On Line appendix for Choosing the variables to estimate singular DSGE models

Fabio Canova
EUI and CEPR

Filippo Ferroni
Banque de France, University of Surrey

Christian Matthes
Federal Reserve Bank of Richmond

November 12, 2013

This on-line appendix reports the results obtained when changing some of the parameters appearing in the two criterion function and when the DGP features different shocks. The next section has a brief discussion; tables A.1-A.7 the results. Figure A.1 the slope of the convoluted likelihood function in certain dimensions for the DGP and two different vectors of variables. Figure A.2 presents the responses to monetary shocks in the GDP and in the five estimated models.

1 Robustness

We have examined whether the essence of the conclusions change when we alter the nuisance parameters present in each of the procedures. Here we describe the results obtained for a subset of these exercises. The basic conclusions we have derived in the text hold also in these alternative setups.

1.1 Six Observables

We have examined what happens if six (rather than four) shocks drive the economy. Thus, we have added price markup and wage markup shocks to the list of disturbances and repeated the analysis maintaining a maximum of seven observable variables. Tables 3, 4, and 5 report the results obtained with the three approaches.

It is still true that output, consumption and investment must be present among the observables when estimating the structural parameters of the model. Adding hours, inflation and the real wage seems the best option, as this combination is at the top of the ordering according to the information analysis, and is among the top ones both with rank and the elasticity analyses. Notice that, with six shocks, the rank analysis becomes less informative (six of the seven combinations are equivalent according to this criteria) and the relative differences in the "elasticity" function for top combinations decrease. The $p_t(\theta)$ statistic instead is still quite sharp in distinguishing the best vector of variables from the others.

1.2 Increasing the sample size

The sample size used in the exercises is similar to the one typically employed in empirical studies. However, in the elasticity and the information analyses, sample uncertainty may be important for the conclusions. For this reason, we have repeated the exercises using T=1500. The results are reported in the first panel of table 1 and in the second panel of table 2.

Sampling variations seems to be a minor issue. The ordering of the first four top combinations in the information analysis is the same when T=150 and T=1500: the averaging approach we use to construct $p_t(\theta)$ helps in this respect. There are some switches in the ordering obtained with the elasticity analysis, but the top combinations with the smaller T are still the best with T=1500.

1.3 Changing the variance of the convolution error

The variance of the convolution error is important as it contaminates the information present in the density of the data. In the baseline exercises, we have chosen it to be of the same order of magnitude as the variance of the structural shocks. This is a conservative choice and adds considerable noise to the likelihood function. In the fourth panel of table 1 and in the third panel of table 2, we report the top four combinations obtained with four observables when the variance of the convolution error is arbitrarily set to $\Sigma_u = 0.01 * I$.

There are no changes in the top four combinations when the $p_t(\theta)$ statistics is used. This is expected since convolution error is averaged out. This is not the case for the elasticity analysis - we have conditioned here on a particular realization of u_t. Nevertheless, even with this criteria, the vector which was best in the baseline scenario is still the preferred one in the alternative scenario we consider.

1.4 Changing the step size in the numerical derivatives

In computing numerical derivatives in both the rank and elasticity analysis we have to select the step size g, which defines the radius of the neighborhood of the parameter vector over which identification is measured. Since the choice is arbitrary, we have repeated the exercise using $g = 0.001$ - which implies a much smaller radius. Choosing a smaller g has minor effects on the elasticity analysis (see second panel of table 1) but affects the conclusions of the rank analysis: now all the combinations have similar rank and either they fail to achieve identification (the case of the unrestricted model) or achieve identification (the case of restricted model). Thus, it seems that, once restrictions are imposed, in a very small neighborhood of the chosen parameter vector the parameter vector is identifiable. However, since this is not the case as we make the neighborhood slightly larger, weak identification seems a pervasive feature of this model.

1.5 Quadratic information distance

Rather than measuring the informativeness of an observable vector with the $p_t^j(\theta)$ statistics, we have also considered, as an alternative, the following quadratic measure of distance:

$$Q_j(\theta, e^{t-1}, u_t) = \sum_{t=1}^{T} (Z_t - W_{jt})' \Sigma_q^{-1} (Z_t - W_{jt}) \tag{1}$$

where $\Sigma_q = \Sigma_{y_j} + \Sigma_y + 2\Sigma_u$, which is the sum of the conditional covariance matrices of Z_t and W_{jt}. While this choice is somewhat arbitrary, it is a useful metric to check to what extent our results depend on the exact measure of information used. We are seeking the combination of variables which minimizes $Q_j(\theta, e^{t-1}, u_t)$, integrating out both the history of the shocks and the convolution error.

When four observables are used (see first panel of table 2), the top four combinations obtained with the $p_t^j(\theta)$ and the $Q_j(\theta)$ statistics are the same. The ordering of the two best combinations is reversed but differences in the relative informativeness of the two vectors is small. When six observables are used (see table 5), the same conclusion holds. However, the difference between the two best vectors, which was large under the $p_t^j(\theta)$ measure, is substantially reduced with the $Q_j(\theta)$ measure.

1.6 Different DGP

The results we have obtained may be dependent on the DGP. To check for this possibility we have eliminated the government spending shock from the model and inserted instead a preference shock. The preference shock is specified as Smets and Wouters (2007).

Tables 6 and 7 report the top combinations using the rank and the information criteria, respectively. As is it clear, the overall ranking of best combinations is not affected by the change. The vector (y,c, i, w) is now the preferred one according to both the rank and the information analysis.

1.7 Other exercises

We have also examined what happens when we change the tolerance level r in the rank analysis and found only minor differences with the baseline scenario. We have also considered the case when the neutral technology shock has a unit root. Having a process with one unit root reduces the number of structural parameters to be estimated, but adding trend information should be irrelevant for both rank and information analysis, so long as the trend is correctly specified, since the trend has no information for any parameter other than the variance of the neutral technology shock. Indeed, we confirm that the ordering of the best combinations is independent of the presence of trends in the data.

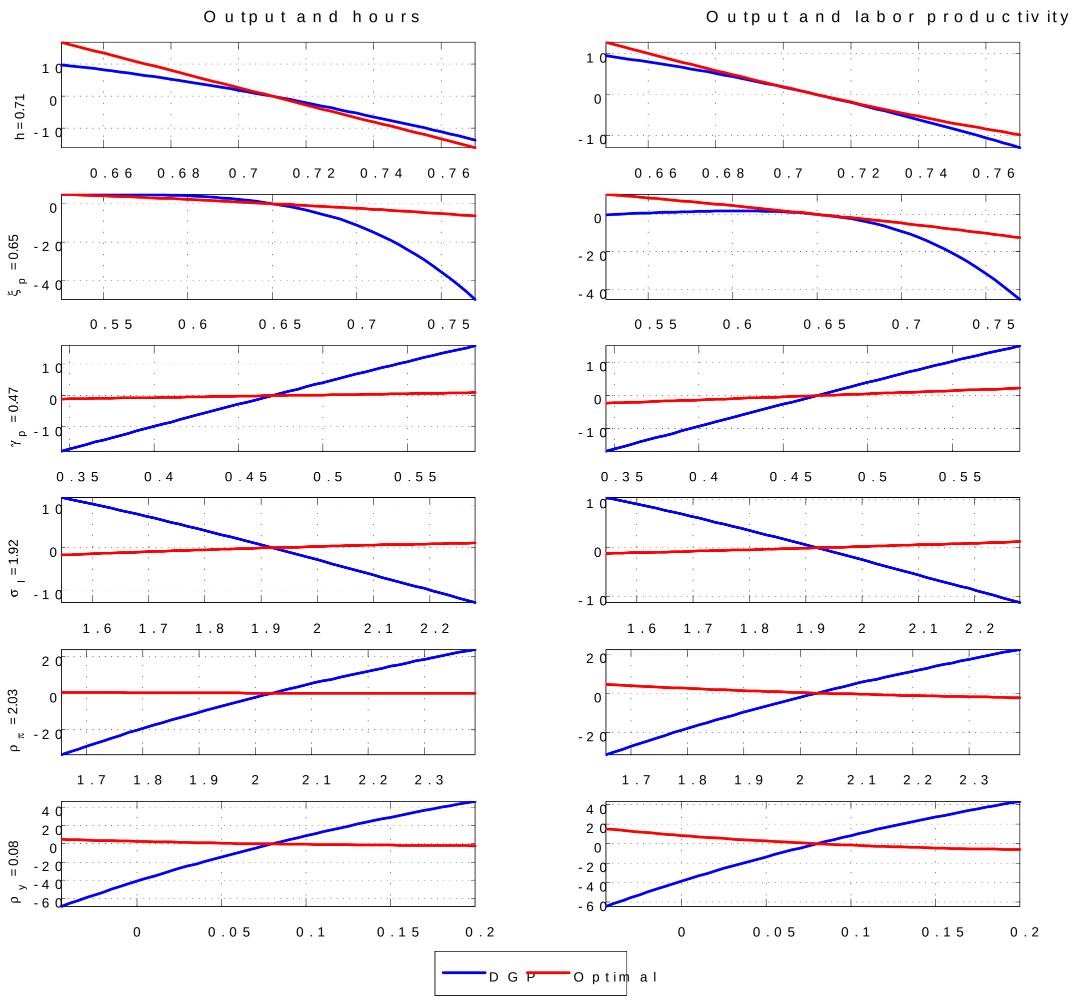

Figure A.1: One dimensional convoluted likelihood slope; DGP and various vectors.

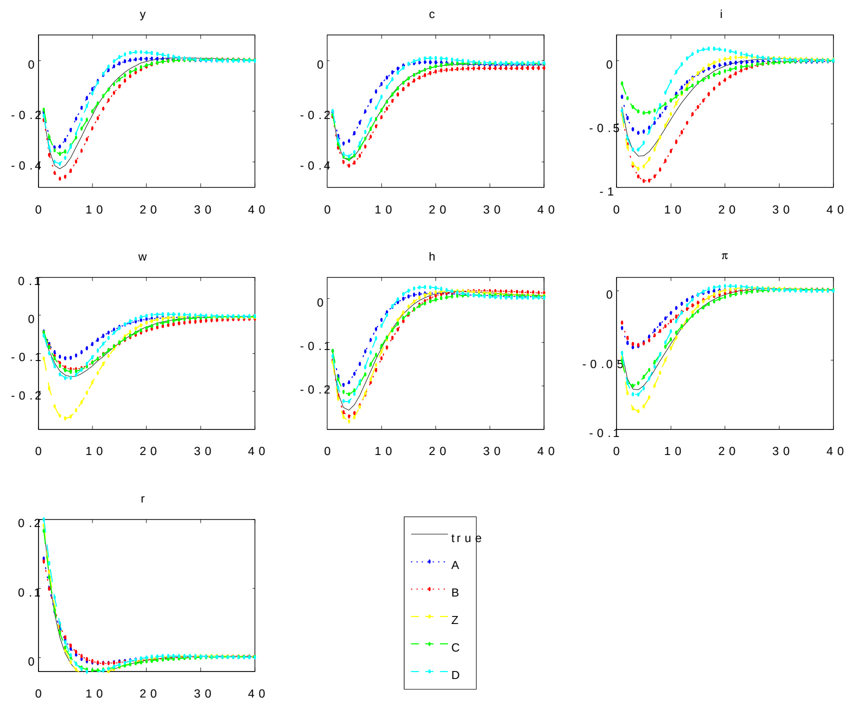

Figure A.2:True and estimated response to monetary shocks

Order	Cumulative Deviation	Weighted Square	Ratio
T=1500			
1	(c, i, r, h)	(c, i, r, h)	1.00
2	$(y.c, i, w)$	(c, i, w, h)	1.64
3	(y, c, r, h)	(y, c, i, w)	1.65
4	(c, i, w, h)	(y, c, r, h)	2.15
g=0.001			
1	(c, i, r, h)	(c, i, r, h)	1.00
2	$(y.c, i, w)$	(c, i, w, h)	1.61
3	(c, i, w, h)	(y, c, i, w)	1.91
4	(y, c, r, h)	(y, c, r, h)	2.09
$\Sigma_u = 0.01 * I$			
1	(c, i, r, h)	(c, i, r, h)	1.00
2	(c, i, w, h)	(c, i, w, h)	1.14
3	(y, c, r, h)	(y, c, r, h)	1.65
4	(y, c, i, w)	(y, i, π, r)	3.11

Table 1: Ranking of the four top combinations of variables using elasticity distance. Unrestricted SW model. The first column uses as objective function the sum of absolute deviation of the likelihood curvature of the parameters, the second the weighed sum of square deviations of the likelihood curvature of the parameters. The third the value of the objective function relative to the best combination. The first panel reports results obtained increasing the sample size T, the third, changing the step size g used in computing derivatives, the fourth the magnitude of the convolution error Σ_u.

Order	Quadratic Distance		T=1500		$\Sigma_u = 0.01 * I$	
	Combination	Relative Information	Combination	Relative Information	Combination	Relative Information
1	(y, c, i, w)	1	(y, c, i, h)	1	(y, c, i, h)	1
2	(y, c, i, h)	0.89	(y, c, i, w)	0.87	(y, c, i, w)	0.86
3 (y, c, i, r)	0.6	(y, c, i, r)	0.51	(y, c, i, r)	0.51	
4	(y, c, i, π)	0.59	(y, c, i, π)	0.5	(y, c, i, π)	0.5

Table 2: Ranking based on the $p(\theta)$ statistic. The first two columns present results for the basic setup, the next six columns the results obtained altering nuisance parameters. Relative information is the ratio of the $p(\theta)$ statistic relative to the statistic obtained for the best combination.

Combinations	Unrestricted Rank			Restricted Rank		
	$\Delta_{\Lambda T}$	$\Delta_{\Lambda U}$	Δ	$\Delta_{\Lambda T}$	$\Delta_{\Lambda U}$	Δ
(y, c, i, w, π, r)	227	67	263	229	69	265
(y, c, i, w, π, h)	227	67	263	229	69	265
(y, c, i, w, h, r)	227	67	263	229	69	265
(y, c, i, h, π, r)	227	67	263	229	69	265
(y, c, h, w, π, r)	227	67	262	229	69	264
(y, h, i, w, π, r)	227	67	263	229	69	265
(h, c, i, w, π, r)	227	67	263	229	69	265
Required	229	69	265	229	69	265

Table 3: Ranks for combinations of variables in the unrestricted SW model (columns 2-5) and in the restricted SW model (columns 6-9), where five parameters are fixed $\delta = 0.025$, $\varepsilon_p = \varepsilon_w = 10$, $\lambda_w = 1.5$ and $cg = 0.18$ when there are 6 observables.

Order	Cumulative Deviation	Weighted Square	Ratio
Basic			
1	(c, i, w, π, r, h)	(c, i, w, π, r, h)	1.00
2	(y, c, w, π, r, h)	(y, c, w, π, r, h)	1.27
3	(y, c, i, w, π, h)	(y, i, w, π, r, h)	1.38
4	(y, i, w, π, r, h)	(y, c, i, w, π, h)	1.52
T=1500			
1	(y, c, w, π, r, h)	(c, i, w, π, r, h)	1.00
2	(c, i, w, π, r, h)	(y, c, w, π, r, h)	1.10
3	(y, c, i, w, π, h)	(y, i, w, π, r, h)	1.18
4	(y, i, w, π, r, h)	(y, c, i, w, π, h)	1.40
g=0.001			
1	(c, i, w, π, r, h)	(c, i, w, π, r, h)	1.00
2	(y, c, w, π, r, h)	(y, c, w, π, r, h)	1.45
3	(y, c, i, w, π, h)	(y, i, w, π, r, h)	1.60
4	(y, i, w, π, r, h)	(y, c, i, w, π, h)	1.71
$\Sigma_u = 0.01 * I$			
1	(y, c, i, w, π, r)	(y, c, i, w, π, r)	1.00
2	(y, c, w, π, r, h)	(y, c, w, π, r, h)	1.12
3	(c, i, w, π, r, h)	(c, i, w, π, r, h)	1.21
4	(y, c, i, w, π, r)	(y, c, i, w, π, r)	1.34

Table 4: Ranking of four top combinations of variables using elasticity distance. Unrestricted SW model, six shock system. The first column uses as objective function the sum of absolute deviation of the likelihood curvature of the parameters, the second the weighed sum of square deviations of the likelihood curvature of the parameters. The third the value of the objective function relative to the best combination. The first panel reports the baseline results, the second increasing the sample size T, the third changing the step size g in computing derivatives, the fourth changing the magnitude of the variance of the convolution error Σ_u.

Order	Basic		Quadratic Objective	
	Combination	Relative info	Combination	Relative info
1	(y, c, i, h, w, π)	1	(y, c, i, w, r, h)	1
2	(y, c, i, w, r, h)	0.4	(y, c, i, h, w, π)	0.82
3	(y, c, i, r, π, h)	0.04	(y, c, i, r, π, h)	0.13
4	(y, c, i, π, w, r)	0.02	(y, c, i, π, w, r)	0.02

Table 5: Ranking according to the $p(\theta)$ statistic, 6 observables. The first two columns present the results for the basic setup, the next two columns the results obtained with the alternative objective function. Relative information is the ratio of the $p(\theta)$ statistic relative to the statistics for the best combination.

	Unrestricted Rank(Δ)	Restricted Rank(Δ)	Efficient Combiantion 3 Restrictions, i.e. $(\varepsilon_p, \varepsilon_w)$ and
y, c, i, w	185	188	(λ_w),(ζ_ω),(σ_c)
y, c, i, r	185	188	(ζ_p),(σ_n), (λ_w),(ϕ_p),(ζ_ω)
y, c, w, r	185	188	(c/g),(ψ),
y, c, r, h	185	188	(c/g),(ψ),
y, i, w, r	185	188	(λ_w),(c/g),(ζ_ω),(σ_n),$(\sigma_c$)
y, i, r, h	185	188	(λ_w),(c/g),(ζ_ω),(σ_n),$(\sigma_c$
c, i, w, r	185	188	(ρ_y),$(\rho_{\Delta y})$, (λ_w),(c/g),(ϕ_p),(ζ_ω),(ζ_p)
c, i, r, h	185	188	(β), (λ_w),(c/g),(ϕ_p),(ζ_ω),(ζ_p),(ρ_y)
i, w, pi, r	185	188	(λ_w),(ζ_ω)
i, w, r, h	185	188	(λ_w),(ζ_ω)
y, c, w, pi	184	188	
y, c, w, h	184	188	
y, c, pi, r	184	188	
y, c, pi, h	184	188	
y, i, w, pi	184	188	
y, i, w, h	184	188	
y, i, pi, r	184	188	
y, i, pi, h	184	188	
c, i, w, pi	184	188	
c, i, w, h	184	188	
c, i, pi, r	184	188	
c, i, pi, h	184	188	
i, pi, r, h	184	188	
y, c, i, pi	184	187	
y, c, i, h	184	187	
y, w, pi, r	184	187	
y, w, pi, h	184	187	
y, w, r, h	184	187	
y, pi, r, h	184	187	
c, w, pi, r	184	187	
c, w, r, h	184	187	
c, pi, r, h	184	187	
i, w, pi, h	184	187	
c, w, pi, h	183	187	
w, pi, r, h	183	187	
Required	188		

Table 6: Ranks for combination of variables. The DGP has a technology, an investment, a monetary and and preference shock.

Combination	Relative Information
y,c,i,w	1
y,c,i,h	0.76
y,c,i,r	0.48
y,c,i, π	0.46

Table 7: Ranking according to the $p(\theta)$ statistic. The DGP has a technology, an invetsment, a monetary and and preference shock. Relative information is the ratio of the $p(\theta)$ statistic relative to the statistics for the best combination.