$

where the first part is the elasticity of the likelihood function w.r.t. θ_i, keeping all other parameters constant

Dynare Implementation

- The strength of identification for parameter θ_i is defined through a renormalized version of the curvature as

$$s_i^{dyn} = \theta_i \times s_i(\theta) \overset{(25)}{=} \sqrt{\frac{\theta_i^2}{\left(I(\theta)^{-1}\right)_{(i,i)}}}$$

- Can be interpreted as an "a-priori t-test"
- Taking square of θ assures it to be positive
- The **sensitivity component** contained in this measure (as opposed to the **correlation component**) is defined as

$$\tilde{\Delta}_i = \sqrt{\theta_i^2 \Delta_i} = \sqrt{\theta_i^2 I(\theta)_{(i,i)}}$$

- Identification is often checked over the prior region, e.g. at the prior mean

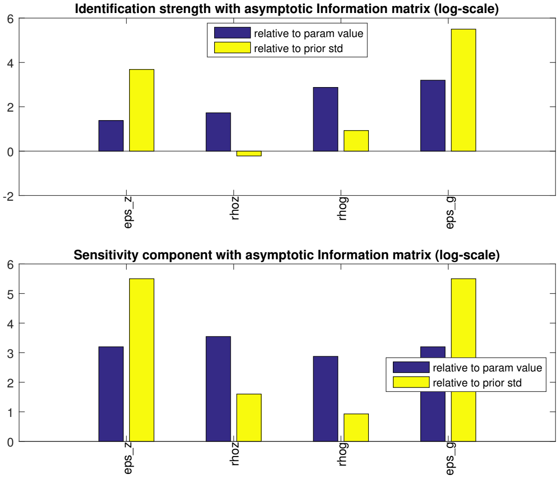
Dynare Implementation

- Because some parameters may have a prior mean of zero, an alternative normalization uses the prior standard deviation $\sigma(\theta_i)$:

$$s_i = \sqrt{\frac{\left(\sigma\left(\theta_i\right)\right)^2}{\left(I(\theta)^{-1}\right)_{(i,i)}}}$$

$$\tilde{\Delta}_i = \sqrt{\left(\sigma\left(\theta_i\right)\right)^2 I(\theta)_{(i,i)}}$$

Dynare Output for Our RBC Model



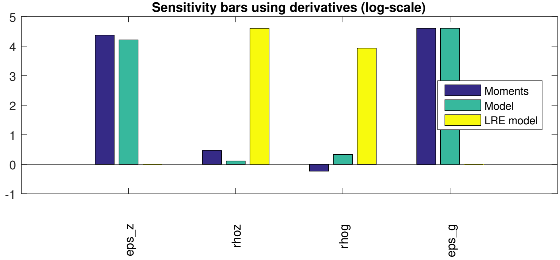
A Different Sensitivity Measure

- Another sensitivity measure is how changes in the elements of the parameter vector θ impact the model moments
- Can be measured locally using the Jacobian, where the entry $\partial m_j / \partial \theta_i$ needs to be normalized
- Can be done by multiplying with the ratio of standard deviations

$$\operatorname{std}(\theta_i) / \operatorname{std}(m_j) \tag{27}$$

- Accounts for different parameter uncertainty by ascribing more importance to more variable parameters (effectively normalizing across parameters i)
- Normalization with standard deviation $\operatorname{std}(m_j)$ allows for comparing the impact of parameter i on differently volatile moments (normalizing across moments j).
- Norm of the columns of the standardized Jacobian yields single aggregate sensitivity measure over all moments j for each parameter i.

Dynare Output for Our RBC Model

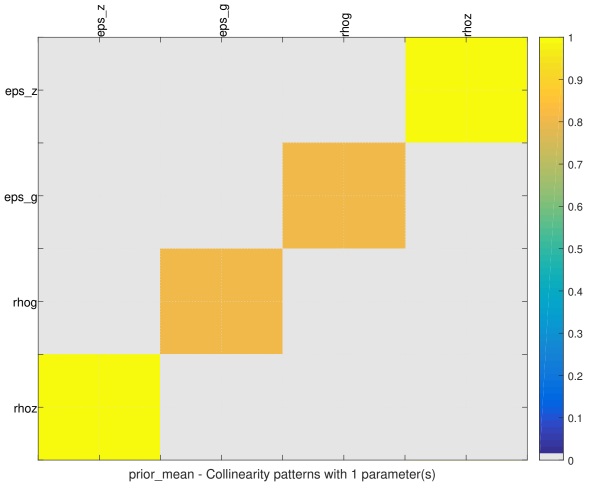


- The respective Jacobian matrices refer to
  1. the moments matrix $(\partial m_T/\partial\theta')$, indicating how well a parameter can be identified due the strength of its impact on the observed moments
  2. the model solution matrices $(\partial\tau/\partial\theta')$, indicating how well a parameter could in principle be identified if all state variables were observed
  3. the Linear Rational Expectations model $(\partial\gamma/\partial\theta')$, indicating trivial cases of non-identifiability due to e.g. some parameters always showing up as a product in the model equations

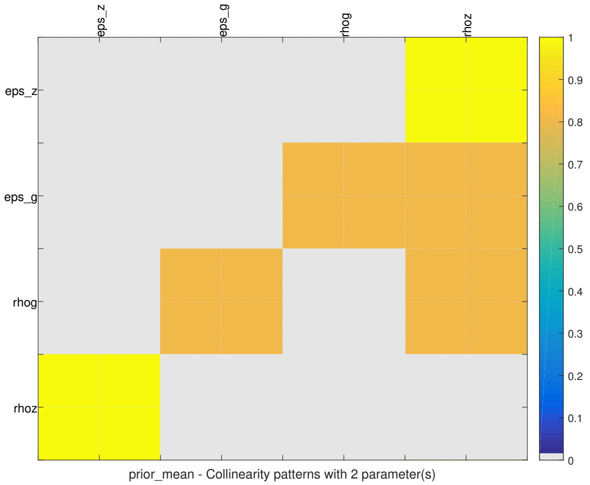
Correlation Component

- Even if the likelihood itself is sensitive, there might be (perfect) collinearity
- Need to look at correlation component
- This analysis is conducted via brute force: For each single parameter, a set of regressions is run of the column of the Jacobian corresponding to the parameter in the row on all possible combinations of other Jacobian columns
- Aim: finding the column (and thus parameter) combination with the highest R^2

Dynare Output for Our RBC Model



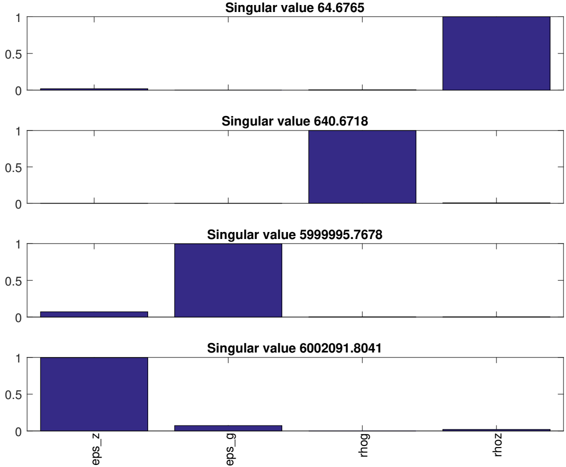
Dynare Output for Our RBC Model



Dynare Output for Our RBC Model

- Following Andrle (2010), identification can also be judged from a singular value decomposition (SVD) of the information matrix
- Provides the size of the singular values and the associated eigenvectors (i.e. parameters)
- Parameter combinations associated with the smallest singular values are closest to being perfectly collinear and thus redundant
- Singular value of 0 implies that the parameter is completely unidentified as it is responsible for the information matrix being rank deficient

Dynare Output for Our RBC Model: SVD



Bibliography I

- Andrle, Michal (2010). "A note on identification patterns in DSGE models". Working Paper Series 1235. European Central Bank.
- Canova, Fabio and Luca Sala (2009). "Back to square one: identification issues in DSGE models". *Journal of Monetary Economics 56* (4), 431–449.
- Cochrane, John H. (2011). "Determinacy and identification with Taylor rules". *Journal of Political Economy 119* (3), 565–615.
- Fernández-Villaverde, Jesús and Juan F. Rubio-Ramírez (2007). "Estimating macroeconomic models: a likelihood approach". *Review of Economic Studies 74* (4), 1059–1087.
- Iskrev, Nikolay (2010). "Local identification in DSGE models". *Journal of Monetary Economics 57*, 189–202.
- ——— (2011). "Evaluating the strength of identification in DSGE models. an a priori approach". Mimeo. Banco de Portugal.

Bibliography II

Ratto, Marco and Nikolay Iskrev (2011). "Algorithms for identification analysis under the Dynare environment: final version of the software". Mimeo. Joint Research Centre, European Commission.