Choosing the variables to estimate singular DSGE models

Fabio Canova
EUI and CEPR

Filippo Ferroni *
Banque de France, Univ Surrey

Christian Matthes
UPF

August 5, 2012

Very Preliminary, please do not quote

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 is to a standard model discussed. Practical suggestions for applied researchers are provided.

Key words: Spectral density, Identification, density ratio, DSGE models.

JEL Classification:

---

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

1 Introduction

The structure of Dynamic Stochastic General Equilibrium (DSGE) models implies that the optimal decision rules typically have a singular format. This occurs because the number of endogenous variables generally exceeds the number of exogenous shocks. A stereotypical example is a basic RBC structure: since the model generates implications for capital, consumption, output, hours, real wages and the real interest rate, and only one shock drives the economy, both the short run dynamics and the long run properties of the endogenous variables are driven by a one dimensional exogenous process. Thus, the covariance matrix of the data produced by the model is singular and likelihood based estimation methods (both of classical or Bayesian inclinations) inapplicable.

The singularity problem can be mitigated if some of the endogenous variables are non-observables; for example, we rarely have data on the capital stock and, occasionally, data on hours is also unavailable. Non-observability of some endogenous variables reduces the extent of singularity problem 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. The introduction of measurement error helps to complete the probability space of the model thus lessening the singularity problem - the number of shocks driving a given number of observable variables is now larger. In general, however, neither non-observability of some endogenous variables nor the addition of justified measurement error is sufficient to completely eliminate the singularity problem. Thus, having more endogenous variables than shocks is an inherent feature of DSGE setups. While singularity is not troublesome for certain limited information estimation approaches, such as impulse response matching, it creates important headaches to researchers interested in estimating the structural parameters by full information likelihood methods.

Two approaches are generally followed in this situation. The first involves enriching the model with additional shocks so as to have as many exogenous shocks as endogenous variables (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 be important, say, to explain output or inflation fluctuations. The second is to transform the decision rules, solving out certain variables from the optimality conditions, until the number of endogenous variables is the same as the number of shocks. This approach is also problematic since the convenient state space structure of the decision rules is lost through this dimensionality reduction, the likelihood becomes an even more complicated function of the structural parameters and can not necessarily be computed with standard Kalman filter recursions. The alternative approach of tossing out certain equations characterizing the decision rules is also unsatisfactory since from a system 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 the combination of variables used in estimation are lacking.

Guerron Quintana (2010) has estimated a standard singular DSGE model, adding enough measurement errors to avoid singularity and using different observable variables, and showed that inference about important structural parameters may be whimsical. To decide which combination of variables should be used in estimation he suggests to use economic hindsight and an out-of-sample MSE criteria. 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 problem in DSGE estimation) are equally weighted.

This paper proposes two complementary and perhaps more relevant 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, our first criteria selects the variables 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 derivative of spectral density matrix of the vector of observables with respect to the parameters. These authors have shown that the stationary solution of a DSGE model has a ARMA representation and therefore a spectral density. Thus, the matrix of derivatives of the spectral density matrix with respect to the structural parameters can be linked to the matrix of derivatives of the ABCD representation of the state space. By examining the ABCD representations corresponding to different sets of observables one can measure local "identification distance" from some benchmark.

Given the ideal rank needed to achieve full identification of a parameter vector, the selected vector of minimizes the discrepancy between the ideal and the actual rank of the spectral matrix. We show how such an approach can be tailored to the question of selecting the restriction needed to make structural parameters identified and, given that a subset of parameters is typically calibrated, what additional restrictions efficiently allow identification of the remaining structural parameters.

The Komunjer and Ng approach is silent about the more subtle issues of weak and partial identification. To deal with these problems, we complement the rank analysis evaluating the difference in local curvature of the convoluted likelihood function of the singular system and of a number of non-singular alternatives. The combination of variables we select is the one making the curvature of the convoluted likelihood in the dimensions of interest as close as possible to the curvature of the convoluted likelihood of the singular system.

The second criteria employs the informational content of the densities of the singular and the non-singular system 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 compare the informational content of different vectors of observables, taking as given the structural parameters. Since the measure of informational distance we construct depends on nuisance parameters (the variance of the structural shocks and of the convolution errors), we take as our criterion function to be optimized the average ratio of densities, where averages are constructed integrating out nuisance parameters.

We apply the methods to select the vector of variables in a singular version of the model of Smets and Wouters (2007) (henceforth SW model). We retain the full structure of nominal and real frictions but we allow only four structural shocks - a technology, an investment specific, a monetary, a fiscal shock - to drive the economy. Since the model features seven observable endogenous variables, the framework can be used to study the identification and information properties of various combination of the endogenous variables.

Although monetary policy plays an important role in the economy, parameter identification and variable informativeness are optimized using only real variables in estimation and including output, consumption and investment seems always the best. These variables help to identify the intertemporal and the intratemporal links present in the economy and thus are useful to correctly measure income and substitution effects in the model. Interestingly, using interest rate and inflation jointly in the estimation makes identification worse and the loss of information due to variable reduction larger. Moreover, when one takes the curvature of the likelihood in the dimensions of interest into consideration, it seems preferable to include the nominal interest rate, rather than the inflation rate, among the vector of observables.

We show that the ordering of various combinations is broadly maintained when parameter restrictions are added to the unrestricted model. We also show that, at least in terms of likelihood curvature, there are important trade-off when deciding to use hours or labor productivity among the observables. Finally, we demonstrate that changes in the setup of the experiment do not alter the main important conclusions of the exercise.

The paper is organized as follows. The next section describes the methodologies used to select the combinations of variables. Section 3 applies the approaches to a singular version of the Smets and Wouter (2007) model. Section 4 analyzes robustness issues. Section 5 concludes providing some practical advice for applied researchers.

2 The selection procedures

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

$$x_t = A(\theta)x_{t-1} + B(\theta)e_t \tag{1}$$

$$y_t = C(\theta)x_{t-1} + D(\theta)e_t \tag{2}$$

$$e_t \sim N(0, \Sigma(\theta))$$

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

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

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 $n_e < n_y$.

From (3) one can generate a number of non-singular structures, using a subset j of endogenous controls $y_{jt} \subset y_t$, and making sure the dimension of the vector of observable variables and of the shocks coincides. Given (3), we 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

$$y_{jt} = \left(C_j(\theta) \left(I - A(\theta)L \right)^{-1} B(\theta)L + D_j(\theta) \right) e_t$$

where $C_j(\theta)$ and $D_j(\theta)$ are obtained selecting the rows corresponding to y_{jt}. The non-singular model j has also a MA(∞) representation, but now $D_j(\theta)e_t e_t' D_j(\theta)'$ has rank $n_e = n_y$.

To study the identification and informational content of various y_{jt} vectors, we examine the properties of the spectral density of y_{jt} and the ratio of convoluted likelihoods of y_{jt} and of y_t.

Komunjer and Ng (2011) derived necessary and sufficient conditions that guarantees local identification of the parameters of a log-linearized solution of a DSGE model. Their approach requires computing of 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 and the derivatives of their linear transformations, T and U, that deliver the same spectral density. Under regularity conditions, they show that two systems are observationally equivalent if there exist two 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 [1]

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

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

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

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

[1] We use slightly different definitions than Komunjer and Ng (2011). They define a system 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 $n_e = n_y$.

$\Delta_{j,\Lambda}(\theta_0)$ defines the local mapping between θ and the tuple $\Lambda(\theta) = [A(\theta), B(\theta), C_j(\theta), D_j(\theta), \Sigma(\theta)]$. 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 $(H_j(L;\theta), \Sigma(\theta))$, where $H_j(L;\theta) = \left(D_j(\theta) + (I - A(\theta)L)^{-1} C_j(\theta)L^{-1}\right)$. A necessary and sufficient condition for local identification at θ_0 is that

$$\text{rank}(\Delta_j(\theta_0)) = n_\theta + n_x^2 + n_e^2 \tag{4}$$

Thus, given a θ_0, we compute the rank of $\Delta_j(\theta_0)$ for each y_{jt} and compare it with $n_\theta + n_x^2 + n_e^2$, the theoretical rank needed to achieve identification of all the parameters. The set of variables which minimizes the discrepancy $(n_\theta + n_x^2 + n_e^2 - \text{rank}(\Delta_j(\theta_0)))$ is the one selected for full information estimation of the parameters.

The rank comparison we perform will hopefully single out combinations of endogenous variables with " good" and "bad" identification content. However, the analysis will not be able to rank combination of observables keeping in mind weak and partial identification problems, which often plague likelihood based estimation of DSGE models (see e.g. Ahn and Schorfheide, 2007 or Canova and Sala, 2009). For this reason we complement the analysis by comparing measures of the elasticity of the convoluted likelihood function with respect to the parameters in the singular and in the non-singular systems - see next paragraph on how to construct the convoluted likelihood. We seek for the combination of variables which makes the curvature of the convoluted likelihood around a pivot point in the singular and non-singular systems "close". We considered two different distance criteria: in the first, absolute elasticity deviations are summed over the parameters of interest. In the second, a weighted sum of the square deviations is considered, where the weights are function of the sharpness of the likelihood of the singular system at θ_0.

The other statistic we employ to select the variables to be used in estimation measures the relative informational content of the original singular system and of a number of non-singular counterparts. To measure the informational content of different non-singular systems, we follow Bierens (2007) and convolute y_{jt} and of y_t with a $n_y \times 1$

random i.i.d. vector. Thus the vector of observables is now

$$Z_t = y_t + u_t \tag{5}$$

$$W_{jt} = Sy_{jt} + u_t \tag{6}$$

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 dimension n_y. For each non-singular structure j, we construct

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

where $\mathcal{L}(m|\theta, y_{1t})$ is the likelihood of m=Z_t, W_{jt}, given the parameters θ, the history of the structural shock up to t-1, e^{t-1}, and the convolution error, u_t. (7) can

be easily computed, if we assume that e_t are normally distributed, since the first and second conditional moments of Z_t and W_{jt} are $\mu_{w,t-1} \equiv E_{t-1}W_t = SC_j(\theta)(I - A(\theta)\ L)^{-1}B(\theta)e_{t-1}$, $\Sigma_{w_j} \equiv V_{t-1}W_t = SD_j(\theta)\Sigma(\theta)D_j(\theta)'S' + \Sigma_u$, $\mu_{z,t-1} \equiv E_{t-1}Z_t = C(\theta)(I - A(\theta)\ L)^{-1}B(\theta)e_{t-1}$ and $\Sigma_z = V_{t-1}Z_t = D(\theta)\Sigma(\theta)D(\theta)' + \Sigma_u$.

Bierens imposes under 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 our framework, we integrate out of (7), both the e^{t-1} and the u_t, and choose the combination of observables j that minimize the average information ratio $p_t^j(\theta)$, i.e.

$$\inf_j p_t^j(\theta) = \int_{e^{t-1}} \int_{u_t} p_t^j(\theta, e_t^{t-1}, u_t) de^{t-1} du_t \tag{8}$$

Given θ, $p_t^j(\theta)$ identifies the combination of variables that produces 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 the convolution errors. Thus, among all the possible combinations of endogenous variables, we select the vector in which the loss of information caused by variable reduction is minimal.

3 An application

In this section we apply our procedures to the singular version of the benchmark DSGE model of Smets and Wouters (2007). This model is selected for the exercise we run because its widespread use in academics, central banks and policy institutions, and because it is frequently adopted to study cyclical dynamics of developed economies and their sources of variations.

We retain all the nominal and real frictions originally present in the model, but we make a number of simplifications to the structure which have no consequences on the conclusions we reach by reduce the computational burden of the experiment. First, we assume that the exogenous shocks are stationary. Since we are working with simulated data, such a simplification involves no loss of generality. The sensitivity of our conclusions to the inclusion of trends in the data is discussed in section 4. 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 (rather than the flexible price equilibrium).

The model typically features a large number of shocks and this makes the number of observable variables and the number of exogenous disturbances equal. Several researchers (for example, Chari, Kehoe and McGrattan (2009) or Sala, Soderstrom, Trigari (2010)) have noticed than some of them have dubious economic interpretations - rather than being structural they are likely to capture potentially misspecified aspects of the model. Here, relative to the SW model, we turn off the price markup, the wage markup and the preference shock which are the disturbances more likely to capture these misspecifications and we consider a model driven by technology, investment specific, government and monetary policy shocks, i.e. $(a_t, i_t, g_t, \epsilon_t^m)$. The basic vector of observable variables coincides with the SW choice of measurable quantities, that is, we have output, consumption, investment, wages, inflation, interest rate and hours worked $(y_t, c_t, i_t, w_t, \pi, r_t, h_t)$.

The log-linearized equations of the model are summarized in table 1 (To be added). To implement the procedures we need to select the θ vector. Table 1 presents our choices: basically, these are the posterior mean of the estimates reported by SW. It is important to stress that any selection 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, and others have economic interpretations, e.g. price indexation or the inverse of Frish elasticity, we focus on a subset of the latter ones when computing elasticity measures.

To construct the convoluted likelihood we need to choose Σ_u, the variance of the convoluted error. We set $\Sigma_u = \kappa * I$, where the scaling factor κ is the maximum of the diagonal elements of $\Sigma(\theta)$, thus making sure that u_t and the innovations in the model e_t have similar scale. In the sensitivity analysis section we describe what happen when a different κ is used. When constructing the ratio $p_t^j(\theta)$ we simulate 500 samples, where the history e_t^{t-1} and the convolution error u_t are random, and average the resulting $p_t^j(\theta, e_t^{t-1}, u_t)$ over the simulated samples.

We also need to select a sample size for the likelihood computation. We set $T = 150$, so as to have a data set comparable to those available in empirical work. In the sensitivity analysis section we discuss what happens to our ranking if the sample size is large, $T = 1500$.

We also need to set the size of the step needed to 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. In the baseline exercises we set g=0.01. When computing the rank of the spectral density, we also need to select the "tolerance level" (see Komunjer and Ng (2011)). This tolerance level is important in defining identifiability of the parameters. As suggested by the authors, we set it equal to the step of the numerical derivatives, q=g= 0.01. We examine the sensitivity of the results with respect to the choice of g and q in section 4.

3.1 The results of the rank analysis

The model features 29 parameters, 12 predetermined states and four structural shocks. Thus the Komunjer and Ng's condition for identification of all structural parameters is that the rank of $\Delta(\theta_0)$ is 189.

We start considering the unrestricted specification and all possible combinations of observables, and ask whether there exists four dimensional vectors of observables 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 is 35, i.e. $\begin{pmatrix} 7 \\ 4 \end{pmatrix} = \frac{7!}{4!(7-4)!} = 35$.

The first columns of table 2 presents a subset of the 35 possible combinations of observables and the second column the rank of the matrix of derivatives, $\Delta_j(\theta_0)$. Combinations are presented according to the rank of the matrix of derivatives, from large to small. To facilitate the examination, we divide the table into two parts:

combinations with large rank, i.e. $\text{rank}(\Delta_j) \geq 185$, and combinations with low rank: $\text{rank}(\Delta_j) < 185$. Clearly, no combination guarantees full parameter identification. This is a well known result (see e.g. Iskrev, 2009, or Komunjer and Ng, 2011) and our analysis confirm this fact. Interestingly, the combination containing the real variables, (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 information for the structural parameter. Conversely, real wages appears more often in low rank combinations than in large rank ones suggesting that this variable has relatively low identification power. For consumption, output or hours worked, the conclusions are much less clear cut. Turning to nominal variables, notice that among the large rank combinations interest rate appears more often than inflation (7 vs. 4), suggesting a mild preference for interest rates over inflation, as far as parameter identification is concerned. More striking is the result that including both inflation and interest rate in the vector of observables makes parameter identification poor; indeed, all combinations featuring these two variables are in the low rank region and four have the lowest rank, i.e. 183.

The third column of table 2 repeats the exercise calibrating some of the 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 other are implicitly selected by statistical agencies (e.g. the depreciation rate of capital). For this reason we have repeated the rank calculation exercise fixing the depreciation rate, $\delta = 0.025$, the good markets and labor marker Kimball aggregators, $\varepsilon_p = \varepsilon_w = 10$, elasticity of substitution labor, $\lambda_w = 1.5$ and government consumption output share $c/g = 0.18$, 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. $\text{rank}(\Delta_\Lambda(\theta_0)) = n_\theta = 29$, they are not sufficient to guarantee local identification of the structural parameters. In general, while the ordering obtained in the unrestricted case is generally preserved, differences combinations of variables have now more similar spectral ranks.

Finally, we examine whether there are parameters restrictions that allow some non-singular 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 of table 2, the parameter restriction that generates identification for each combination of observables we consider. A black space means that there are no parameter restrictions able to generate full parameter identification for that combination. 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 2 reports the parameters restrictions that achieve identification for each combination of observables.

From column 4 one can observe that, for some combinations, the extra restriction is not enough to achieve full parameter identification. In addition, the combinations of variables which were best in the unrestricted case are still the combinations 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, the parameter that has to be fixed to achieve identification is elasticity of capital utilization adjustment costs. Column 5 indicates that at least four restrictions are need to identify the vector of structural parameters of the SW model and that the goods and labor market aggregator, ε_p and ε_w, cannot be estimated either individually or jointly for any combination of observables. In general, the largest (unrestricted) rank combinations are more likely to produce identification with a tailored use of parameter restrictions.

3.2 The results of the elasticity analysis

As mentioned, the rank analysis is unsuited to detect weak and partial identification problem that often plague estimation of the structural parameters of the DSGE model. To address investigate issue 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 3 presents the four best combinations minimizing the "elasticity" distance described in the previous section. We focus attention on six parameters, which are often the object of discussion among macroeconomists: the habit persistence, the inverse of the Frish elasticity of labor supply, the price stickiness and the price indexation parameters, the inflation and output coefficients in the Taylor rule. As it is clear comparing table 2 and table 3, maximizing the rank of the spectral density does not necessarily make the curvature of the convoluted likelihood in the singular and non-singular system close. The vector of variables which is best according to the "elasticity" criteria is consumption, investment, hours and the nominal interest rate, but combinations including only real variables have objective functions which are close. As shown in figure 1, 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 Frish elasticity and of the two Taylor coefficients worse. Note that, even the best combination of variables makes the curvature of likelihood quite flat as far as the inflation coefficient in the Taylor rule is concerned. Thus, while there does not exist a combination which simultaneously avoid weak identification problems in all six parameters, different combinations of variables may reduce weak identification problems in different 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 within the class of "good" curvature combinations.

It is worth also mentioning that while there are no theoretical reasons to prefer the use of any two variables among output, hours and labor productivity, and the ordering the best models is unaffected, 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 the simultaneous use of output and labor productivity, in place of output and hours in the vector of observables can help to reduce the flatness of the likelihood function in the dimensions represented by the inflation and the output coefficients in the Taylor rule, 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 4 observables according to the information statistic (8). As in table 3, we also provide the value of the average objective function for that combination relative to the best. Since the average log of the likelihood ratio statistic is negative for all parameter combinations, the maximum value is the smallest in absolute value, and the ratio is smaller or equal to 1.

The table suggests that an econometrician interested in estimating the structural parameters of this model by maximum likelihood 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, while among the top four, 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 $p_t(\theta)$ measure can sharply distinguish various options.

Interesting, the identification and the informational analysis 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 among the top combinations according to the rank analysis.

To estimate the structural parameters of this model it is therefore 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 used. Nevertheless, it is a fact that, 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 of the Phillips curve are difficult to identify no matter what combination is used, the use of real variables allows us to pin down fairly well the intertemporal and intratemporal links of the model which crucially determine income and substitution effects.

4 Robustness

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

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 shocks of the model and repeat the analysis maintaining a maximum of seven observable variables. Tables (5, (6),(7) 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 $p_t(\theta)$ statistics instead is still quite sharp in distinguishing the best vector of variables from the others.

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 matter for the conclusions. For this reason we have repeated the exercise using T=1500. The results are reported in the second panel of table (3) and in the third panel of table (5).

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.

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 third panel of table (3) and in the four panel of table (5) 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 we consider.

Changing the step in the numerical derivatives In computing numerical derivatives in both the rank and elasticity analysis we had to select the step g, which defines the radious 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 much smaller radius. Choosing a smaller g has minor effects on the elasticity analysis (see fourth panel of table (3)) 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 in a very small neighborhood of the chosen parameter vector, the parameter vector is identifiable (once restrictions are imposed). However, since this is not the case as we make the neighbourhood slightly larger, weak identification seems a feature of this model.

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

$$Q_j(\theta, e^{t-1}, u_t) = \sum_{t=1}^{T} (Z_t - W_{jt})' \Sigma_q^{-1} (Z_t - W_{jt}) \tag{9}$$

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 in the exercises. We are looking for the combination of variables minimizing $Q_j(\theta, e^{t-1}, u_t)$, integrating out both the history of the shocks and the convolution error.

When four observables are used (see table (5), 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 with the $Q_j(\theta)$ statistics but differences in the relative informativeness of the two vectors is small according to both statistics. When six observables are used (see table (7), 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.

Other exercises We have also examined what happens when we change the tolerance level in the rank analysis and found only minor differences with the baseline scenario. We have also considered the case when the neutral 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 parameters 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.

5 Conclusions and suggestions for empirical practice

This paper proposes methods to select the observables to be used in the estimation of the structural parameters when one feel 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 equations and insists in working with a singular DSGE model. The methods we suggest measure the identification and the information content of various vectors of observables, are easy to implement, and seem able to distinguish "good" from "bad" combinations of observable variables. Interestingly, and despite the fact that the statistics we consider are derived from different principles, the best combinations of variables these methods deliver are pretty much the same.

Although monetary policy plays an important role in the DSGE model we consider, parameter identification and variable informativeness are optimized using only real variables in estimation and including output, consumption and investment seems the best. These variables help to identify the intertemporal and the intratemporal links in the model and thus are useful to get the right income and substitution effects. Interestingly, using interest rate and inflation jointly in the estimation make identification worse and the loss of information larger. When one takes the curvature of the likelihood in the dimensions of interest into consideration, we find it preferable to include the nominal interest rate rather than the inflation rate in the list of observables.

We show that the ordering of various combinations is broadly maintained when parameter restrictions are added, and this is true both when these restrictions are chosen following the conventional wisdom or when they are selected to optimize some efficiency criteria. We show that, at least in terms of likelihood curvature there are important trade-off when deciding to use hours or labor productivity among the observables. Finally, we demonstrate that our conclusions are broadly invariant to changes in the setup of the experiment.

While our conclusions are sharp, an econometrician working in a real world application should also naturally consider whether the measurement of a certain variables is reliable or not. Our study only asks what set of observables is preferable on theoretical grounds. In practice, the analysis we have performed can be undertaken even when some measurement errors are preliminary added to the model.

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 of the model, one wants to examine which features of the model is well identified and what is the information content of different variables. Seen through these lenses, the analysis we perform here complements those of Canova and Paustian (2011) and of Mueller (2010).

References

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

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.

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.

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.

Mueller, U., 2010. Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models, Princeton University, manuscript.

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.

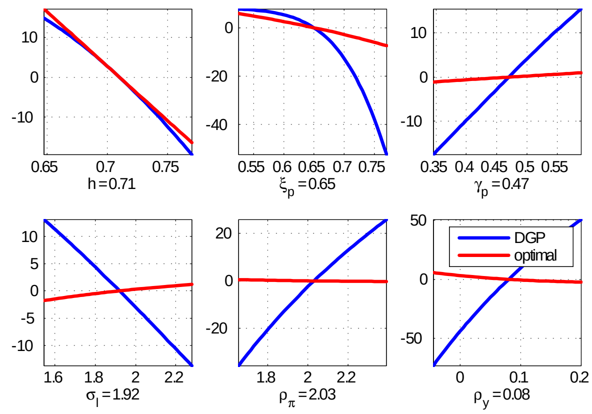

Figure 1: One dimensional convoluted likelihood of the DGP and of the optimal combination.

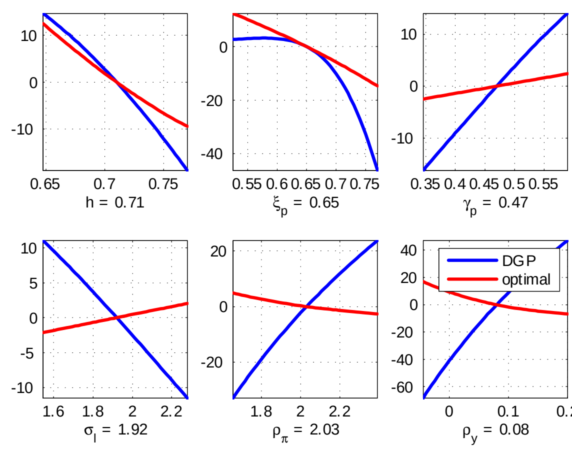

Figure 2: One dimensional convoluted likelihood of the DGP and of the optimal combination using labor productivity.

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 1: Parameters description and values used.

	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
y, i, π, r	184	188	ψ	
i, w, π, h	184	188	ψ	
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		

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

Order	Cumulative Deviation	Weighted Square	Ratio
Basic			
1	(c, i, r, h)	(c, i, r, h)	1.00
2	$(y.c, i, w)$	(c, i, w, h)	1.65
3	(c, i, w, h)	(y, c, i, w)	1.91
4	(y, c, r, h)	(y, c, r, h)	2.12
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 3: Ranking of 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 the baseline results, the second increasing the sample size, the third, changing the step size in computing derivatives, the fourth the magnitude of the convolution error.

Order	Basic		Quadratic Distance		T=1500		$\Sigma_u = 0.01 * I$	
	Combination	Relative Information	Combination	Relative Information	Combination	Relative Information	Combination	Relative Information
1	(y, c, i, h)	1	(y, c, i, w)	1	(y, c, i, h)	1	(y, c, i, h)	1
2	(y, c, i, w)	0.89	(y, c, i, h)	0.89	(y, c, i, w)	0.87	(y, c, i, w)	0.86
3	(y, c, i, r)	0.52	(y, c, i, r)	0.6	(y, c, i, r)	0.51	(y, c, i, r)	0.51
4	(y, c, i, π)	0.5	(y, c, i, π)	0.59	(y, c, i, π)	0.5	(y, c, i, π)	0.5

Table 4: Ranking based on the $p(\theta)$ statistic. The first two column present the results for the basic setup, the next six columns the results obtained altering some nuisance parameters. Relative information is the ratio of the $p(\theta)$ statistic relative to the best combination.

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 5: Rank conditions 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 $c/g = 0.18$. SW model with 6 shocks.

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 6: 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, the third, changing the step size in computing derivatives, the fourth changng the magnitude of the convolution error.

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 7: Ranking according ot the $p(\theta)$ statistic, 6 observables.The first two column present the results for the basic setup, the next two columns the results obtained altering some nuisance parameters. Relative information is the ratio of the $p(\theta)$ statistic relative to the best combination.