Choosing the variables to estimate singular DSGE models

Fabio Canova
EUI and CEPR

Filippo Ferroni *
Banque de France, University of Surrey

Christian Matthes
UPF and Barcelona GSE

March 3, 2013

Abstract

We propose two methods to choose the variables to be used in the estimation of the structural parameters of a singular DSGE model. The first selects the vector of observables that optimizes parameter identification; the second the vector that minimizes the informational discrepancy between the singular and non-singular model. An application to a standard model is discussed and the estimation properties of different setups compared. Practical suggestions for applied researchers are provided.

Key words: ABCD representation, Identification, Density ratio, DSGE models.

JEL Classification: C10, E27, E32.

---

*Corresponding author, e-mail: `filippo.ferroni@banque-france.fr`. The views expressed in this paper do not necessarily reflect those of the Banque de France.

We would like to thank H. Van Djik, two anonymous referees, Z. Qu, M. Ellison, F. Kleibergen, A. Justiniano, G. Primiceri and V. Curdia and the participants of numerous seminars and conferences for comments and suggestions.

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models feature optimal decision rules that are singular. This occurs because the number of endogenous variables generally exceeds the number of exogenous shocks. For example, a basic RBC structure generates implications for consumption, investment, output, hours, real wages, the real interest rate, etc. Since both the short run dynamics and the long run properties of the endogenous variables are driven by a one dimensional exogenous technological process, the covariance matrix of the data is implicitly assumed to be singular.

The problem can be mitigated if some endogenous variables are non-observables - for example, data on hours is at times unavailable - since the number of variables potentially usable to construct the likelihood function is smaller. In other cases, the data may be of poor quality and one may be justified in adding measurement errors to some equations. This lessens the singularity problem since the number of shocks driving a given number of observable variables is now larger. However, neither non-observability of some endogenous variables nor the addition of justified measurement error is generally sufficient to completely eliminate the problem. While singularity is not troublesome for limited information structural estimation approaches, such as impulse response matching, it creates important headaches to researchers using full information likelihood methods, both of classical or Bayesian inclinations.

Two approaches are generally followed in this situation. The first involves enriching the model with additional shocks (see e.g. Smets and Wouters, 2007). In many cases, however, shocks with dubious structural interpretation are used with the only purpose to avoid singularity and this complicates inference when they turn out to matter, say, for output or inflation fluctuations (see Chari et al., 2009, Sala et al., 2010, Chang et al., 2013). The second is to solve out variables from the optimality conditions until the number of endogenous variables equals the number of shocks. This approach is also problematic: the convenient state space structure of the decision rules is lost, the likelihood is an even more nonlinear function of the structural parameters and can not necessarily be computed with standard Kalman filter recursions. In addition, with k endogenous variables and $m < k$ shocks, one can form many non-singular systems with only m endogenous variables and, apart from computational convenience, solid principles to choose which combination should be used in estimation are lacking.

Guerron Quintana (2010), who estimated a standard DSGE model adding enough measurement errors to avoid singularity, shows that estimates of the structural parameters may depend on the observable variables and suggests to use economic hindsight and an out-of-sample MSE criteria to decide the combination to be employed in estimation. Del Negro and Schorfheide (forthcoming) indicate that the information set available to the econometrician matters for forecasting in the recent recession. Economic hindsight may be dangerous, since prior falsification becomes impossible. On the other hand, a MSE criteria is not ideal as variable selection procedure since biases (which we would like to avoid in estimation) and variance reduction (which are a much less of a concern in DSGE estimation) are equally weighted.

This paper proposes two complementary criteria to choose the vector of variables to be used in the estimation of the parameters of a singular DSGE model. Since Canova and Sala (2009) have shown that DSGE models feature important identification problems that are typically exacerbated when a subset of the variables or of the shocks is used in estimation [1], our first criterion selects the variables to be used in likelihood based estimation keeping parameter identification in mind. We use two measures to evaluate the local identification properties of different combinations of observable variables. First, following Komunjer and Ng (2011), we examine the rank of the matrix of derivatives of the ABCD representation of the solution with respect to the parameters for different combinations of observables. Given an ideal rank, the selected vector of observables minimizes the discrepancy between the ideal and the actual rank of this matrix. Since a subset of parameters is typically calibrated, we show what additional restrictions allow the identification of the remaining structural parameters.

The Komunjer and Ng approach does not necessarily deliver a unique candidate and it is silent about the subtle issues of weak and partial identification [2]. Thus, we complement the rank analysis by evaluating the difference in the local curvature of the convoluted likelihood function of the singular system and of a number of non-singular alternatives which fare well in the rank analysis. The combination of variables we select makes the average curvature of the convoluted likelihoods of the non-singular and singular systems close in the dimensions of interest.

[1] Earlier work discussing identification issues in single equations of DSGE models include Mavroedis (2005), Kleibergen and Mavroedis (2009), and Cochrane (2011).

[2] Recent work describing how to construct confidence regions which are robust to weak identification problems include Guerron Quintana, et al. (2012), Andrews and Mikusheva (2011) and Dufour et al. (2009).

The second criterion employs the informational content of the densities of the singular and the non-singular systems and selects the variables to be used in estimation to make the information loss minimal. We follow recent advances by Bierens (2007) to construct the density of singular and non-singular systems and to compare the informational content of vectors of observables, taking the structural parameters as given. Since the measure of informational distance depends on nuisance parameters, we integrate them out prior to choosing the optimal vector of observables.

We apply the methods to select observables in a singular version of the Smets and Wouters (2007) (henceforth SW) model. We retain the full structure of nominal and real frictions but allow only a technology, an investment specific, a monetary and a fiscal shock to drive the seven observable variables of the model. In this economy, parameter identification and variable informativeness are optimized including output, consumption and investment and either real wages of hours worked among the observables. These variables help to identify the intertemporal and the intratemporal links in the model and thus are useful to correctly measure income and substitution effects, which crucially determine the dynamics of the model in response to the shocks. Interestingly, using interest rate and inflation jointly in the estimation makes identification worse and the loss of information due to variable reduction larger. When one takes the curvature of the likelihood into consideration, the nominal interest rate is weakly preferable to the inflation rate.

We also show that, in terms of likelihood curvature, there are important trade-offs when deciding to use hours or labor productivity together with output among the observables and demonstrate that changes in the setup of the experiment do not alter the main conclusions of the exercise.

Our ranking criteria would be irrelevant if the conditional dynamics obtained with different vectors of observables were similar. We show that the best and the worst combinations of variables indeed produce different responses to shocks and that approaches that tag on measurement errors or non-existent structural shocks to use a larger number of observables in estimation, may distort parameter estimates and jeopardize inference.

The paper is organized as follows. The next section describes the methodologies. Section 3 applies the approaches to a singular version of a standard model. Section 4 estimates models with different variables and compares the dynamic responses to interesting shocks. Section 5 concludes.

2 The selection procedures

The log-linearized decision rules of a DSGE model have the state space format

$$\begin{aligned} x_t &= A(\theta)x_{t-1} + B(\theta)e_t && (1)\\ y_t &= C(\theta)x_{t-1} + D(\theta)e_t && (2)\\ e_t &\sim N(0, \Sigma(\theta)) \end{aligned}$$

where x_t is a $n_x \times 1$ vector of predetermined and exogenous states, y_t is a $n_y \times 1$ vector of endogenous controls, e_t is a $n_e \times 1$ vector of exogenous innovations and, typically $n_e < n_y$. Here $A(\theta), B(\theta), C(\theta), D(\theta), \Sigma(\theta)$ are matrices, function of the vector of structural parameters θ. Assuming left invertibility of $A(\theta)$, one can solve out the x_t's and obtain a MA representation for the vector of endogenous controls:

$$y_t = \Big(C(\theta)\left(I - A(\theta)L\right)^{-1} B(\theta)L + D(\theta)\Big)\, e_t \equiv H(L,\theta)e_t \tag{3}$$

where L is the lag operator. Thus, the time series representation of the log-linearized solution for y_t is a singular MA(∞) since $D(\theta)e_te_t'D(\theta)'$ has rank $n_e < n_y$[3].

From (3) one can generate a number of non-singular structures, using a subset j of endogenous controls, $y_{jt} \subset y_t$, simply making sure the dimensions of the vector of observable variables and of the shocks coincide. Given (3), one can construct $J = \begin{pmatrix} n_y \\ n_e \end{pmatrix} = \frac{n_y!}{(n_y-n_e)!n_e!}$ non-singular models, differing in at least one observable variable. Let the MA representation for the non-singular model $j = 1, \ldots, J$ be

$$y_{jt} = \Big(C_j(\theta)\left(I - A(\theta)L\right)^{-1} B(\theta)L + D_j(\theta)\Big)\, e_t \equiv H_j(L,\theta)e_t \tag{4}$$

where $C_j(\theta)$ and $D_j(\theta)$ are obtained from the rows corresponding to y_{jt}. The non-singular model j has also a MA(∞) representation, but the rank of $D_j(\theta)e_te_t'D_j(\theta)'$ is $n_e = n_y$. Our criteria compare the properties of y_t and those of y_{jt} for different j.

Komunjer and Ng (2011) derived necessary and sufficient conditions that guarantee local identification of the parameters of a log-linearized solution of a DSGE model. Their approach requires calculating the rank of the matrix of the derivatives of $A(\theta)$, $B(\theta)$, $C_j(\theta)$, $D_j(\theta)$ and $\Sigma(\theta)$ with respect to the parameters θ and of the

[3](3) does not require assumptions about the dimensions of n_x, n_y and n_e which would be needed to compute, for example, the VARMA representation used in e.g. Kascha and Mertens (2008).

derivatives of the linear transformations, T and U, that deliver the same spectral density for the observables. Under regularity conditions, they show that two systems are observationally equivalent if there exist triples $(\theta_0, I_{n_x}, I_{n_e})$ and (θ_1, T, U) such that $A(\theta_1) = TA(\theta_0)T^{-1}$, $B(\theta_1) = TB(\theta_0)U$, $C_j(\theta_1) = C_j(\theta_0)T^{-1}$, $D_j(\theta_1) = D_j(\theta_0)U$, $\Sigma(\theta_1) = U^{-1}\Sigma(\theta_0)U^{-1}$, with T and U being full rank matrices[4].

For each combination of observables y_{jt} , define the mapping

$$\delta_j(\theta, T, U) = \left(vec(TA(\theta)T^{-1}),\ vec(TB(\theta)U),\ vec(C_j(\theta)T^{-1}),\ vec(D_j(\theta)U),\ vech(U^{-1}\Sigma(\theta)U^{-1})\right)'$$

We study the rank of the matrix of the derivatives of $\delta_j(\theta, T, U)$ with respect to $(\theta,\ T,\ U)$ evaluated at $(\theta_0, I_{n_x}, I_{n_e})$, i.e. for $j = 1, ..., J$ we compute the rank of

$$\begin{aligned}\Delta_j(\theta_0) \equiv \Delta_j(\theta_0, I_{n_x}, I_{n_e}) &= \left(\frac{\partial\delta_j(\theta_0, I_{n_x}, I_{n_e})}{\partial\theta}, \frac{\partial\delta_j(\theta_0, I_{n_x}, I_{n_e})}{\partial T}, \frac{\partial\delta_j(\theta_0, I_{n_x}, I_{n_e})}{\partial U}\right) \\ &\equiv (\Delta_{j,\Lambda}(\theta_0),\ \Delta_{j,T}(\theta_0),\ \Delta_{j,U}(\theta_0))\end{aligned}$$

$\Delta_{j,\Lambda}(\theta_0)$ defines the local mapping between θ and $\Lambda(\theta) = [A(\theta), B(\theta), C_j(\theta), D_j(\theta), \Sigma(\theta)]$, the matrices of the decision rule. When $\text{rank}(\Delta_{j,\Lambda}(\theta_0)) = n_\theta$, the mapping is locally invertible. The second block contains the partial derivatives with respect to T: when $\text{rank}(\Delta_{j,T}(\theta_0)) = n_x^2$, the only permissible transformation is the identity. The last block corresponds to the derivatives with respect to U: when $\text{rank}(\Delta_{j,U}(\theta_0)) = n_e^2$, the spectral factorization uniquely determines the duple $(H_j(L;\theta), \Sigma(\theta))$. A necessary and sufficient condition for local identification at θ_0 is that

$$\text{rank}(\Delta_j(\theta_0)) = n_\theta + n_x^2 + n_e^2 \tag{5}$$

Thus, given a θ_0, we compute the rank of $\Delta_j(\theta_0)$ for each y_{jt} vector. The vector of variables j minimizing the discrepancy between rank $(\Delta_j(\theta_0))$ and $n_\theta + n_x^2 + n_e^2$, the theoretical rank needed to achieve identification of all parameters, is the one selected for full information estimation of the parameters.

The rank comparison should single out combinations of endogenous variables with different identification content. However, ties may result. Furthermore, the setup is not suited to deal with weak and partial identification problems, which often plague likelihood based estimation of DSGE models (see e.g. An and Schorfheide, 2007 or

[4] We use slightly different definitions than Komunjer and Ng (2011). They define a system to be singular if the number of observables is larger or equal to the number of shocks, i.e. $n_e \leq n_y$. Here a system is singular if $n_e < n_y$ and non-singular if $n_e = n_y$.

Canova and Sala, 2009). For this reason, we also compare measures of the elasticity of the convoluted likelihood function with respect to the parameters in the singular system and in the non-singular systems which are best according to the rank analysis - see next paragraph on how to construct the convoluted likelihood. We seek the combination of variables which makes the average curvature of the convoluted likelihood around θ_0 in the singular and non-singular systems close. We considered two distance criteria:

$$D1 = \sum_{i=1}^{q} |\frac{\partial logL(\theta_i)}{\partial\theta_i} - \frac{\partial logL^*(\theta_i)}{\partial\theta_i}| \tag{6}$$

$$D2 = \sum_{i=1}^{q} (\frac{\frac{\partial logL(\theta_i)}{\partial\theta_i} - \frac{\partial logL^*(\theta_i)}{\partial\theta_i}}{\sum_{i=1}^{q}(\frac{\partial logL(\theta_i)}{\partial\theta_i} - \frac{\partial logL^*(\theta_{i0})}{\partial\theta_i})^2} * \frac{L^*(\theta_{i0})}{H(\theta_{i0})})^2 \tag{7}$$

where $L^*(\theta_i)$ is the value of the convoluted likelihood of the original singular system and $H(\theta_{i0})$ the curvature at the true parameter value α_i. In the first case, absolute elasticity deviations are summed over the parameters of interest. In the second, we consider a weighted sum of the square deviations is considered, where the weights depend on the sharpness of the likelihood of the singular system at θ_0.

The other statistic measures the relative informational content of the original singular system and of a number of non-singular counterparts. To measure the informational content, we follow Bierens (2007) and convolute y_{jt} and of y_t with a $n_y \times 1$ random iid vector. Thus, the vectors of observables are now

$$Z_t = y_t + u_t \tag{8}$$

$$W_{jt} = Sy_{jt} + u_t \tag{9}$$

where $u_t \sim N(0, \Sigma_u)$ and S is a matrix of zeros, except for some elements on the main diagonal, which are equal to 1. S insures that Z_t and W_{jt} have the same dimension n_y. For each non-singular structure j, we construct

$$p_t^j(\theta_0, e^{t-1}, u_t) = \frac{\mathcal{L}(W_{jt}|\theta_0, e^{t-1}, u_t)}{\mathcal{L}(Z_t|\theta_0, e^{t-1}, u_t)} \tag{10}$$

where $\mathcal{L}(m|\theta_0, e^{t-1}, u_t)$ is the density of $m = (Z_t, W_{jt})$, given the parameters θ, the history of the structural shock e^{t-1}, and the convolution error u_t. (10) can be easily computed, if we assume that e_t are normally distributed, since the first and second conditional moments of W_{jt} and Z_t are $\mu_{w,t-1} = SC_j(\theta_0)(I - A(\theta_0)\ L)^{-1}B(\theta)e_{t-1}$,

$\Sigma_{w_j} = SD_j(\theta_0)\Sigma(\theta_0)D_j(\theta_0)'S' + \Sigma_u$, $\mu_{z,t-1} = C(\theta_0)(I - A(\theta_0)\ L)^{-1}B(\theta_0)e_{t-1}$ and $\Sigma_z = D(\theta_0)\Sigma(\theta_0)D(\theta_0)' + \Sigma_u$.

Bierens imposes mild conditions that make the matrix $\Sigma_{w_j}^{-1} - \Sigma_z^{-1}$ negative definite for each j, and p_t^j well defined and finite, for the worst possible selection of y_t. Since these conditions do not necessarily hold in our framework, we integrate both the e^{t-1} and the u_t out of (10), and choose the combination of observables j that minimize the average information ratio $p_t^j(\theta_0)$, i.e.

$$\inf_j p_t^j(\theta_0) = \inf_j \int_{e^{t-1}} \int_{u_t} p_t^j(\theta_0, e_t^{t-1}, u_t) de^{t-1} du_t \tag{11}$$

$p_t^j(\theta_0)$ identifies the observables producing a minimum amount of information loss when moving from a singular to a non-singular structure, once we eliminate the influence due to the history of structural shocks and to the convolution error.

2.1 Discussion

The rank analysis is straightforward to undertake. However, the computation of the rank of $\Delta_j(\theta_0)$ may be problematic when the matrix is of large dimension and potentially ill-conditioned. Thus, care needs to be used. We recommend users to try to measure the rank of $\Delta_j(\theta_0)$ in different ways (for example, compute the condition numbers or the ratio of the sum of the smallest h roots to the sum of all the roots of the matrix) to make sure that results are not spurious. Also, one needs to make sure that the combination j satisfies the regularity conditions of Kommunjer and Ng, otherwise the ranking of vectors may not be appropriate.

(11) is related to standard entropy measures. In fact, if we take the log of $p_t^j(\theta, e^{t-1}, u_t)$, our information measure resembles the Kullback-Leibler information criteria (KLIC). Hence, our criterion implicitly takes into account the fact that the density of the approximating system is misspecified relative to the one of the singular system.

Both criteria are valid only locally around some θ_0. While it is far from clear how to render the identification criteria global, it is relatively simple to make our information measure global. Suppose we have prior $\mathcal{P}(\theta)$ on the structural parameters. One can then construct

$$q_t^j(e^{t-1}, u_t) = \frac{\int \mathcal{L}(W_{jt}|\theta, e^{t-1}, u_t)\mathcal{P}(\theta)d\theta}{\int \mathcal{L}(Z_t|\theta, e^{t-1}, u_t)\mathcal{P}(\theta)d\theta} \tag{12}$$

that is, one can average the densities of the singular and the non-singular model with respect to the parameters using $\mathcal{P}(\theta)$ as weight. The combination j that achieves $\inf_j \int_{e^{t-1}} \int_{u_t} q_t^j(e_t^{t-1}, u_t) de^{t-1} du_t$ can be found using Monte Carlo methods. We have decided to stick to our local criteria because the ranking of various combination of observables does not depend on the choice of θ, except in some knife-edge cases. The example in the next subsection highlights what these situations may be.

Since our criteria only require the ABCD representation of the (log)-linearized solution of the model, they are implementable prior to the estimation of the model and do not require the use of any vector of actual data. Given the decision rule, the procedures ask what combination of observables makes identification and information losses minimal, locally around some prespecified parameter vector. Thus, our analysis is exploratory in scope and similar in spirit to prior predictive exercises, sometimes used to study the properties of models (see e.g. Faust and Gupta, 2011).

Our scope is to identify observables with particular characteristics since this helps us to understand better the role these variables play in the model. As an alternative, one could also think of choosing linear combinations of observables that optimize certain statistical criteria. For example, one could choose the first m principal components of the observables (see Andrle, 2012), or optimize the linear weights to obtain the best identification and information properties. Such an approach has the advantage of using all the information the model provides. The disadvantage is that the variables used for estimation have no economic interpretation. For comparison, in the application section we also present results obtain using the first m static and dynamic principal components of the vector of observables. Since dynamic principal components are two sided moving averages of the observables, we maintain comparability with our original analysis by projecting them on the available information at each t.

2.2 An example

To illustrate what our criteria deliver in a situation where we can explicitly derive the decision rules, consider the simplified version of simple three equations New Keynesian model used in Canova and Sala (2009):

$$x_t = a_1 E_t x_{t+1} + a_2(i_t - E_t \pi_{t+1}) + e_{1t} \tag{13}$$

$$\pi_t = a_3 E_t \pi_{t+1} + a_4 x_t + e_{2t} \tag{14}$$

$$i_t = a_5 E_t \pi_{t+1} + e_{3t} \tag{15}$$

where x_t is the output gap, π_t the inflation rate, i_t the nominal interest rate, e_{it} is an iid demand shock, e_{2t} is an iid supply shock, and e_{3t} an iid monetary policy shock and E_t represents expectation, conditional on the information at time t. To make the model singular, let $e_{2t} = 0, \forall t$, so that $n_e = 2 < n_y = 3$.

Since there are no endogenous states, $A = B = C = 0$, so that the decision rules for (x_t, π_t, i_t) are just functions of the shocks and

$$\begin{bmatrix} x_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 & a_2 \\ a_4 & a_2 a_4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{3t} \end{bmatrix} \equiv D e_t \tag{16}$$

The forward looking parameters (a_1, a_3, a_5) disappear from the decision rules since the shocks are iid and there are no endogenous states. Thus, the rank of $\Delta(\theta_0)$ in (16) is 6 and it is deficient by 3 (the ideal rank is $n_\theta + 0 + n_e^2 = 9$).

Depending on which observables we select, the rank of $\Delta_j(\theta_0)$ could also be deficient by 3 if (x_t, π_t) or (π_t, i_t) are used, or by 4, if (x_t, i_t) are used, whenever a_4 is different from one. In this situation, our rank analysis would prefer (x_t, π_t) or (π_t, i_t) as observables. When $a_4 = 1$, however, all combinations are equivalent because, trivially, output and inflation have exactly the same information from the parameters.

Whenever $a_4 \approx 1$, the likelihood function of (x_t, π_t) has weak information about a_4, since the two observables have similar MA structure, but a system with (π_t, i_t) is unaffected. Hence, our elasticity analysis would lead us to prefer (π_t, i_t) as observables whenever inference about the slope of the Phillips curve parameter a_4 is important.

It is also easy to see what the informational analysis will give us. The convoluted (singular) system is

$$\begin{bmatrix} x_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 & a_2 & 1 \\ a_4 & a_2 a_4 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{3t} \\ u_t \end{bmatrix} \equiv D_1 v_t \tag{17}$$

The convoluted non-singular system including (x_t, π_t) is

$$\begin{bmatrix} x_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 & a_2 & 1 \\ a_4 & a_2 a_4 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{3t} \\ u_t \end{bmatrix} \equiv D_2 v_t \tag{18}$$

The convoluted non-singular system including (π_t, i_t) is

$$\begin{bmatrix} x_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ a_4 & a_2 a_4 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{3t} \\ u_t \end{bmatrix} \equiv D_3 v_t \tag{19}$$

Assuming normality and unit variance for the structural and the convoluted shocks, the population log likelihood of (17) will be proportional to $D_1^T D_1$ and the population log likelihoods of (18) and (19) will be proportional to $D_2^T D_2$ and $D_3^T D_3$. Thus, whenever $a_2 > 0$, the loss of information is minimized selecting (x_t, π_t). When $a_2 = 0$, the vectors (x_t, π_t), (π_t, i_t) produce the same loss of information.

In sum, this example shows two important points: the variables one may want to choose in estimation depend on the focus of the investigation - if two studies focus on different parameters of the same model, the optimal vector of observables may be different; the ranking of observable vectors our criteria deliver may depend on the true parameter values, but in a step-wise, discontinuous fashion.

3 An application

We apply our procedures to a singular version of Smets and Wouters' (2007) model. This model is selected because of its widespread use for policy analyses in academics and policy institutions, and because it is frequently adopted to study the cyclical dynamics and their sources of variations in developed economies.

We retain the nominal and real frictions originally present in the model, but we make a number of simplifications, which reduce the computational burden of the experiment, but have no consequences on the conclusions we reach. First, we assume that all exogenous shocks are stationary. Since we are working with the decision rules of the model, such a simplification involves no loss of generality. The sensitivity of our conclusions to the inclusion of trends in the disturbances is discussed in the on-line appendix. Second, we assume that all the shocks have an autoregressive representation of order one. Third, we compute the solution of the model around the steady state.

The model features a large number of shocks and this makes the number of observable variables equals the number of exogenous disturbances. Several researchers (for example, Chari, et al., 2009, or Sala, et al., 2010) have noticed than some of these shocks have dubious economic interpretations - rather than being structural they are likely to capture potentially misspecified aspects of the model. Relative to the SW model, we turn off the price markup, the wage markup and the preference shocks, which are the disturbances more likely to capture these misspecifications (see e.g., Chang et al., 2013), and we consider a model driven by technology, investment specific, government and monetary policy shocks, i.e. $(\epsilon_t^a, \epsilon_t^i, \epsilon_t^g, \epsilon_t^m)$. The sensitivity of the results to changes in the type of shocks we include in the model is described in the on-line appendix. The vector of endogenous controls coincides with the SW choice of measurable quantities; thus, we need to select four observables among output y_t, consumption c_t, investment i_t, wages, w_t, inflation π_t, interest rate r_t and hours worked h_t.

The log-linearized optimality conditions are in table 1 and our choices for the θ_0 vector are in table 2. Basically, the parameters used are the posterior estimates reported by SW, but any value would do it and the statistics of interest can be computed, for example, conditioning on prior mean values. Since there are parameters which are auxiliary, e.g. those describing the dynamics of the exogenous processes, while others have economic interpretations, e.g. price indexation or the inverse of Frisch elasticity, we focus on a subset of the latter ones when computing elasticity measures.

To construct the convoluted likelihood we need to choose the variance of the convolution error. We set $\Sigma_u = \kappa * I$, where κ is the maximum of the diagonal elements of $\Sigma(\theta)$, thus insuring that u_t and e_t have similar scale. When constructing the ratio $p_t^j(\theta)$ we simulate 500 samples and average the resulting $p_t^j(\theta, e_t^{t-1}, u_t)$. We also need to select a sample size when computing $p_t^j(\theta, e_t^{t-1}, u_t)$. We set $T = 150$, so as to have a data set comparable to those available in empirical work. We comment on what happens when a different κ is used and when $T = 1500$ in the on-line appendix.

We also need to set the size of the step when compute the numerical derivatives of the objective function with respect to the parameters - this defines the radius of the neighborhood around which we measure parameter identifiability. Following Komunjer and Ng (2011), we set g=0.01. When computing the rank of the spectral density, we also need to select the "tolerance level" for computing the rank of a matrix, which we set it equal to the step of the numerical derivatives, r=g= 0.01. The on-line appendix examines the sensitivity of the results with respect to the choice of g and r.

3.1 The results of the rank analysis

The model features 29 parameters, 12 predetermined states and four structural shocks. Thus, the condition for identification of all structural parameters is that the rank of $\Delta(\theta_0)$ is equal to 189. We start with an unrestricted specification and ask whether there exists four dimensional vectors that ensure full identifiability and, if not, what combination gets 'closest' to meet the rank condition. The number of combinations of four observables out of a pool of seven endogenous controls is $\left(\begin{array}{c}7\\4\end{array}\right) = \frac{7!}{4!(7-4)!} = 35$.

The first column of table 3 presents a subset of the 35 combinations and the second column the rank of $\Delta_j(\theta_0)$. Clearly, no combination guarantees full parameter identification. Hence, our rank analysis confirms a well known result (see e.g. Iskrev, 2010, Komunjer and Ng, 2011, Tkachenko and Qu, 2012) that the parameter vector of the SW model is not identifiable. Interestingly, the combination containing (y, c, i, w), has the largest rank, 186. Moreover, among the 15 combinations with largest rank, investment appears in 13 of them. Thus, the dynamics of investment are well identified and this variable contains useful identification information for the structural parameters. Conversely, real wages appears often in low rank combinations suggesting that this variable has relatively low identification power. Among the large rank combinations the nominal interest rate appears more often than inflation (7 vs. 4). More striking is the result that identification is poor when both inflation and interest rate are among the observables; indeed, all combinations featuring these two variables are in the low rank region and four have the lowest rank.

The third column of table 3 repeats the exercise calibrating some of the structural parameters. It is well known that certain parameters cannot be identified from the dynamics of the model (e.g. average government expenditure to output ratio) and others are implicitly selected by statistical agencies (e.g. the depreciation rate of capital). Thus, we fix the depreciation rate, $\delta = 0.025$, the good markets and labor market aggregators, $\varepsilon_p = \varepsilon_w = 10$, elasticity of substitution labor, $\lambda_w = 1.5$, and government consumption share in output $cg = 0.18$, as in SW (2007). Even with these five restrictions, the remaining 24 parameters of the model fail to be identified for any combination of the observable variables. While these five restrictions are necessary to make the mapping from the deep parameters to the reduced form parameters invertible, i.e. $\text{rank}(\Delta_\Lambda(\theta_0)) = n_\theta - 5 = 24$, they are not sufficient to guarantee local identification. Note that the ordering obtained in the unrestricted case is preserved, but different combinations of variables have now more similar ranks.

Finally, we examine whether there are parameter restrictions that allow some nonsingular system to identify the remaining vector of parameters. We proceed in two steps. First, we consider adding one parameter restriction to the five restrictions used in column 3. We report in column 4 the restriction that generates identification for each combination of observables. A blank space means that there are no parameter restrictions able to generate full parameter identification for that combination of observables. Second, we consider whether "any" set of parameter restrictions generate full identification; that is, we search for an 'efficient' set of restrictions, where by efficient we mean a combination of four observables that generates identification with a minimum number of restrictions. The fifth column of table 3 reports the parameters restrictions that achieve identification for each combination of observables.

From column 4 one can see that an extra restriction is not enough to achieve full parameter identification in all cases. In addition, the combinations of variables which were best in the unrestricted calculation are still those with the largest rank in this case. Thus, when the SW restrictions are used and an extra restriction is added, large rank combination generate identification, while for low rank combinations one extra restriction is insufficient. Interestingly, for most combinations of observables, the parameter that has to be fixed is elasticity of capital utilization adjustment costs. Column 5 indicates that at least four restrictions are need to identify the vector of structural parameters and that the goods and labor market aggregators, ε_p and ε_w, cannot be estimated either individually or jointly for any combination of observables. Thus, the largest (unrestricted) rank combinations are more likely to produce identification with a tailored use of parameter restrictions.

The first four static principal components of the seven variables track very closely the (y, c, i, w) combination and efficient identification requires the same restrictions. Dynamic principal components appear to be poorer and they seem to span the space of lower rank combinations. Thus, for this system, there seems to be little gain in using principal components rather than observable variables.

3.2 The results of the elasticity analysis

As mentioned, the rank analysis is unsuited to detect weak and partial identification problems that often plague the estimation of the structural parameters of the DSGE model. To investigate potential weak and partial identification issues we compute the curvature of the convoluted likelihood function of the singular and non-singular systems and examine whether there are combinations of observables which have good rank properties and also avoid flatness and ridges in the likelihood function.

Table 4 presents the four best combinations minimizing the "elasticity" distance.

We focus attention on six parameters, which are often the object of discussion among macroeconomists: the habit persistence, the inverse of the Frisch elasticity of labor supply, the price stickiness and the price indexation parameters, the inflation and output coefficients in the Taylor rule. Notice first, that the format of the objective function is irrelevant: the top combinations are also the best according to the second criterion. Also, by comparing tables 3 and table 4, it is clear that maximizing the rank of $\Delta_j(\theta_0)$ does not necessarily make the curvature of the convoluted likelihood in the singular and non-singular system close in these six dimensions. The vector of variables which is best according to the "elasticity" criterion (consumption, investment, hours and the nominal interest rate) was in the second group in table 3 but the top combination in that table ranks close second. In general, the presence of the nominal interest rate helps to identify the habit persistence and the price stickiness parameters; excluding the nominal rate and hours in favor of output and the real wage (the second best combination) helps to better identify the price indexation parameter at the cost of making the identifiability of the Frisch elasticity worse. Interestingly even the best combination of variables makes the curvature of likelihood quite flat, for example, in the dimension represented by the inflation coefficient in the Taylor rule. Thus, while there does not exist a combination which simultaneously avoids weak identification in all six parameters, different combinations of observables may reduce the extent of the problem in certain parameters. Hence, depending on the focus of the investigation, researchers may be justified in using different vectors of the observables to estimate the structural parameters.

It is worth also mentioning that while there are no theoretical reasons to prefer any two variables among output, hours and labor productivity, and the ordering the best models is unaffected in the rank analysis, there are important weak identification trade-offs in selecting a group of variables or the other. For example, comparing figures 1 and 2, one can see that if labor productivity is used in place of hours, the flatness of the likelihood function in the dimensions represented by the inflation and the output coefficients in the Taylor rule is reduced, at the cost of worsening the identification properties of the habit persistence and the price stickiness parameters.

3.3 The results of the information analysis

Table 5 gives the best combinations of observables according to the information statistic (11). As in table 4, we also provide the value of the average objective function for that combination relative to the best.

An econometrician interested in estimating the structural parameters of this model should definitely use output, consumption and investment as observables - they appear in all four top combinations. The fourth observable seems to be either hours or real wages, while combinations which include interest rates or inflation fare quite poorly in terms of relative informativeness. In general, the performance of alternative combinations deteriorates substantially as we move down in the ordering, suggesting that the information measure can sharply distinguish various options.

Interestingly, the identification and the informational analyses broadly coincide in the ordering of vectors of observables: the top combination obtained with the rank analysis (y, c, i, w) fares second in the information analysis and either second or third in the elasticity analysis. Moreover, three of the four top combinations in table 5 are also among the top combinations in table 3. Finally, note that also in this case the performance of the first four static principal components is very similar to the one of the (y, c, i, w) vector and that dynamic principal components appear to be inferior to their static counterparts.

3.4 Summary

To estimate the structural parameters of this model it is necessary to include at least three real variables and output, consumption and investment seem the best for this purpose. The fourth variable varies according to the criteria. Despite the monetary nature of this model, jointly including inflation and the nominal rate among the observables make things worse. We can think of two reasons for this outcome. First, because the model features a Taylor rule for monetary policy, inflation and the nominal rate tend to comove quite a lot. Second, since the parameters entering the Phillips curve are difficult to identify no matter what variables are employed, the use of real variables at least allows us to pin down intertemporal and intratemporal links, which crucially determine income and substitution effects present in the economy.

As the on-line appendix shows, changes in nuisance parameters present in the two procedures, in the sample size, in the choice of shocks of entering the model and in the specifications of the informational distance do not affect these results.

4 How different are the specifications?

To study how different the "best"and the "worst"specifications are in practice, we generate 150 data points for output (y), consumption (c), investment (i), wages (w), hours worked (h), inflation (π) and interest rate (r) using the SW model driven by four structural shocks ($\epsilon_t^a, \epsilon_t^i, \epsilon_t^g, \epsilon_t^m$) and parameters as in table 2. We then estimate the structural parameters of the following five models:

- Model A: Four structural shocks and (y, c, i, w) as observables (this is the best combination of variables according to the rank analysis).
- Model B: Four structural shocks and (y, c, i, h) as observables (this is the best combination of variables according to the information analysis).
- Model Z: Four structural shocks and (c, i, π, r) as observables (this the worst combination of variables according to the rank analysis).
- Model C: Four structural shocks, three measurement errors, attached to output, interest rates and hours and all seven observable variables.
- Model D: Seven structural shocks - the four basic ones plus price markup, wage markup and preference shocks, all assumed to be iid - and all seven observable variables

and compute responses to interesting shocks. We want to see i) how the best models (A and B) fare relative to the DGP and to the worst model (Z); ii) how standard alternatives augmenting the true set of shocks with either artificial measurement errors (C) or with artificial structural errors (D), fare in comparison to the best models and the DGP. In ii) we are particularly interested in whether the presence of three artificial shocks distorts the responses to true disturbances and in whether responses to these artificial shocks display patterns that could lead investigators to confuse them with the structural disturbances.

The likelihood of the simulated sample is combined with the prior distribution of the parameters to obtain the posterior distribution in each case. The choice of priors closely follows SW(2007) and posterior distributions are obtained using two independent chains of 100,000 draws using the MH algorithm. Table 6 presents the true parameter values and the vector of highest posterior 90 % credible sets for the five models. For model A,B and Z 23 structural parameters are estimated. For model C, we estimate 22 structural parameters (the wage stickyness parameters is kept fixed) and the standard deviation of the three measurement errors; for model D we estimate 20 structural parameters (i.e. the monetary policy coefficients are set to their true values) and the standard deviation of the artificial preference, the price and the wage markups shocks [5].

The table confirms that models A and B are different from model Z. Even if the three models are all correctly specified in terms of structure, there are sizable differences in the magnitude and the precision of credible sets. In models where the observables feature large spectral ranks or high information, estimates of the structural parameters are typically more accurate. For example, in Models A and B there are four credible sets that do not include the true parameter values, while in model Z nine credible sets fail to include the true parameter values. In addition, in model Z important objects, such as the price stickiness and the price indexation parameters, which crucially determine the slope of the price Phillips curve, are poorly estimated.

It is worth mentioning that in model Z, which uses (c, i, π, r) as observables in estimation, the government process is very poorly estimated: both the autoregressive coefficient and the standard deviation of the process are underestimated. One reason for this outcome is that, in the DGP, g_t enters only in the feasibility constraint. In model Z we only have information about c_t and i_t, which is insufficient to precisely disentangle g_t from y_t. This misspecification has important consequences for the transmission of government expenditure shocks. Figure 3 displays the responses of the seven variables to a positive spending impulse. Stars represent the 90 percent credible sets in Model A, red circles the 90 percent credible sets of Model Z and the black solid line the true responses. While Model A correctly identifies the propagation mechanism of a spending shock, in the model Z the shock has almost no impact on the variables and the true response is never contained in the credible sets we present.

[5] We fix some of the parameters at the true values in models C and D since they turned out to be poorly identified and the MCMC routine encountered numerical difficulties.

Estimates of the parameters of models C and D are characterized by different degrees of misspecification. In model C, we have added (non-existent) measurement error to output, hours worked and the nominal interest rate. Since measurement error is iid and the structural model is correctly specified, one would expect this addition not to make a huge difference in terms of parameter estimates. In model D, we have added white noise structural shocks to the dynamics of the original data generating process. While the shocks that perturb the price Phillips curve, the wage Phillips curve and the Euler equation are iid, the structure we estimate is misspecified, making estimates of the structural parameters potentially biased.

Table 6 suggests that distortions are present in both setups but larger in model D. For example, the posterior sets do not include the true parameters in 13 out of 22 cases for model C; in 14 out of 20 cases in model D. Furthermore, posterior credible sets are tight thus incorrectly attributing large informational content to the likelihood. Hence, augmenting the original model with measurement or structural shocks to employ more variables in estimation, does not seem to help to produce more accurate estimates of the structural parameters.

Why are there distortions? First, while the estimated standard deviation of the three additional shocks is low compared to the standard deviation of the original shocks, it is a-posteriori different from zero. Thus, while the estimation procedure recognizes that these shocks have smaller importance relative to the four original shocks, it wants to give them a role because the prior heavily penalizes their non-existence. Second, the significance of artificial shocks implies that the properties of other shocks are misspecified. For example, in model C, the standard deviation of the technology disturbance is underestimated. Figure 4, which presents the responses to a technology shock in the true model (black line) and the highest 90 percent credible sets in Model B (blue stars) and C (red circles), shows that also the transmission mechanism is altered.

The situation is worse when the estimated model features structural shocks that the true model does not possess. Figure 5 gives a glimpse of what may happen in this case. First, notice that the responses of the variables of the system to a (non-existent) price markup shock are small but a-posteriori significant. Second, the responses have the same shape (but different magnitude) as those that would be estimated in case price markup shocks were truly a part of the DGP (compare dotted and solid lines). Thus, it is possible to obtain perfectly reasonable patterns of responses even though the shocks which are supposed to drive them are not present in the DGP and the patterns look very much like those one would obtain if the shocks where present. This conclusion holds also for the other two shocks, which we have erroneously added in Model D.

Finally, we would like to mention that despite the fact that neither Model A nor Model B uses any nominal variables in estimation, the responses to a monetary shocks produced by these model match very well those of the DGP (see on-line appendix). Hence, we confirm that capturing the intertemporal and the intratemporal links present in the model is enough to have the dynamics of the endogenous variables in response to all the shocks are also well captured.

Overall, it seems a bad idea to add measurement errors to the model to be able to use more variables in the estimation. Relative to a setup where a reduced number of variables is chosen in some meaningful way, impulse responses are tighter but strictly more inaccurate. Similarly, it seems far from optimal to complete the probability space of the model by artificially inserting structural shocks. Given standard prior restrictions, their presence will distort parameter estimates and impulse responses in two ways: they will take away importance from the true shocks; they will have responses which will look reasonable even if their true effects are zero. In this sense, our conclusions echo results derived by Cooley and Dweyer (1998) in a different framework.

To conclude, we would like to mention that the results obtained in this section are conditional on one particular vector of time series generated by the model. To check whether the conclusions hold when sampling uncertainty is taken into account, we have conducted also a small Monte Carlo exercise where for each model estimation is repeated on 50 different samples and credible sets are constructed using 90 percent of the posterior median estimates. Results are unchanged.

5 Conclusions and practical suggestions

This paper proposes criteria to select the observable variables to be used in the estimation of the structural parameters when one feels uncomfortable in having a model driven by a large number of potentially non-structural shocks or does not have good reasons to add measurement errors to the decision rules, and insists in working with a singular DSGE model. The methods we suggest measure the identification and the information content of vectors of observables, are easy to implement, and seem able to effectively rank combinations of variables. Interestingly, and despite the fact that the statistics we employ are derived from different principles, the best combinations of variables these methods deliver are pretty much the same.

In the model we consider, parameter identification and variable informativeness are optimized including output, consumption and investment among the observables. These variables help to identify the intertemporal and the intratemporal links in the model and thus are useful to correctly measure income and substitution effects, which crucially determine the dynamics of the model in response to the shocks. Interestingly, using interest rate and inflation jointly in the estimation makes identification worse and the loss of information due to variable reduction larger. When one takes the curvature of the likelihood into consideration, the nominal interest rate is weakly preferable to the inflation rate.

We also show that, in terms of likelihood curvature, there are important trade-offs when deciding to use hours or labor productivity together with output among the observables and demonstrate that changes in the setup of the experiment do not alter the main conclusions of the exercise.

The estimation exercise we perform indicates that the best models our criteria select capture the conditional dynamics of the singular model reasonably well while the worst models do not. Furthermore, the practice of tagging-on measurement errors or non-existent structural shocks to use a larger number of observables in estimation may distort parameter estimates and jeopardize inference.

While our conclusions are sharp, an econometrician working in a real world application should certainly consider whether the measurement of a variables is reliable or not. Our study only asks what set of observables is preferable, when a singular model is assumed to be the DGP. In practice, the analysis can be undertaken also when some justified measurement error is preliminarily added to the model.

In designing criteria to select the variables for estimation, we have taken as given that researchers have a set of shocks they are interested in studying. One may also consider the alternative of a researcher with no strong a-priori ideas about which disturbances the theory should specify. In this case, our variable selection procedures can be nested in a more general approach which would involve taking a vector of data, characterizing the principal components of the one-step ahead prediction error and selecting those explaining a certain prespecified variance of the data (as in Andrle, 2012).

Then, one would perform prior predictive analysis to select the theoretical shocks that are more likely to generate the second order properties produced in the data by the principal components of the shocks one has selected. Once this is done, our procedures can then be applied to select the endogenous variables used for estimation, given the empirically selected vector of structural shocks.

Our selection criteria implicitly assume that all variables are equally relevant from an economic point of view. That may not always be the case and one may have a set of core variables and a set of ancillary variables, potentially relevant to characterize a phenomena. For example, in a model featuring macro-financial linkages, the macro variables could be held fixed and one may want to choose the vector of financial variables that best inform researchers on this link. In this situation, our selection criteria can be used to select relevant variables from the latter set.

The approaches are designed with the idea that a researcher wants to use the likelihood function for inferential purposes. If this is not the case, the spectral methods of Qu and Tkachenko (2011) can be employed to estimate the structural parameters, since the spectral density is well defined object that can be optimized, even in a singular system.

One way of interpreting our exercises is in terms of prior predictive analysis (see Faust and Gupta, 2011). In this perspective, prior to the estimation of the structural parameters, one may want to examine which features of the model are well identified and what is the information content of different vector of observables. Seen through these lenses, the analysis we perform complements those of Canova and Paustian (2011) and of Mueller (2010).

References

An, S. and Schorfheide, F. , 2007. Bayesian analysis of DSGE models. Econometric Reviews, 26, 113-172.

Andrle, M., 2012. Estimating DSGE models using likelihood based dimensionality reduction. IMF manuscript.

Bierens, H., 2007. Econometric analysis of linearized DSGE models. Journal of Econometrics, 136, 595-627.

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

Canova, F. and Paustian, M., 2011. Business cycle measurement with some theory. Journal of Monetary Economics, 58, 345-361.

Chari, V.V., Kehoe, P. and McGrattan, E., 2009. New keynesian models: Not yet useful for policy analyses. American Economic Journals: Macroeconomics, 1, 242-266.

Chang, Y. , S. Kim and Schorfheide, F., 2013. Labor market heterogeneity and the Lucas critique, Journal of the European Economic Association, 11, 193-220.

Cochrane, J., 2011. Determinacy and identification of Taylor rules. Journal of Political Economy, 119, 565-615.

Cooley, T. and Dweyer, M., 1989. Business cycle analysis without too much theory: A look at structural VARs. Journal of Econometrics, 83, 57-88.

Del Negro, M. and Schorfheide, F., forthcoming. DSGE model-based forecasting, Handbook of Economic Forecasting, volume 2.

Faust, J., and Gupta, A. , 2011. Posterior predictive analysis for evaluating DSGE models. John Hopkins University manuscript.

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

Kascha, C. and Mertens, K., 2010. Business cycle analysis and VARMA. Journal of Economic Dynamics and Control, 33, 267-282.

Kleibergen, F. and Mavroedis, S., 2009. Weak instrument robust tests in GMM and the New Keynesian Phillips curve. Journal of Business and Economic Statistics, 27, 293-311.

Komunjer, I. and Ng, S., 2011 Dynamic identification of DSGE models. Econometrica, 79, 1995-2032.

Iskrev, N., 2010. Local identification in DSGE Models. Journal of Monetary Economics, 57, 189-202.

Mavroedis, S., 2005. Identification issues in forward looking models estimated by GMM with an application to the Phillips curve. Journal of Money, Credit and Banking, 37, 421-448.

Mueller, U., 2010. Measuring prior sensitivity and prior informativeness in large Bayesian models. Princeton University manuscript.

Qu, Z. and Tkachenko, D., 2012. Identification and frequency domain QML estimation of linearized DSGE models. Quantitative Economics, 3, 95-132.

Sala, L., Soderstrom, U., and Trigari, A., 2010. Output gap, the labor wedge, and the dynamic behavior of hours. IGIER working paper.

Smets, F. and Wouters, R., 2007. Shocks and frictions in US business cycles: a Bayesian approach. American Economic Review, 97, 586-606.

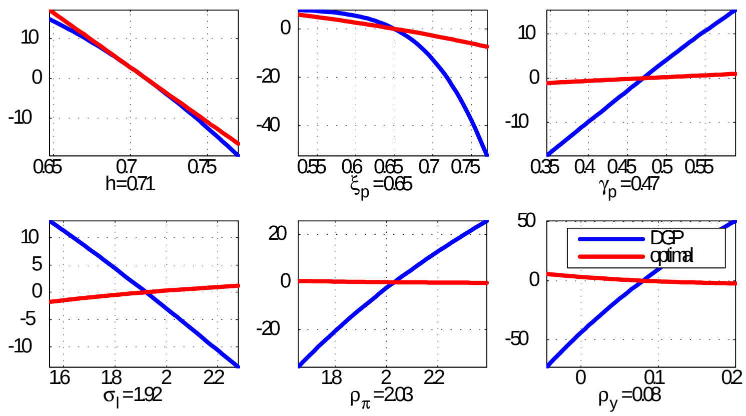

Figure 1: One dimensional convoluted likelihood of the DGP and of (y_t, c_t, r_t, h_t).

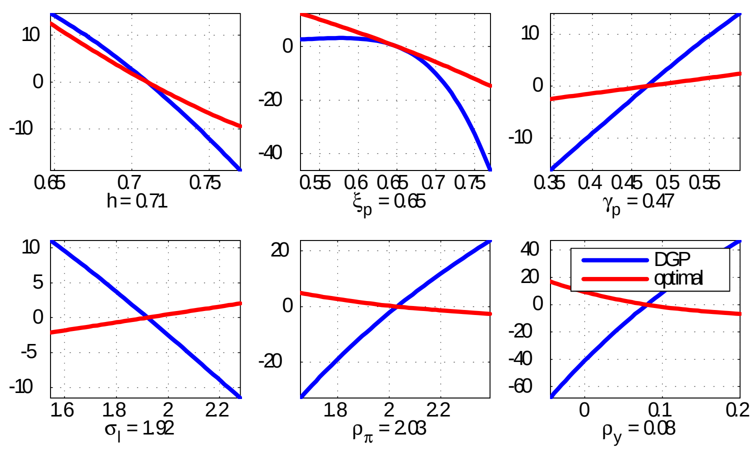

Figure 2: One dimensional convoluted likelihood of the DGP and of $(y_t, c_t, r_t, \frac{y_t}{h_t})$.

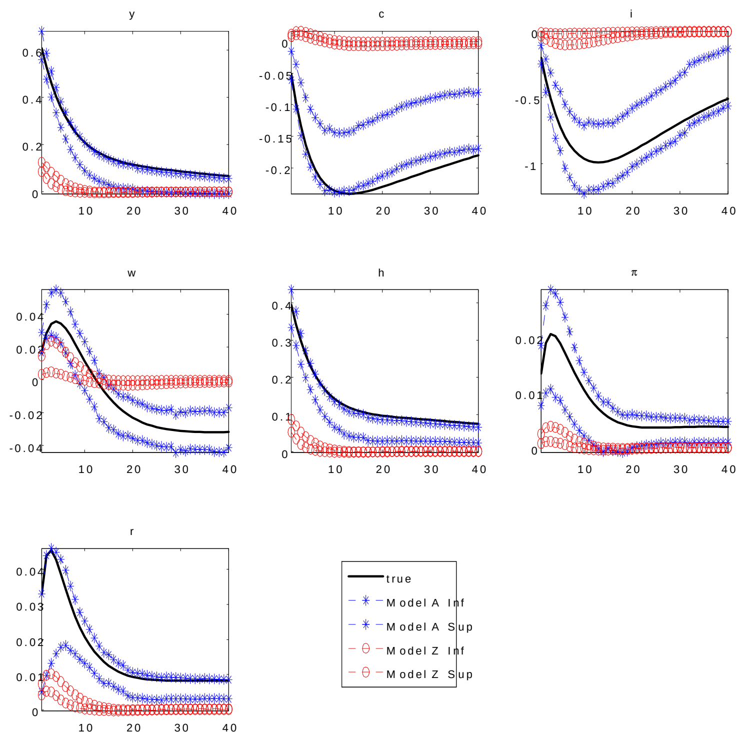

Figure 3: Impulse response to a government spending shock. Stars represent highest 90 percent credible sets for Model A; red circles the highest 90 percent credible sets for Model Z; the black solid line the true impulse response. From top left to bottom right, are the response of output y, consumption c, investment i, real wage w, hours h, inflation π and nominal interest rate r.

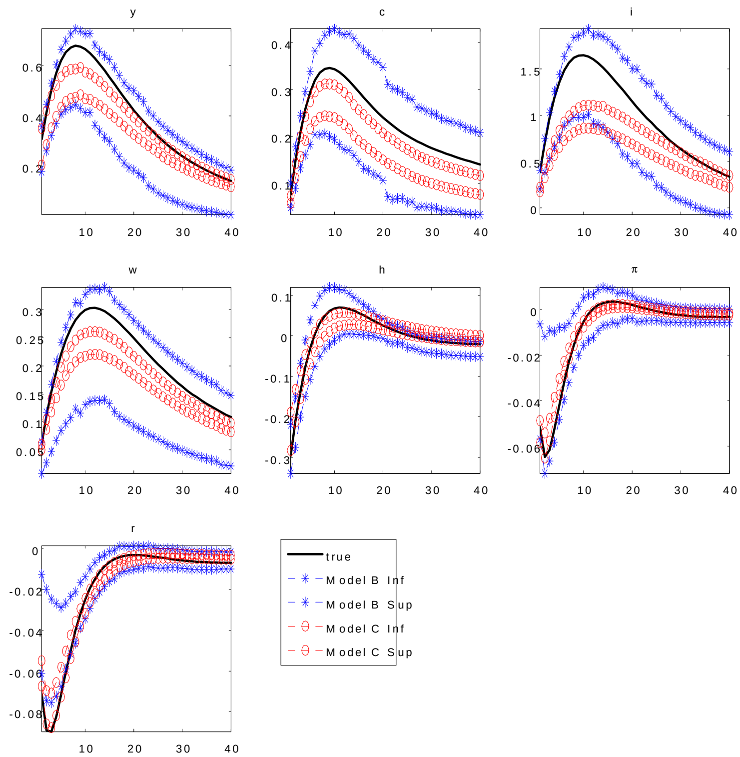

Figure 4: Impulse response to a technology shock. Stars represent the highest 90 percent credible sets for Model B; red circles the highest 90 percent credible sets for Model C; the black solid line the true impulse response. From top left to bottom right, are the response of output y, consumption c, investment i, real wage w, hours h, inflation π and nominal interest rate r.

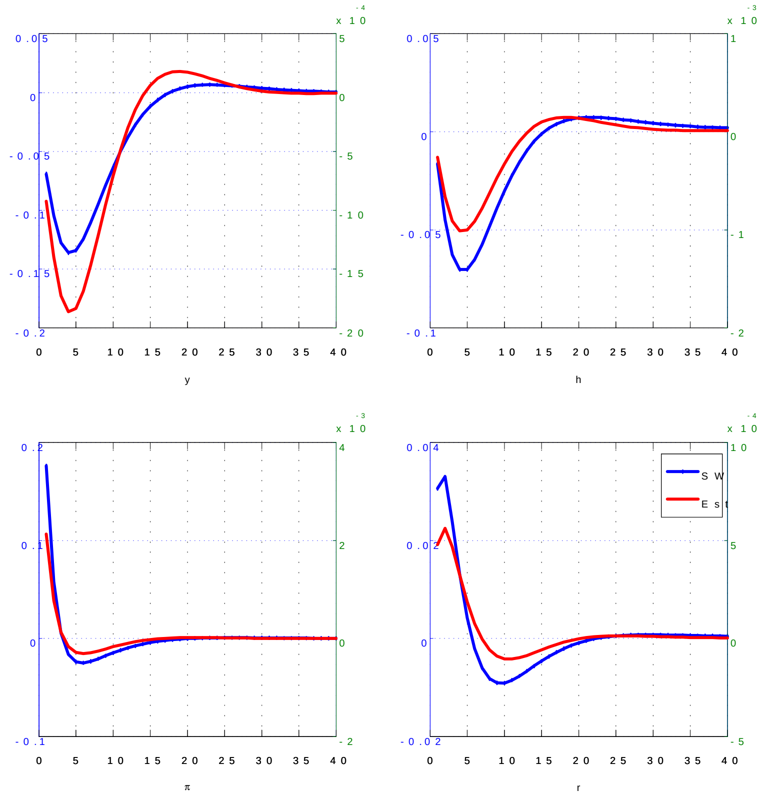

Figure 5: Impulse response to a price markup shock. Stars represent the estimated responses in a SW model when there are 7 structural shocks; dotted lines the estimated responses in a SW model with four structural shocks. From top left to bottom right, are the response of output y, hours h, inflation π, and nominal interest rate r.

y_t	$=\alpha\phi_p k_t + (1-\alpha)\phi_p h_t + \phi_p \epsilon_t^a$
y_t	$=\epsilon_t^g + c/y c_t + i/y i_t + r^k k/y z_t$
k_t	$= k_{t-1}^s + z_t$
k_t	$= \omega_t + h_t - \frac{\psi}{1-\psi} z_t$
mc_t	$= \alpha \frac{\psi}{1-\psi} z_t + (1-\alpha)\omega_t - \epsilon_t^a$
k_t^s	$= (1-\delta)k_{t-1}^s + i/k\ i_t + i/k\ \varphi \epsilon_t^i$
π_t	$= \frac{\beta}{1+\beta i_p} E_t \pi_{t+1} + \frac{i_p}{1+\beta i_p}\pi_{t-1} + \frac{k_p}{1+\beta i_p} mc_t + \epsilon_t^p$
ω_t	$= \frac{1}{1+\beta i_\omega}\omega_{t-1} + \frac{\beta}{1+\beta i_\omega} E_t \left(\omega_{t+1} + \frac{1}{1+\beta i_\omega}\pi_{t+1}\right) + \frac{i_\omega}{1+\beta i_\omega}\pi_{t-1} + \pi_t - \frac{k_\omega}{1+\beta i_\omega}\omega_t^\mu + \epsilon_t^\omega$
ω_t^μ	$= \omega_t - \left(\sigma_n h_t + \frac{1}{1+h}(c_t - hc_{t-1})\right)$
c_t	$= \frac{1}{1+h}(E_t c_{t+1} - hc_{t-1}) + c_1(h_t - E_t h_{t+1}) - c_2(r_t - E_t \pi_{t+1}) + \epsilon_t^b$
q_t	$= -(r_t - E_t \pi_{t+1}) + \frac{\sigma_c(1+h)}{(1-h)}\epsilon_t^b + E_t(q_1\ z_{t+1} + q_2\ q_{t+1})$
i_t	$= \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} E_t i_{t+1} + \frac{1}{\varphi(1+\beta)} q_t + \epsilon_t^i$
r_t	$= \rho_R r_{t-1} + (1-\rho_R)(\rho_\pi \pi_t + \rho_y y_t + \rho_{\Delta y}\Delta y_t) + \nu_t^r$

Table 1: Log-linear equations, Smets and Wouter (2007) model. Variables without the time subscript are steady state values, variable with time subscript are deviation from the steady state. $k_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p((\phi_p-1)e_p+1)}$, $k_\omega = \frac{(1-\zeta_\omega)(1-\zeta_\omega\beta)}{\zeta_\omega((\phi_\omega-1)e_\omega+1)}$, $c_1 = \frac{(\sigma_c-1)\omega^h n/c}{\sigma_c(1+h)}$ and $c_2 = \frac{1-h}{\sigma_c*(1+h)}$, $q_1 = \frac{r^k}{r^k+1-\delta}\frac{\psi}{1-\psi}$ and $q_2 = \frac{1-\delta}{r^k+1-\delta}$. In the version of model we consider $\epsilon_t^a = \rho_a\ \epsilon_{t-1}^a + \nu_t^a$ (technology), $\epsilon_t^i = \rho_i \epsilon_{t-1}^i + \nu_t^i$ (investment specific), $\epsilon_t^r = \nu_t^r$ (Taylor rule), $\epsilon_t^g = \rho_g \epsilon_{t-1}^g + \nu_t^g + \rho_{ga}\nu_t^a$ (government spending), $\epsilon_t^\omega = 0$ (wage markup) and $\epsilon_t^p = 0$ (price markup),$\epsilon_t^b = 0$ (preference).

θ	Description	Value
δ	depreciation rate	0.025
ε_p	good markets kimball aggregator	10
ε_w	labor markets kimball aggregator	10
λ_w	elasticity of substitution labor	1.5
cg	gov't consumption output share	0.18
β	time discount factor	0.998
ϕ_p	1 plus the share of fixed cost in production	1.61
ψ	elasticity capital utilization adjustment costs	5.74
α	capital share	0.19
h	habit in consumption	0.71
ζ_ω	wage stickiness	0.73
ζ_p	price stickiness	0.65
i_ω	wage indexation	0.59
i_p	price indexation	0.47
σ_n	elasticity of labor supply	1.92
σ_c	intertemporal elasticity of substitution	1.39
φ	st. st. elasticity of capital adjustment costs	0.54
ρ_π	monetary policy response to π	2.04
ρ_R	monetary policy autoregressive coeff.	0.81
ρ_y	monetary policy response to y	0.08
$\rho_{\Delta y}$	monetary policy response to y growth	0.22
ρ_a	technology autoregressive coeff.	0.95
ρ_g	gov spending autoregressive coeff.	0.97
ρ_i	investment autoregressive coeff.	0.71
ρ_{ga}	cross coefficient tech-gov	0.52
σ_a	sd technology	0.45
σ_g	sd government spending	0.53
σ_i	sd investment	0.45
σ_r	sd monetary policy	0.24

Table 2: Parameters description and values used in the DGP.

	Unrestricted Rank(Δ)	Restricted Rank(Δ)	Resticted and Restriction on	Efficient Restrictions Four parameters fixed, ε_p, ε_w and
y,c,i,w	186	188	ψ	(λ_w,ψ), (ϕ_p,ψ), (ψ,ζ_ω), (ψ,ζ_p), (ψ,σ_n), (ψ,σ_c), (ψ,ρ_π), (ψ,ρ_y)
y,c,i,π	185	188	ψ	(ψ,ϕ_p), (ψ,ζ_ω), (ψ,ζ_p), (ψ,σ_n)
y,c,r,h	185	188	ψ	(ψ,ζ_p), (ψ,i_ω), (ψ,ρ_π), (ψ,ρ_y), (ζ_p,σ_c), (i_ω,σ_c), (σ_c,ρ_π), (σ_c,ρ_y)
y,i,w,r	185	188	ψ	(λ_w,ψ), (ψ,ζ_ω), (ψ,ρ_y)
c,i,w,h	185	188	ψ,σ_c,ρ_i	(λ_w,ψ), (ψ,ζ_ω), (ψ,ρ_y)
c,i,π,h	185	188	ψ	(λ_w,ψ), (cg,ψ), (ψ,ζ_ω), (ψ,σ_c)
c,i,r,h	185	188	$\zeta_\omega,\zeta_p,i_\omega$	(λ_w,ψ), (cg,ψ), (ψ,σ_n), (ψ,ζ_ω), (ψ,σ_c)
y,c,i,r	185	187		(λ_w,ψ), (cg,ψ), (ψ,ζ_ω), (ψ,σ_c)
y,c,i,h	185	187		(λ_w,ψ), (cg,ψ), (ψ,ζ_ω), (ψ,σ_c)
i,w,r,h	185	188	ψ	(λ_w,ψ), (cg,ψ), (ψ,ζ_ω), (ψ,σ_c)
y,i,w,h	185	188	ψ	
y,i,π,h	185	188	ψ	
y,i,r,h	185	188	ψ,ρ_i	
y,c,w,r	185	188	ψ	
y,i,w,π	185	188	ψ	y,i,π,r
i,π,r,h	184	188	ψ	
c,w,r,h	184	188	ψ	
y,c,w,π	184	187		
y,c,w,h	184	187		(ζ_p,σ_c), (cg,ψ), (ϕ_p,ψ), (cg,σ_c), (ϕ_p,σ_c), (ψ,ζ_p)
y,c,π,r	184	187		
y,c,π,h	184	187		(ϕ_p,ψ), (cg,ψ)
y,w,π,r	184	187		(cg,ζ_ω), (ϕ_p,ψ), (ϕ_p,ζ_ω), $(\psi,\rho_{\Delta y})$
y,w,π,h	184	187		
y,w,r,h	184	187		
y,π,r,h	184	187		
c,i,w,π	184	187		
c,i,w,r	184	188		
c,π,r,h	184	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		
Required	189	189		
Static PC	186	184	ψ	(λ_w,ψ), (ϕ_p,ψ), (ψ,ζ_ω), (ψ,ζ_p), (ψ,σ_n), (ψ,σ_c), (ψ,ρ_π), (ψ,ρ_y)
Dynamic PC	184	188	ψ	

Table 3: Rank conditions for combinations of observables 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 $cg = 0.18$ are fixed. The fourth columns reports the extra parameter restriction needed to achieve full parameter identification; a blank space means that there are no parameter restrictions able to guarantee identification. The last column reports the efficient restrictions that generates identification, in addition to fixing $(\varepsilon_p, \varepsilon_w)$.

Order	Cumulative Deviation	Ratio	Weighted Square	Ratio
1	(c, i, r, h)	1.00	(c, i, r, h)	1.00
2	$(y.c, i, w)$	1.49	(c, i, w, h)	1.65
3	(c, i, w, h)	1.87	(y, c, i, w)	1.91
4	(y, c, r, h)	2.04	(y, c, r, h)	2.12

Table 4: 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.

Order	Combination	Relative Information
1	(y, c, i, h)	1
2	(y, c, i, w)	0.89
3	(y, c, i, r)	0.52
4	(y, c, i, π)	0.5
	PC static	0.84
	PC dynamic	0.65

Table 5: Ranking based on the $p(\theta)$ statistic. Relative information is the ratio of the $p(\theta)$ statistic relative to the statistic obtained for the best combination.

Parameter	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)

Table 6: True parameter values and highest posterior 90 percent credible sets for the common structural parameters of the five models.