A Note on Identification Patterns in DSGE Models

Michal Andrle*

First draft: September, 2009
This draft: January 2010

Abstract

This paper comments on selected aspects of identification issues of DSGE models. It suggests the singular value decomposition as a useful tool for detecting local non- and weak identification. The decomposition is useful for checking rank conditions of identification, identification strength and, importantly, it offers 'identification patterns' of the parameter space. With respect to other methods of identifiability structure of the parameter space of a given model the singular value decomposition is particularly easy to apply and offers an intuitive interpretation. We suggest a simple algorithm for detecting non- and weak identification and an algorithm for finding a set of most identifiable set of parameters. We also demonstrate that use of bivariate and multiple correlation coefficients of parameters is only limited check for identification problems.

Keywords: DSGE, identification, information matrix, rank, singular value decomposition,
J.E.L. Classification: F31, F41

*Czech National Bank, Monetary and Statistics Dept., Macroeconomic Forecasting Division. The main part of the paper was written while I was working for European Central Bank, DG-Research, Econometric Modelling Division in 2009. Email: name_surname(at)yahoo.com. I would like to thank Jan Brůha, Antti Ripatti and Anders Warne for useful comments and discussions. I am indebted to Juha Kilponen for his enormous encouragement, invaluable comments and suggestions. All opinions are of the author and should not be associated with the CNB or the ECB.

Introduction

This paper is about a method and pictures that go with it.[1] The method suggested deals with inspection of 'identification patterns' in Dynamics Stochastic General Equilibrium (DSGE) models build by economists and used for research and policy analysis, but is not limited only to these models.

DSGE models are nowadays typically estimated relying on formal econometric full- or limited information methods instead of less formal calibration of parameters. In order to obtain meaningful results of estimation, parameters must be 'identified', in the since that different set of structural parameters should not result into observationally equivalent outcomes. Importantly, identification is relevant in both classical (frequentist) and Bayesian approach to statistical inference.

We show that identification of DSGE models effectively boils down into a problem of invertible linear transformations, i.e. whether one can recover structural parameters θ from a set of reduced-form parameters $\tau(\theta)$ that are functionally dependent on those structural ones, or of a parameter sensitivity of selected criterion function. In the context of the problem order and rank conditions naturally arise.

Our goal is to find out non-identified and weakly identified parametrizations, their structures and suggest the most plausible restrictions on unidentified parameters. To understand what parameters or combinations of structural parameters are the cause of the problem, we propose to inspect four basic subspaces of linear maps which allow to locate identifiable and non-identifiable subspaces in the parameter space. We suggest that the *singular value decomposition* is natural candidate for locating the unidentified parameter subspace and delivers also insight into the 'strength' of the identification. Further, we demonstrate that bivariate and multiple correlation measures may provide rather missleading view about identification and demonstrate a method for detecting best identified parameters based on rank-revealing factorizations.

The issue of identification is of course well-known in the econometrics literature, e.g. Fisher (1966), Rothemberg (1971), Hannan (1971) or Hsiao (1983) to name but few classics. A research on 'weak' identification has been recently stimulated by Staiger and Stock (1997), or Stock, Wright and Yogo (2002).

The identification problem related to DSGE models has been perhaps underresearched for a while and seem still neglected by some researches in the area. Importantce of the issue was reminded by Canova and Sala (2006) and Canova and Sala (2009), lucid paper by Cochrane (2007) or investigation by Iskrev (2008) and Iskrev (2009b) with focus on estimation. Cochrane (2007) and Beyer and Farmer (2007) rise the important issue of observationally equivalent structures and types

[1] The opening sentence is inspired by Strang (1993)

of equilibria in connection –with often arbitrary– lag length restrictions, a point raised previously by Pesaran (1987) and Sargent (1978). These papers alert to the importance of inspection of border-line lag specifications, since absence or presence of sufficient lag length may prevent or deliver identifiability for important structural parameters.

The very recent contribution of Komunjer and Ng (2009) states necessary and sufficient order and rank conditions for local identifiability of DSGE models, so we do not restate the results. The commonality of this paper and Komunjer and Ng (2009) is the departure from well-established literature on identification of linear and non-linear dynamic state-space models in an engineering literature.[2] The above mentioned paper adjusts some of the results by an engineering classics Glover and Willems (1974) or Grewal and Glover (1976), *inter alia,* to specific conditions of DSGE models – e.g. possible stochastic singularity and lack of observed input. These are important contributions, providing both necessary and sufficient conditions for local-identifiability of (minimal) state-space realizations from auto-covariance generating function (ACGF). The method suggested in this paper complements the optimal selection of restriction in unidentified models.

After some progress in the analysis of the identification via the nullspace of the linear map and the singular value decomposition and its applications, the effort to treat the issue formally we have found a related work on multicollinearity and identification. In the field of regression analysis Belsley, Kuh and Welsch (1980) discuss multicollinearity detection and mention the use of SVD. Vajda, Rabitz, Walter and Lecourtier (1989) use eigenvalue decompositions of the Information matrix in chemical engineering models motivated by principal components, and recently Van Doren, den Hof, Jansen and Bosgra (2008) use the singular value decomposition for the analysis of the Information matrix as it is also suggested bellow. The use of the SVD for detecting multicollinearity and near collinearity is thus not novel, though we treat the issue in greater detail, with stronger relation to subspaces and also suggest an algorithm for a parameter subset selection.

The plan of the paper is as follows. The first sections defines the issue of identification and its importance. The second section focuses on the analysis of the presence of unidentifiability and its sources by inspecting the rank and the nullspace of the linear map. Section three investigates the strength of identifiability in relationship with collinearity detection, provides algorithm for parameter subset selection and analyzes the limitations of correlation measures. Section four demonstrates the method using two well-established DSGE models and then we conclude.

[2]In engineering literature by 'identification' of the model it is often understood estimation of the model, yet structural identification is also dealt with. Another important difference is often the availability of both output and input of the system in engineering, which is not the case in economics.

1 Identification and Observational Equivalence

1.1 Identification Defined

Identification is related to observational equivalence and with ability to draw inference on parameters from the data.

Let Y be the set of observations and let structure S be a complete probability specification of Y in the form $S = F(Y, \theta)$ where $\theta \in \Theta \subset \mathbb{R}^n$ is the vector of parameters, Θ being the parameter space. Two structures, $S^0 = F(Y, \theta^0)$ and $S^* = F(Y, \theta^*)$ are said to be observationally equivalent, if $F(Y, \theta^0) = F(Y, \theta^*)$ for almost all Y. The structure is identified if this equality means $\theta^0 = \theta^*$, and unidentified otherwise.

Often the inspection of global identification for the whole parameter space is difficult, so we say that the structure is locally identified if there exists an open neighbourhood of θ^0 containing no other $\theta \in \Theta$ which produces observationally equivalent structure. This paper treats only local identification of DSGE models, that is identification nearby the specific value of θ in the parameter space Θ.

In a lucid paper Rothemberg (1971) proved, subject to some regularity conditions, that θ^0 is *locally identified* for a given structure if and only if the Information matrix evaluated at θ^0 is not singular, i.e. of defficient rank. Further, note that Rothemberg (1971) stated the results for two important cases – (i) the case of Information matrix without existence of the *reduced form* structure and (ii) in case of existence of the reduced form, where reduced form parameters $\tau \in \mathcal{T} \subset \mathbb{R}^m$ may help to establish the identification of the structural parameters. We thus assume existence of reduced form parameters and the mapping $T(\theta, \tau) = 0$.

In case where $T(\theta, \tau) = 0$ exists and, importantly, if reduced form parameters are identified, the necessary and sufficient condition for identification is that $T \equiv \partial T / \partial \theta'$ is of full rank. The question is thus whether we can find a unique solution from τ to θ.

Bayesian View of the Identification Problem First we briefly comment on a possible Bayesian view on the identification issues. The problem of identification does not seem as clearcut under the Bayesian paradigm as under a Classical one, see Aldrich (2002) for the enlightening review and discussion of how likelihood and identification went Bayesian. In our view the issue must be divided into 'genuine Bayesians' and 'Bayesians out-of-convenience', who consider prior information only as a method of regularization of the optimization problem of the likelihood.

Aldrich (2002) discusses the difficult evolution of 'identification' under the Bayesian paradigm where parameters may be estimable even when data are completely

uninformative abou these. There were requests for broadening the concept of identification for Bayesians, view that "underidentifiability causes no real difficulty in the Bayesian approach" (Dréze 1972) followed by recognition that "... it was misleading to use the word 'identification' in defining a property of the prior density for the parameters of unidentified models. [I] agree with Kadane's view that 'identification is a property of the likelihood function and it is the same whether considered classically or from the Bayesian approach'" in (Dréze 1975).

The uninformativeness of the data for parameters can be treated as marginal uninformativeness or conditional uninformativeness, see Poirier (1998). It is a common view that equality of marginal posterior with a prior distribution is a sign of lack of identification and uninformativeness of the data about a parameter. The likelihood is non- or weakly responsive for changes in the parameter and the prior information dominates no matter what the sample size is, beliefs may not be revised about the parameter. On the other hand, when there is an explicit or an implicit dependence among parameters the data may be marginally informative even for conditionally unidentified parameter, hence the difference of marginal posterior from prior density is not sufficient sign of identification.

In case of 'Bayesians out-of-convenience' the importance of identification as a property of the likelihood or data informativeness in general are important, since as it is argued in Gelman, Carlin, Stern and Rubin (2004, Chapter 4) or recently by Guerron-Quintana, Inoeu and Kilian (2009) or Moon and Schorfheide (2009) the large sample inference and frequency properties of Bayesian inference are importantly affected by identification problems. The key fact is that the likelihood does not dominate the prior information as sample size grows.

We limit ourselves to identification in terms of *whether the data may become informative for parameters* and thus we explore the properties of likelihood or other criterion functions.

1.2 DSGE Models Case

We write down a linear or linearized DSGE model in a standard state-space form as

$$X_t = \mathbb{C}_1(\theta) + \mathbb{T}(\theta)X_{t-1} + \mathbb{R}(\theta)\varepsilon_t \quad (1)$$

$$Y_t = \mathbb{C}_2(\theta) + \mathbb{Z}(\theta)X_t + \mathbb{H}(\theta)\varepsilon_t, \quad (2)$$

where the state-space parameters are functionally related to set of structural parameters θ as indicated by the notation and $\mathbb{E}[\varepsilon\varepsilon'] = \mathbb{S}(\theta)$. We declare the set of *reduced form parameters* as

$$\tau \equiv \{\text{vec } \mathbb{C}_1(\theta); \text{vec } \mathbb{C}_2(\theta); \text{vec } \mathbb{T}(\theta); \text{vec } \mathbb{R}(\theta); \text{vec } \mathbb{Z}(\theta); \text{vec } \mathbb{H}(\theta); \text{vec } \mathbb{S}(\theta)\}.$$

The properties of the mapping $T(\theta, \tau) = 0$ are crucial for identification of DSGE models. Due to its highly non-linear nature we inspect the Jacobian of the map evaluated at particular θ so we explore linear map $\tau = T\theta$. Note that the non-uniqueness of the solution of this map is sufficient to bring unidentification. Its uniqueness is only necessary condition for identificiation since τ may not be well identified.[3]

We take the model under inspection as given. We do not follow the implication of unidentificiation or weak identification for possible modification of a model economic structure, etc. At this moment we also consider the *set of observables as given,* although it is clear that the identification analysis is always conditional on the set of observables of the model. Note that this is not the same as distinguishing limited information and full-information methods of estimation (Canova and Sala 2009), it concerns the model's proper transfer function and thus also full-information methods.

Estimation Methods The method discussed bellow is not limited to a particular estimation method. It focuses on two basic ingredients – the Hessian of the criterion function and, in if plausible, the mapping from structural parameters to reduced form parameters.

In particular we shall focus on the Information matrix based on the log-likelihood function of the state-space model, we do not restrict ourselves to a particular way of obtaining the likelihood function of the model, see e.g. Harvey (1989). Likelihood functions convey full information and it is the key building block of Bayesian estimation of DSGE models that became popular. We carry out all exercises without use of the data and we do not estimate the model. We evaluate identification in certain areas of parameter space. However, the number of time periods T is an important as it enters the likelihood function as a parameter.

We shall thus explore properties the information matrix

$$\begin{aligned} R(\theta) &\equiv \mathbb{E}\left\{\left(\frac{\partial L}{\partial \theta'}\right)'\left(\frac{\partial L}{\partial \theta'}\right)\right\} = \mathbb{E}\left\{\left(\frac{\partial \tau}{\partial \theta'}\right)'\left[\left(\frac{\partial L}{\partial \tau'}\right)'\left(\frac{\partial L}{\partial \tau'}\right)\right]\left(\frac{\partial \tau}{\partial \theta'}\right)\right\} \quad (3) \\ &= \mathbb{E}\left\{\left(\frac{\partial \tau}{\partial \theta'}\right)' R(\tau)\left(\frac{\partial \tau}{\partial \theta'}\right)\right\}, \quad (4) \end{aligned}$$

and the mapping $T\theta = \tau$, where $\theta \in \Theta \subset \mathbb{R}^n$ is the vector of structural parameters and $\tau \in \mathcal{T} \subset \mathbb{R}^m$ is the vector of reduced form parameters – if plausible – and $R(\tau)$ is defined as the information matrix with respect to reduced form parameters.

[3]Recall that state-space models can be related by similarity transforms, which are unique in case of minimal systems, so the state X_t is not identified up to rotation, see e.g. Kailath (1980) or Harvey (1989), and for the role in identification Glover and Willems (1974) or Komunjer and Ng (2009). Further, we do not make any further assumptions in terms of observability and controllability since we care about the presence of structural identifiability, though these are related, see e.g. DiStefano (1977).

2 Identification and Linear Maps

The identification issue generally boils down to inspecting a rank of a matrix, which often is an Fisher Information-matrix, Hessian of a particular estimation criterion function with respect to θ or a linear map from structural parameters θ to reduced form parameters τ such as $\tau = T\theta$. Ideally we need to inspect the rank of the matrix and –if possible– determine the causes of the rank-deficiency of the matrix. In case of reduced form parameters the question may also be posed about the invertibility of the linear map (its matrix) T.

Let us first focus on the intuitive case of the mapping $T\theta = \tau$, where T is the Jacobian. Let $\theta \in \Theta \subset \mathbb{R}^n$ and let $\tau \in \mathcal{T} \subset \mathbb{R}^m$, which implies that the matrix of the map T is $(m \times n)$. To investigate the mapping one can use standards results from linear algebra. In our view Simmons (2003), Strang (1988), Axler (1997) and Golub and van Loan (1996) are useful general references, Strang (1993) is a joy to read.

2.1 Important Subspaces and SVD

The relationship $T\theta = \tau$ is fully described by linear transformation, its matrix T and its *four fundamental subspaces* – null space of T, null(T), range of T and null space and range of T', whose dimensions obey many important rules, most importantly,

$$\dim \text{range}(T) + \dim \text{null}(T) = n. \tag{5}$$

We define rank of T as $\text{rank}(T) = \dim \text{range}(T)$. A not self-evident fact is that the rank of the column space is equal to the rank of the row space of the matrix, hence the concept of rank is unambiguous, i.e. $\dim \text{range}(T) = \dim \text{range}(T')$.

Recall that the null space of linear transform is a subspace of its domain and consists of all vectors which are the solution to $T\theta = 0$. Importantly, when the range is of dimension r the null space is of dimension $n - r$.

It is clear that only when τ is in the range(T), i.e. in the column space of T, we can solve the problem $T\theta = \tau$. If $m < n$ the problem for obtaining unique structural parameters from reduced ones is ill posed. This is due to the fact that *order condition* does not hold and it is impossible for the columns of T to be independent. For at least one solution we require $m \leq n$, but in order to hope for at most one solution we require $m \geq n$.

These facts are quite well known, so for *exact identification* we search for a *unique* solution and thus we require the linear map determined by matrix T to be of a full rank. Importantly, we demonstrate that the inspection of subspaces associated with T bring insight into the problem, due to their orthogonal properties. The null space

of T is orthogonal to range of T' and also range of A is orthogonal to left-null space, null (T'), which has dimension of $m - r$.

An absence of null (T') implies that a solution can always be found and an absence of null (T) indicates that the solution is *unique*. The absence of nullspace of T is easy to check via rank conditions guaranteeing unique solution, when T is not singular.

When there is not a unique solution and thus the null space of T is not empty, it is crucial to explore the structure of the null space, which determines subspace of unidentifiable parameters. We may inspect the *row Echelon form* to find out free elements. Consequently, we need to find some 'suitable' basis of the subspace. The linear map is expressed by a matrix T, yet linear maps are defined with respect two basis – domain and target. When the basis are not stated explicitly standard basis can be assumed. Operators (maps from a space to itself) require only one basis. Most results on matrix factorisations, where the mapping is expressed to some 'nice' basis, concern square non-singular matrices. It is useful to have more general tool to analyze the matrix of a linear mapping. Such a useful tool is the *Singular value decomposition* (SVD).

Singular value decomposition If A is a real $(m \times n)$ matrix, then there exist orthogonal matrices

$$U = [u_1, \ldots, u_m] \in \mathbb{R}^{m \times m} \qquad \text{and} \qquad V = [v_1, \ldots, v_n] \in \mathbb{R}^{n \times n}$$

such that

$$U'AV = \Sigma = \text{diag}(\sigma_1, \ldots, \sigma^p) \in \mathbb{R}^{m \times n}, \qquad p = \min\{m, n\}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, (Golub and van Loan 1996).

Using an SVD, we can find a very 'nice' bases for the subspaces needed. We have obtained orthonormal basis vectors and the matrix of the linear map is diagonal with respect to both bases. Importantly we can find the SVD factorisation for any $(m \times n)$ matrix, not necessarily square, or symmetric or well-behaved any other way. Moreover, no complex numbers are involved after applying SVD to real matrix. The cost is, however, that we have two orthogonal bases, which do not cancel mutually. For some problems this might be an issue, but in our case this is not. The SVD is intimately connected to the *eigenvalue decomposition* (EVD), which for square and symmetric matrices –such as the Fisher information matrix– delivers the same decomposition, with a diagonal matrix and a single basis.

The SVD has numerous very useful and interesting properties and consequences. We need only few for our purpose. As it is common, σ_i are labeled as *singular values* and vectors u_i,v_i are the i-th left and right singular vectors, respectively. Further

we have $A = U\Sigma V'$ and thus $Av_i = \sigma_i u_i$ and $A'u_i = \sigma_i v_i$ for i respecting dimensions of U, V.[4]

The SVD is a standard way of determining a *rank* of the matrix, one of our goals. The rank of a matrix A is r such that

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = 0, \tag{6}$$

that is, the rank of a matrix r is equal to number of nonzero singular values. This is why the diagonal form under new basis is so useful – the inputs from the subspaces associated with the zero singular values are annihilated. The matrix of rank r can thus be rewritten factorized as

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} = U_1 \Sigma_1 V_1' = \sum_{i=1}^{r} \sigma_i u_i v_i', \tag{7}$$

where U_1 is $(m \times r)$, Σ_1 is $(r \times r)$ and V_1' is $(r \times n)$. The individual matrices accumulated in the sum are of rank $r = 1$. Importantly, with respect to near-collinearity and compression, the approximation for a particular order of the sum of rank-one matrices can be understood as optimal.

Once the rank is determined one knows whether the identification problem becomes immediate. If the matrix is rank-deficient, i.e. $r < \min\{m, n\}$, it means that null space of A is not empty. If it is not empty, it structure will indicate the suspects. We have that

$$\begin{aligned} \text{null}(A) &= \text{span}\{v_{r+1}, \ldots, v_n\} & (8) \\ \text{range}(A) &= \text{span}\{u_1, \ldots, u_r\}, & (9) \end{aligned}$$

so the 'criminals' form a basis for the null space of A. Those unidentified parameter combinations can be found in V_2' and as a bonus – they are sorted according to degree of their 'crime' and are 'dressed' as normalized unit vectors.

In what follows we label the *structure of the null space* of the linear map under consideration as an *identification pattern.* Next sections demonstrate in more detail how identification patterns can be used and interpreted.

2.2 Identification Patterns, Nullspace and Restrictions

As we have demonstrated, inspecting several subspaces associated with a linear map suggest insights into roots of (near) singularity. Assume we have already decomposed the Information matrix $R(\theta)$ or map T and found it of rank r, $r < \min\{m, n\}$ and

[4]Hence $\|Av_i\| = \sigma_i$. Also $u_i = Av_i/\sigma_i$ and it is a unit eigenvector of AA', so σ_i^2 are positive eigenvalues of AA'.

thus rank-deficient. Thus we have matrices $V = [V_1 \ V_2]$ and $U = [U_1 \ U_2]$ that capture the necessary information about the map. In case of the square symmetric matrix we have $U = V$.

The rows of V_2' provide an orthonormal basis of the nullspace of the map and identify directions in the parameter space where the parameters are structurally unidentifiable. The columns of U_1 constitute the orthonormal basis for the range (column space) of the linear map (the Information matrix) and constitutes a mapping from the original parameter space Θ to lower dimensional parameter space $\mathcal{K}$ of dimension $(n - r)$.

Rank-deficiency implies a *necessity of restriction of the parameter space* of a particular order. The inspection of the nullspace and the range of the map suggests what these restrictions should be in order to achieve local identification. Basically one needs to get rid of the members of the nullspace. Consider a restriction from $\theta \in \Theta \subset \mathbb{R}^n$ into $\kappa \in \mathcal{K} \subset \mathbb{R}^r$ given by $\phi(\kappa, \theta) = 0$. Then $\partial\theta/\partial\kappa' = U_1$ or simply $\theta = U_1\kappa$ where U_1' is $(r \times n)$ matrix. Reparameterizing the model in terms of κ and calculating the Information matrix $R(\kappa)$, an $(r \times r)$ matrix, we get

$$
\begin{aligned}
R(\kappa) &= \mathbb{E}\left\{\left(\frac{\partial L}{\partial \kappa'}\right)'\left(\frac{\partial L}{\partial \kappa'}\right)\right\} && (10)\\
&= \mathbb{E}\left\{\left(\frac{\partial \theta}{\partial \kappa'}\right)'\left[\left(\frac{\partial L}{\partial \theta'}\right)'\left(\frac{\partial L}{\partial \theta'}\right)\right]\left(\frac{\partial \theta}{\partial \kappa'}\right)\right\} && (11)\\
&= \mathbb{E}\left\{\left(\frac{\partial \theta}{\partial \kappa'}\right)'\left[U\Sigma V'\right]\left(\frac{\partial \theta}{\partial \kappa'}\right)\right\} = \mathbb{E}\left\{\left(\frac{\partial \theta}{\partial \kappa'}\right)'\left[U_1\Sigma_1 V_1'\right]\left(\frac{\partial \theta}{\partial \kappa'}\right)\right\} && (12)\\
&= \mathbb{E}\{U_1'U_1\Sigma_1U_1U_1'\} = \mathbb{E}\{\Sigma_1\}, && (13)
\end{aligned}
$$

which is obviously of rank r as required. This rank condition is related to rank conditions of stacked $[R(\theta);\ \phi(.)]$ as in Rothemberg (1971).

To highlight the intuition, assume a very special case of $R(\tau) = 1$, perfectly identified reduced form parameters, and thus $R(\theta) = \mathbb{E}\{T'T\}$. Then using SVD to analyze $T = U\Sigma V'$ delivers

$$
R(\theta) = \mathbb{E}(V\widetilde{\Sigma}V'), \tag{14}
$$

where $\widetilde{\Sigma} = \Sigma^2$ and V is shared by the map from structural to reduced form parameters and the Information matrix. In this case the results of identification exploration would be identical whether one explores $R(\theta)$ or T.

Since the nullspace of the linear map is usually much smaller than the range (column space) of the map it is easier to analyze and derive restrictions from there. The nullspace is spanned by columns of $V_2 = [v_{r+1}, \ldots, v_n]$, an $(n \times n - r)$ matrix. From right to the left the $(n \times 1)$ vectors provide the basis for the spaces associated with the associated eigenvalue – zero or numerically zero. As indicated, $\|v_k\| = 1$

and $v_i v_j' = 0$ for $i \neq j$. Since for each v_i in the nullspace $Tv_i = 0$, using the column view it is intuitive that elements of v_i determine linear dependent columns.

Note that this type of exploration of, for instance, the identity matrix would result into the set of identical singular values and all elements of V, U with unitary vectors forming the standard basis.

To enhance the intuition further, we demonstrate three important and clear cases which can help in understanding of identification patterns. Assume for the moment that in the vector of parameters θ the element θ_1 is completely redundant from the system, not affecting the criterion function and thus completely unidentified. Furthermore, assume that parameters θ_2 and θ_3 affect criterion function importantly, yet are perfectly positively collinear and enter the system as $(\theta_2 + \theta_3)$. Finally let θ_4, θ_5 enter the system as $(\theta_4 - \theta_5)$, while θ_i for $i = \{6, \ldots, n\}$ are well identified and unrelated to $\theta_{1,\ldots,5}$. The nullspace given by V_2 is then as follows

$$V_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -s & 0 \\ 0 & s & 0 \\ 0 & 0 & w \\ 0 & 0 & w \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix}. \tag{15}$$

We have three obvious *identification patterns.* First, for θ_1 we see that first column of V_2 suggest that the parameter is strictly unrelated to other parameters and that it is completely unidentified, since it is fully mapped to nullspace. It is not identified due to lack of influence. The second column of V_2 implies perfect dependency between the two parameters, while indicating no similarity of these on the criterion function to other parameters. An identical increase in one parameter together with identical decrease in the other does not affect anything, trivially. The third pattern is obvious.

Thus we need to adopt three restrictions, respecting the structure of V_2 in order to make the system identified in a lower-dimensional space $\mathcal{K}$, smaller by three dimensions. Thus we must restrict θ_1 to a particular value and at least one element from (θ_2, θ_3), (θ_4, θ_5) pairs. Fixing e.g. $\theta_{1,4,5}$ obviously does not solve the identification problem.

When the problem is the *selection of a subset of variables* to restrict, the goal is to leave those close to orthogonal with others, so that the conditioning number of the sub-problem is maximized. Bellow we introduce a method based on the QR with column-pivoting decomposition of V_1' which delivers a sorted list of parameters (columns) that are as much independent as possible given the original problem.

More complex patterns For truly rank-deficient problems (non-identification) the properties of the nullspace of T can be traced into very detail. The dimension of the nullspace (nullity) equals to the number of *free parameters,* corresponding to non-pivotal columns of the matrix. Working out full details analytically is always possible, yet gets more difficult with increased dimension.

The procedure is as follows – (i) identify basic and free parameters of the model, (ii) construct the solutions by setting each free parameter to unity and others to zero, one by one. The free parameters are the easily determined by the inspection of *row Echelon form.*

For large dimensional parameter spaces the inspection of fully unidentified and weakly identified patterns can be facilitated by plots of v_i's so the spatial patterns and strength of dependence can be checked, as we demonstrate in next section. Also our experience with having a simple program that produces list of unidentified parameters and those weakly identified proved to be useful.

3 Identification Strength

This section discusses the case where all parameters are identified, yet they may be *weakly identified* in the sense that their individual impact on the criterion function is very small or there are some linear combinations of parameters very close to linear dependence. Again the SVD with its properties is the ideal tool for analyzing the strength of parameters identification – finding close-linear combinations, sorting them according to their importance and selecting a subset of parameters (columns) with maximal linear independence.

At this stage we have determined whether there is some clear-cut identification problem embodied in our linear mapping T by calculating the rank of the matrix and isolating basis vectors for the nullspace to uncover the cause of the trouble.

The computational task of rank determination is however difficult since computers calculations do not operate under ideal precision. A singular values thus may be close to zero, but not a true zero. Luckily again the SVD is an extremely valuable tool for determination of the numerical rank of a matrix, see e.g. Golub and van Loan (1996) or Higham (1996).

Let us use the label *approximate nullspace* for the nullspace of the linear map for the basis associated with set of singular values considered as small by the researcher. In what follows we show how to explore approximate nullspace, how to interpret results from the SVD in terms of weak identification and, finally, provide an algorithm for selection of subset of parameters to be estimated.

3.1 Near collinearity

The problem of a *near linear dependence* or a near singularity is crucial for strength of identification. It is well known that the linearly independent vectors (spaces) are orthogonal, i.e. perpendicular. The *cosine* of angle of two vectors is equal to *correlation* of two vectors and closely related to their dot product –

$$\text{corr}(x, y) = \cos\alpha = \frac{x'y}{(\|x\|\|y\|)}. \tag{16}$$

Basically what SVD does is to make similar things more similar and dissimilarities become more dissimilar and decorrelates the identification patterns, without using a 'correlation' analysis. The collinearity can be among multiple columns, hence bivariate relationships or correlations are only necessary, not sufficient signs of rank deficiency or weak identification. Further, recall that dim range(T) = dim range(T'), so rows are considered as well.

For multiple columns of the $(m \times n)$ matrix $T = [t_1 \dots t_n]$ the full collinearity is present in the case

$$\sum_{i=1}^{n} \gamma_i t_i = 0, \tag{17}$$

that is – the linear combinations of columns is zero for $\{\gamma_i\}_{i=1}^n$ which are not all zero. Often, we do not have perfect collinearity, but only strong collinearity. As we repeatedly point out, it would be useful to find out the linear dependencies that are sorted out according to their distance to zero.

That is, basically, what SVD delivers. Note that since $T = U\Sigma V'$ we have for $j \in [1, p]$, $p = \min\{m, n\}$ and in our case $p = n$, and $Tv_i = \sigma_i u_i$ where v_i, u_i are of unit length. For $\sigma_i \to 0$ it is clear that $\sigma_i u_i \to 0$ which implies that $Tv_i \to 0$. Now we clearly need to adopt again the column-view of the operation to see that

$$Tv_i = [t_1, \dots, t_n]v_i = \sum_{k=1}^{n} t_k v_{k,i} \to 0 \qquad \text{and} \qquad \|Tv_i\| = \sigma_i. \tag{18}$$

The elements of each right singular vector v_i associated with σ_i determine the coefficients γ_k determining the combination closest to zero, i.e. closest to collinearity. For important patterns of multicollinearity most of γ_k coefficients go fast towards zero. Since we have $\|Tv_i\| = \sigma_i$ we can re-normalize coefficients in the linear form and express this as norm of residual, where $\varepsilon_{i,k} \equiv t_k - (\hat{\gamma}_1 t_1 + \dots + \hat{\gamma}_{k-1} t_{k-1} + \hat{\gamma}_{k+1} t_{k+1} + \dots + \hat{\gamma}_n t_n)$ and $\|\varepsilon_{k,i}\| = \sigma_i / |\gamma_k|$. For perfect collinearity the linear combination attains zero norm of the error and thus *multiple correlation coefficient* between t_k and the rest of selected vectors is unity in that case. In other cases, the SVD produces linear combinations whose coefficients are not coincident with those obtained from

multiple correlation coefficient, but always less or equal.

The use of the SVD is closely related to *principal components analysis* (PCA). The identification of identification patterns can thus be regarded also as indicating principal components of the Information matrix if this is the case.

Scaling of the Problem When analyzing the matrix T or $R(\theta)$ using singular value decomposition, the scaling of the matrix matters for the strength of identification analysis. Clearly it does not matter for detecting strict linear combinations since these are immune to rescaling. SVD is however dependent on scaling and thus rescaling of columns is often advisable if units of parameter differ greatly. This may sometimes be the case with DSGE models, yet the situation is not extreme.

For analyzing linear map of structural to reduced form parameters T rescaling of columns to have equal length is advisable. A common scaling is also by the absolute values of individual parameters, e.g. by $\Gamma = [|\theta_1|, \ldots, |\theta_n|]$ and $\Gamma T \Gamma$. In case of $R(\theta)$ one may decide to rescale to $R_r(\theta) = \Gamma^{-1/2} R(\theta) \Gamma^{-1/2}$ where $\Gamma = \mathrm{diag} R(\theta)$ and thus both row and column scaling is carried out. Obviously, the rescaling using variances from the Information matrix is feasible only after perfect linear collinearity is extracted from the system.

The condition number is invariant to multiplication of the matrix by a constant – note that in case of determinant opposite is the truth.[5] The condition number is however not invariant to the rescaling of individual columns. Partition columns of A as $A = [A_1 \ \alpha a_k]$, where a_k is one column and α is scalar, then by letting $\alpha \to 0$ the condition number $\kappa(A) \to \infty$, an effect referred to as *artificial ill-conditioning.* Thus large differences in scaling of the columns make condition numbers large.

The question is what is thus the most 'natural' scaling for the problem. Our view is that either no scaling or scaling by the absolute value of parameters – parameters as used in the estimation. The reason is that in case of searching for extreme of the criterion function during the estimation a *gradient-based method* rely on the score vector and its relation to SVD of the Hessian. The information embodied in the SVD can be also successfully used to enhance the optimization routine.

To complicate things a little bit more, one must be aware that by appropriate rescaling the structure of V_2 changes, which for non-precise calculation is an issue. It may be not sufficient always to inspect the structure of V_2, since for any nonsingular matrix A we have $XV_2A = 0$ with altered zero structure of linear dependency, so the problem should be put into linear transformation invariant setup, see Belsley and Klema (1974), where inspection of $G \equiv -V_{22}V_{21}^{-1}$ is analyzed.

[5] Although zero determinant implies linear dependence, the use of determinants for the purpose is numerically very unstable and also a small value of determinant has nothing to do with near collinearity.

Conditioning of the Problem It is a standard result in computation to check for *conditioning* of a matrix, which –simply put– determines sensitivity of the problem $T\theta = \tau$ to small variations in T and τ. Ill-conditioned problems display extreme sensitivity. Condition number is a distance to singularity, nearness of trouble, and the condition number κ equals to ratio of largest to smallest singular value of the linear map matrix.[6] The ratio of the first and the i-th singular value is called *condition index.* The steepness of profile of sorted singular values conveys a lot of information about the decay of identifiability. For singular problems $\kappa \to \infty$, see e.g. Golub and van Loan (1996) or a lucid treatment in Higham (1996), inter alia.

3.2 Parameter subset selection

It would be rather desirable to find a set of k individual parameters out of n, which are the best identified. Those $n - k$ are then restricted in a preferred way. So how should one choose them?

Say we know we have linearly dependent (or collinear) columns of the matrix, whose columns correspond to structural parameters. Let us assume that the nullity of the map is one – hence, we need to eliminate one parameter. It turns out that it often may be easier to find what parameters to retain in the model, than determine those to be discarded. We would like to keep parameters corresponding to set of columns that are as much as possible independent, i.e. that minimize the condition number – given a particular scaling, which is important.

To find the set of k columns to retain, where $k \leq r$ one might experiment and iterate on orderings to get $A = [A_1 \ A_2]$, which is our new column-partition of A so that condition number of A_1 is as small as possible. Interestingly there are algebraic procedures that avoid iteration and deliver plausible results. We follow Golub, Klema and Stewart (1976) as an example of well-established and straightforward procedure to produce 'enough' linearly independent columns of the matrix by manipulating its row-space.[7]

ALGORITHM

[6] At this stage, we note some details on implementation. First, the conditioning of the problem is norm-dependent, our statements relates to 2-norm. Calculating the SVD in, for instance, `Matlab` by Mathworks uses the command `[U S V] = svd(A)`. Further, if you check the rank of the matrix, then again SVD is used and the smallest non-zero singular value is checked with entered tolerance, default tolerance being related to machine ε. Furthermore when the condition number is checked for (deafult) 2-norm the SVD is used again! Hence, once rank conditions are involved the costs with computing an SVD are paid, so why not to use the 'full package' then.

[7] The problem of subset selection is well-known in linear algebra and computer science. Given a matrix $A \in \mathbb{R}^{m \times n}$ and positive k, we want to choose k columns of A forming a matrix $B \in \mathbb{R}^{m \times k}$ such that the residual $\|A - BB^{+}A\|_{\xi}$ is minimized given the combination n^k possibilities and ξ is either spectral or Frobenius norm. A^{+} denotes Moore-Penrose generalized inverse. There are both *deterministic* and *randomized* algorithms for this difficult problem. We follow one of the algorithms that belongs to the oldest, most straightforward, deterministic and with good empirical behavior.

1. Compute SVD of the matrix $T = U\Sigma V'$ and determine k components, parameters that become the new free parameters. Feasible choice of k is $k \leq r = \text{rank}(T)$.
2. Calculate rank-revealing QR factorization with column-pivoting[8]

$$V_1'P = QR, \tag{19}$$

   where V_1' is $(k \times m)$ matrix in $V' = [V_1' \; V_2]'$ and P is the *permutation matrix.*
3. Choose the *subset* of k components of the parameter set θ as $\widehat{\theta} = P'\theta$.

Note the algorithm suggests what parameters retain and what parameters not, yet these groups are not necessarily sorted. However we suggest a heuristic procedure for parameter sorting in such a way that the condition number is monotonically decreasing. The procedure runs backwards. First, select $k = r$ columns of the matrix that are regular. Always partition the matrix into two groups where only one vector is not to be selected. Reduce the matrix and proceed with the selected columns in the same way until the number of columns is reduced to the last one. In case of symmetric matrices one can eliminate each time a corresponding row and column, hence ignore a particular parameter from the system, treating it as fixed. This heuristic backward procedure results into unique ordering, though its results are very sensitive to scaling of the matrix under investigation.

The permutation matrix P permutes the columns of matrix T in such a way that $TP = [T_1 \; T_2]$, where T_1 is $(m \times k)$ and $\sigma_i(T_1) \leq \sigma_i(T_1)$, where $\sigma_k(.)$ denotes the i-th largest singular value. This method is due to Golub et al. (1976) and more complex modifications of the algorithm exist. QR decomposition with column-pivoting, where the rank-deficiency is treated by taking QR decomposition of column-permuted matrix, see e.g. Golub and van Loan (1996).

The logic of the method works as follows. Imagine taking an SVD of column-permuted T, i.e.

$$TP = [T_1 \; T_2] = U \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tilde{V}' = U \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} \tilde{V}_{11}' & \tilde{V}_{21}' \\ \tilde{V}_{12}' & \tilde{V}_{22}' \end{bmatrix} \tag{20}$$

and being lucky so that $\tilde{V}' = I$, where I is identity matrix. Then the 'important columns' become $T_1 = U[\Sigma_1 \; ; 0]$ and $\sigma_i(T_1) = \sigma_i(T)$ for $1 \leq i \leq k$. If we are not as lucky, then V_{11} becomes the key player here (which is obvious for $\Sigma_2 = 0$) and Golub et al. (1976) prove that the largest singular value of T_1 is between lower and

[8] In `Matlab` the command `[q r p] = qr(A)` delivers QR with column pivoting with the permutation matrix P.

upper bonds given by

$$\frac{\sigma_k(T)}{\|\tilde{V}_{11}^{-1}\|} \leq \sigma_k(T_1) \leq \sigma_k(T). \tag{21}$$

So having $\tilde{V}_{11}$ well-conditioned as much as possible is important. Yet note that the SVD is decomposition of column-permuted matrix T, hence $\tilde{V} = VP$ and thus column permutation matrix P may be chosen so that $\tilde{V}_{11}$ is well-conditioned. This is achieved by the QR column-pivoting factorization of the $H \equiv [V_{11}' \; V_{21}']$ from $T = U\Sigma V$ as $Q'HP = [R_{11} \; R_2]$ where well-conditioned R_{11} is upper triangular and since multiplication by ortogonal matrix preserves norms $\|\tilde{V}_{11}^{-1}\| = \|R_{11}\|$. How small can $\|\tilde{V}_{11}^{-1}\|$ actually be can be determined from properties of V and $\|\tilde{V}_{11}^{-1}\| \leq \sqrt{1+k(n-k)}$ if $|\det\tilde{V}_{11}|$ is maximal.

The singular value decomposition itself is not a tool for subset selection or sparsity enhancement, but it can be made a part of the procedure. It decomposes a matrix A into a layers, sum of p matrices of rank one. This is the principle of principle components analysis, which is closely related to singular and eigenvalue decompositions and, the procedures discussed in the paper may be treated as as principal component analysis of the identification problem.

Graphical investigation One can –and we find it very useful– explore identification patterns graphically. We produce plots ('heat maps') of individual identification patterns embodied in v_i so that we plot $v_i v_i'$ matrix with individual elements values represented by a particular color on a scale. Plotting the v_i and its associated singular value often allows quick inspection of parameters involved in non- or weak identification. Obviously one can group v_is or plot the whole nullspace.

3.3 The problem of correlation measures

Correlations measures may not be always reliable measures of weak identification and so this section is devoted to issues related to using correlation measures as tools to detect multicollinearity and thus as measures to detect identification problems. The discussion uses theoretical arguments, a numerical example and makes use of some results in Golub and van Loan (1996), Belsley and Klema (1974) and Belsley et al. (1980).

The use of correlation as a measure of collinearity is very intuitive, since the correlation of two vector amounts to a cosine of the angle of these vectors in n dimensional space. As the bivariate correlation $\rho_{i,j}$ gets closer to ± 1 the vectors are becoming more and more collinear.

However, the singularity may results from not only two, but more vectors involved in a linear combination. In this case one may find very low bivariate correlations

among vectors, while the matrix is effectively singular – as easily detected using the SVD. An intuitive solution to the problem is then a coefficient of multiple correlation.

Multiple correlation coefficient (Anderson 2003) is defined as the maximum correlation between x_i and the linear combination αX, where X is appropriate matrix and α is a vector. To get the maximum correlation, the vector α is formed by projection coefficients. Anderson (2003, pp.38, 145) also provides several useful formulas for calculating the multiple correlation coefficient. For instance it can be shown that for covariance matrix $S = [s_{ij}]$ the multiple regression coefficient for the first-left vector x_1 with respect to others can be expressed using formula

$$1 - R_i^2 = \frac{|S|}{s_{11}|S_{22}|}, \tag{22}$$

where $|.|$ denotes determinant. As we have already mentioned, determinant is extremely unreliable measure of collinearity, yet if it is zero $R_1^2 \to 1$ and one is tempted to carry out detailed limit analysis of the ratio in the formula. The problem is that small, but nonzero determinant may have nothing to do with collinearity. Another venue must be taken.

Importantly Marquardt (1970) demonstrates that when R is correlation matrix, then the diagonal elements of R^{-1} contains the *variance inflation factors,* (VIF), where $\text{VIF}_i = 1/(1 - R_i^2)$. When the VIF is large the multiple correlation coefficient increases. Assume now that matrix B has normalized columns (in a 'regression form') so that $R = B'B$ is the correlation matrix and thus diagonal of $(B'B)^{-1}$ features the VIFs. Taking the SVD we get $B = U\Sigma V'$ and thus $R^{-1} = V\Sigma^{-2}V'$, so that the individual VIFs can be expressed as

$$\text{VIF}_k = \sum_{j=1}^{n} \frac{v_{kj}^2}{\sigma_j^2} = \sum_{j=1}^{n} \left(\frac{v_{kj}}{\sigma_j} \right)^2, \tag{23}$$

where σ_k and v_k are singular values and right-singular vectors of B. When there is not perfect-collinearity, the elements of v_k are not true zeros, though some of the may be small. Note that for each VIF_k all singular values are used – including those potentially very small in case of near collinearity. Then we have that for some r-th element

$$\sigma_r \to 0 \text{ and } v_{kr} \neq 0 \qquad \text{VIF}_k \to \infty \qquad R_k \to 1. \tag{24}$$

A single near-collinear relationship can thus blow-off all variance inflation factors and thus all multiple correlation coefficients among vectors in the matrix. Further, Belsley et al. (1980) demonstrate that this behavior is shared by all bivariate correlations and thus partial correlation coefficients.

Numerical example To demonstrate these facts in a simple setup, we choose to inspect matrix B of the form

$$B = [B_1\ b_5] = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.20 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.40 \\ 0.00 & 0.00 & 0.00 & 1.00 & -0.70 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}, \tag{25}$$

where the first four columns are formed as $B_1 = I_4 + \mathcal{E}$, where I_4 is identity and $\mathcal{E}$ draw from Gaussian distribution with almost zero variance and $b_5 = 0.2b_2 + 0.4b_3 - 0.7b_4 + \epsilon$, so that the matrix is not singular and features one nearly collinear relationship where three columns are involved.

The resulting correlation matrix, VIFs and R_i^2 coefficients are then

$$R = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.24 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.48 \\ 0.00 & 0.00 & 0.00 & 1.00 & -0.84 \\ 0.00 & 0.24 & 0.48 & -0.84 & 1.00 \end{bmatrix}, \text{ VIF} = 10^{12} \begin{bmatrix} 0.00 \\ 0.40 \\ 1.60 \\ 4.92 \\ 6.93 \end{bmatrix} \quad R^2_{1:5} = \begin{bmatrix} 0.9679 \\ 0.8501 \\ 0.9923 \\ 0.8290 \\ 0.9941 \end{bmatrix}. \tag{26}$$

The matrix is not extremely badly behaved since the few first columns are almost mutually orthogonal and there is only one small singular value, since $diag(\Sigma) = [1.3\ 1\ 1\ 1\ 2.42 \times 10^{-7}]$.

The facts above may perhaps partly explain why multiple correlation coefficients in Iskrev (2009a) for the model of Smets and Wouters (2007) are all but three, out of 39, larger than 0.99 for all parameter combinations since few truly unidentified parameters are kept in the set of regressors.

4 Identification Examples

We demonstrate the method explained above using three examples of well-established models – a small-scale model by An and Schorfheide (2006), a medium-scale model by Smets and Wouters (2007) and –in the appendix– a small open economy model as popularised by Monacelli (2003) and Justiniano and Preston (2004).

We have also used the method for other DSGE models, for instance, in Andrle, Hlédik, Kameník and Vlček (2007–2008) or Steinbach, Mulhoe and Smit (2009)[9] to identify the strength of identification and identification patterns.[10]

[9]The author thanks Research Dept. of the South African Reserve Bank for warm hospitality during November 2008.

[10]For computations we have used the IRIS-Toolbox for Matlab by Jaromír Beneš, an objected-oriented

4.1 An-Schorfheide (2006) Model

The log-linear version of the model by An and Schorfheide (2006) consists essentially of a closed economy IS curve without lagged term of output, a forward-looking Phillips curve,a consumption identity and an interest rate rule with smoothing, contemporaneous effect of inflation and output growth. The model features three measurement variables – output growth, inflation and nominal interest rates. There are three stochastic shocks – exogenous government spending, technology shock and uncorrelated monetary policy innovation.

We analyze only local identification at coefficient values from the first data generating process (DGP1). Following An and Schorfheide (2006) we define reduced form parameter κ for the slope of the Phillips curve, since it individual components are not identified and we would trivially obtain rank-deficient Information matrix with right-singular vectors pointing to these components, associated with zero singular values. Contrary to authors we do not analyze the intercept parameters γ, r^A, π in measurement equation and focus only on the set of ten parameters $\theta = \{\tau, \kappa, \phi_1, \phi_2, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_z\}$, denoting the coefficient of risk aversion, slope of the Phillips curve, interest rate rule weights on inflation and output growth, interest rate smoothing parameter and persistence and standard deviations of exogenous stochastic processes. For computing Information matrix we use the $T = 80$. observations.

An and Schorfheide (2006, pp. 19–20) comment that the visual inspection of prior and posterior distributions of their estimation indicate that the sample contains little information on the risk-aversion coefficient τ and policy rule coefficients ϕ_1, ϕ_2, whereas data are informative about the slope of the Phillips curve κ and also autocorrelation and standard deviation of stochastic shocks.

Inspection of log-likelihood sensitivities with respect to individual parameters[11], (An and Schorfheide 2006, Fig. 14), indicates that the log-likelihood is rather flat with respect to τ, ρ_g and ϕ_2 in the neighbourhood of the true parametrization. On the other hand the curvature of $\sigma_R, \sigma_g, \sigma_z$ or ρ_R is reasonable. Authors provide the sensitivity with respect to components of κ, which are not flat, though it is the linear dependence that prevents the identification of these parameters.

On the basis of An and Schorfheide (2006) we would expect the methods presented in this paper to indicate good local identification of the slope of the Phillips curve κ, standard deviations of stochastic shocks and weaker identification of risk-aversion parameter τ and interest rate rule parameters ϕ_1, ϕ_2. The issue is how the possible interaction among parameters affects the information from the log-likelihood curvature with respect to individual parameters.

toolbox for developing and using DSGE models – `www.iris-toolbox.com`.

[11] This is a version of the often used 'happy faces' plot.

We analyze the Information matrix without any scaling. It is of full rank and identification patterns (i.e. righ-singular vectors) plots are at Fig. 1. Since the Information matrix is not scaled the condition number is large and the profile of condition indices sharply drops after the sixth dimension as can be easily seen from singular values of identification patterns.

The plots seem to support the results of An and Schorfheide (2006) that parameters $\kappa, \sigma_{R,z,g}$ are (relatively) well identified. Another well identified parameter is interest rate smoothing parameter ρ_R, though we can observe that the parameter interacts with ϕ_2 and τ, which is intuitive. On the other hand the persistence of the government spending is less identified, though not due to collinearity but due to its small impact on the likelihood, i.e. the likelihood is flat with respect to this parameter.

The least identified parameter seems to be the risk-aversion coefficient both due to its small impact on the likelihood and partial confounding with ϕ_1, the interest rate rule weight on inflation as can be viewed from the identification patterns 9 and 10. From the economic point of view the higher τ decreases the impact of interest rate changes on the output gap in the IS curve, which is the driving force of the inflation. To stabilize inflation with lower τ the policy authority needs to increase the weight on inflation ϕ_1 in the interest rate reaction function.

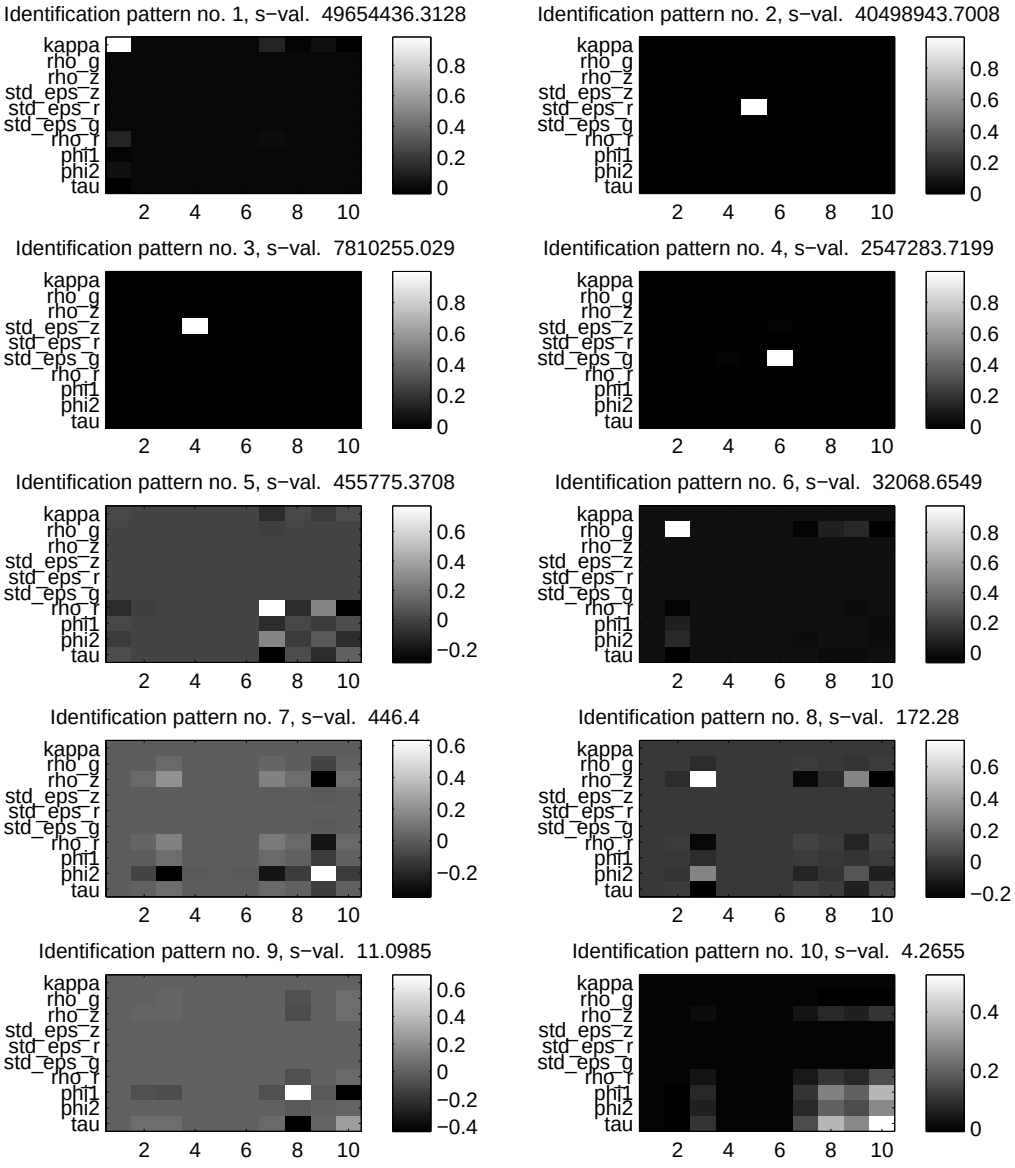
The next step is the application of the heuristic procedure to 'order' the parameters in terms of their identifiability. More precisely, we carry on repeated subset-selection problem using the rank-revealing QR factorization as indicated above. The backward-pass algorithm delivers the following vector of sorted coefficients $\tilde{\theta} = \{\kappa, \sigma_R, \sigma_z, \sigma_g, \rho_R, \rho_g, \phi_2, \rho_z, \phi_1, \tau\}$, which seems to be broadly in line with the identification patterns and the discussion in An and Schorfheide (2006).

4.2 Smets-Wouters (2007) Model

To test the method on the model by Smets and Wouters (2007) we use the model code and prior and posterior mode made public by authors. However our implementation of the model may differ in minor details or by our omissions.

We inspect the FIM using $T = 200$ and both in its unscaled version and in its version scaled by parameter size. We also provide a correlation-version of the FIM –after eliminating singularity– and calculate variance inflation factors (VIFs) and corresponding righ-singular vectors pointing into space associated with large components of VIFs. The FIM was calculated numerically, so it is sensitive to numerical inaccuracies. In the two-step numerical differentiation the differentiation step reflects the absolute value of the parameter, e.g. the numerical step for adjustment costs is larger than for standard deviation of monetary policy shock.

Fig. 1: Identification pattern – (An and Schorfheide, 2006)



Tab. 1: Parameter subset selection – SW-07 model

1	ρ_w	*11*	ξ_p	*21*	σ_c	*31*	α
2	$\bar{\gamma}$	*12*	σ_{eg}	*22*	$r_{\Delta y}$	*32*	σ_l
3	ρ_p	*13*	σ_r	*23*	cgy	*33*	ρ_r
4	ρ_g	*14*	σ_{eb}	*24*	μ_p	*34*	$\bar{\pi}$
5	ρ	*15*	σ_{ea}	*25*	φ	*35*	$\bar{l}$
6	ρ_a	*16*	ξ_w	*26*	ψ		
7	μ_w	*17*	σ_w	*27*	ι_w		
8	λ	*18*	r_π	*28*	r_y		
9	ρ_I	*19*	σ_p	*29*	ρ_b		
10	Φ	*20*	σ_I	*30*	ι_p		

We analyze only those parameters actually estimated by Smets and Wouters (2007), hence the depreciation rate δ, share of government spending g_y, steady-state labor market markup λ_w and Kimbal aggregator parameters ϵ_p and ϵ_w are treated as fixed. The last three parameters lead otherwise to rank-defficiency due to their collinearity with ξ_w and ξ_p, Calvo parameters.

By inspection of 'identification patterns' we can see very complex interactions among virtually all parameters. That implies that it would be difficult and insufficient to rely only on bivariate correlations of coefficients and that multiple correlation coefficients (and thus variance inflation factors) would be high in general.

Notable exceptions with none or smaller interactions, clearly tractable to corresponding right-singular vectors, are $\bar{l}$, $\bar{\pi}$ and $\bar{\gamma}$ which determine steady state level of hours worked, inflation and trend growth rate, and others are ρ_g, σ_b, σ_r or ρ, determining auto-regression of exogenous govt. spending, variance of preference and monetary policy shock or interest rate smoothing parameter, to smaller degree. With little interactions the first two parameters are very poorly identified, whereas the the second group belongs to rather well identified parameters, judged using the analysis of singular values.

The identification patterns of most other parameters are rather complex and would require lengthy analysis, which we attempt to by-pass by application of subset selection algorithm to unscaled and parameter size scaled FIM. The results of parameter ranking are listed in Tab. 1 for the FIM scaled by the relative size of parameters.

It seems that weakly identified parameters are $\bar{l}$, $\bar{\pi}$ determining the steady state, i.e. constant terms, and autocorrelation of monetary policy shocks ρ_r, followed by σ_l and α, determining Frisch labor elasticity and the share of capital in the production of intermediate goods. Calvo parameters for inflation and wages ξ_p and ξ_w do not

belong to best identified parameters, but seem identified better than indexation parameters ι_p and ι_w. The parameter affecting intertemporal elasticity of substitution σ_c also does not belong to the best identified parameters.

Among the best identifiable parameters one can find autoregression coefficient of wage and inflation cost-push shocks ρ_w, ρ_p, government exogenous process persistence ρ_g, interest rate smoothing parameter ρ and persistence of technology shock ρ_a, habit formation parameter λ or investment specific shock persistence ρ_I. In our implementation also the trend parameter $\bar{\gamma}$ seems to be well identified, we have excluded β.[12]

Concerning parameters in the interest rate rule the smoothing parameter ρ is best identified, followed by inflation coefficient r_π, output-gap difference $r_{\Delta y}$ and the least identified coefficient on level of the output gap r_y.

The identification of the Smets and Wouters (2007) model is explored in detail in Iskrev (2008) and Iskrev (2009a), where the first paper – among other measures– uses bivariate correlation analysis and the second multiple correlation coefficient analysis. The conclusions from the set of extremely large multiple correlation analysis, most above 0.99, point towards rather poorly identified coefficients of the model and presence of large structural redundance of the model's features. However opposed to Smets and Wouters (2007) or our analysis author allows for λ_w, extremely poorly identified parameter, to enter these multiple correlation coefficients. Our analysis suggests that it is few of close collinearities that cause multiple correlation coefficient to blow-off.

[12] As for implementation of the model $\bar{\gamma}$ parameter is rescaled differently than in Smets and Wouters (2007) and thus results may be affected.

5 Conclusion

This paper demonstrates both simple and potentially powerful method for exploring the 'identification patterns' of DSGE models. The identification patterns are associated with the nullspace of a particular linear map investigated. The binary problem of either identification or the strength of the identification is showed to be naturally analyzed with the conditioning of the map, which also sorts the identification patterns in terms of their strength.

The method is able to indicate identified and weakly identified patterns of the parameter space, while suggesting whether this is due to lack of influence of the parameter or its interactions with other parameters. The method seems very useful and can be used for a-priori investigation of the model identification before the model is estimated at all. The local nature of the method can make a step towards global identification search using simple pseudo- or random simulation schemes or at particular border parametrisations. The method is not dependent on a particular estimation method.

The location of the unidentified subspace of the parameter space, sorting of identification patterns with respect to their strength and determination of the rank condition of the identification problem – for all these tasks we demonstrate that singular value decomposition of the matrix of the linear map is a strong candidate. In contrast to an eigenvalue decomposition the singular value decomposition may be applied also to only rectangular, not square, matrices of Jacobians (linear maps) from structural to reduced form parameters to analyze in-depth the rank condition of identification.

We also suggest a heuristic method for ordering the parameter vector in terms of individual element's 'identifiability' by carrying out a repeated sub-set selection problem using rank revealing factorizations, which takes into account both flatness of the criterion function and parameter confounding.

References

Aldrich, J., "How Likelihood and Identification went Bayesian," *International Statistical Review*, 2002, *70* (1), 79–98.

An, S. and F. Schorfheide, "Bayesian Analysis of DSGE Models," 2006. Federal Reserve Bank of Philadelphia, WP No. 06-5.

Anderson, T.W., *An Introduction to Multivariate Statistical Analysis,* 3^{rd} *ed.*, Wiley Series in Probability and Statistics, New Jersey, 2003.

Andrle, M., T. Hlédik, O. Kameník, and J. Vlček, "Putting in Use the New Structural Model of the CNB," 2007–2008. Czech National Bank Working Paper, forthcoming.

Axler, S., *Linear Algebra Done Right,* 2^{nd} *ed.*, Springer, 1997.

Belsley, D., E. Kuh, and R. Welsch, *Regression Diagnostics – Identifying Influential Data and Sources of Variation*, Wiley Series in Probability and Statistics, New York, 1980.

Belsley, D.A. and V. Klema, "Detecting and Assessing the Problems Caused by Multicollinearity: A Use of the singular-value decomposition," 1974. NBER WP No. 66.

Beyer, A. and R. Farmer, "Indeterminancy of Determinacy and Indeterminacy," *American Economic Review*, 2007, *97*, 524–529.

Canova, F. and L. Sala, "Back to Square One, Identification Issues in DSGE Models," 2006. ECB WP No. 583.

— **and** — , "Back to Square One, Identification Issues in DSGE Models," *Journal of Monetary Economics*, 2009, *56*, 431–449.

Cochrane, J., "Identification with Taylor Rules: A Critical Review," 2007. mimeo.

DiStefano, J.J., "On the Relationship Between Structural Identifiability and the Controllability, Observability Properties," *IEEE Transactions on Automatic Control*, 1977, *AC-22* (4 (August)), 652–652.

Doren, J.F.M. Van, P.M.J. Van den Hof, J.D. Jansen, and O.H. Bosgra, "Determining Identifiable Parametrizations for Large-Scale Physical Models in Reservoir Engineering," 2008. Proceedings of the 17th World Congress, The International Federation of Automatic Control, Seoul, Korea, July 6–11.

Dréze, J., "Econometrics and Decision Theory," *Econometrica*, 1972, *40*, 1–17.

— , "Bayesian Theory of Identification in Simultaneous Equations Models," in S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*, North-Holland 1975, pp. 159–174.

Fisher, F.M., *The Identification Problem in Econometrics*, New York: McGraw-Hill, 1966.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis,* 2^{nd} *ed.*, Chapman & Hall/CRC, 2004.

Glover, K. and J.C. Willems, "Parametrizations of Linear Dynamical Systems:

Canonical Forms and Identifiability," *IEEE Transactions on Automatic Control*, 1974, *AC-19* (6), 640–646.

Golub, G.H. and F. van Loan, *Matrix Computations*, Baltimore: Johns Hopkins University Press, 3^{rd} Ed., 1996.

— **, V. Klema, and G.W. Stewart**, "Rank Degeneracy and Least Squares Problems," 1976. Technical Report TR-456, Dept. of Computer Science, University of Maryland, College Park, Maryland.

Grewal, M.S. and K. Glover, "Identifiability of Linear and Nonlinear Dynamical Systems," *IEEE Transactions on Automatic Control*, 1976, *December* (6), 833–837.

Guerron-Quintana, P., A. Inoeu, and L. Kilian, "Frequentist Inference in Weakly Identified DSGE Models," 2009. Federal Reserve Bank of Philadelphia WP No.09-13.

Hannan, E., "The Identification Problem for Multiple Equation Systems with Moving Average Errors," *Econometrica*, 1971, *39* (5), 751–765.

Harvey, A., *Forecasting, structural time series models and the Kalman filter*, Cambridge: Cambridge UP, 1989.

Higham, N.J., *Accuracy and Stability of Numerical Algorithms*, Philadelphia: SIAM Publications, 1996.

Hoerl, A.H. and R.W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 1976, *42* (1, February), 80–86.

Hsiao, Ch., "Identification," in Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Vol. I*, Elsevier 1983.

Iskrev, N., "How Much Do We Learn from Estimation of DSGE models? A Case Study of Identification in a New Keynesian Business Cycle Model," 2008. mimeo, University of Michigan.

— , "Evaluating the strength of identification in DSGE models. An a priori approach," 2009. Bank of Portugal mimeo, paper presented at ECB WGEM, 30 Nov.

— , "Local Identification in DSGE Models," 2009. Banco de Portugal Working Paper No.7.

Justiniano, A. and B. Preston, "Small Open Economy DSGE Models: Specification, Estimation and Model Fit," 2004. mimeo.

Kailath, T, *Linear Systems*, Englewood Cliffs, N.J.: Prentice-Hall, 1980.

Komunjer, I. and S. Ng, "Dynamic Identification of DSGE Models," 2009. University of Michigan, November.

Marquardt, D.W., "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, 1970, *12*, 591–612.

Monacelli, T., "Monetary Policy in a Low Pass-Through Environment," 2003. ECB WP. No 227.

Moon, H.R. and F. Schorfheide, "Bayesian and Frequentist Inference in Partially Identified Models," 2009. Univ. of Pennsylvania and Univ. of Southern

Cal., mimeo.

Pesaran, H., *The Limits to Rational Expectations*, Blackwell, 1987.

Poirier, D.J., "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 1998, *14*, 483–509.

Rothemberg, T.J., "Identification in Parametric Models," *Econometrica*, 1971, *39* (3), 577–591.

Sargent, T., "Estimation of Dynamic Labor Demand Schedules under Rational Expectations," *Journal of Political Economy*, 1978, *86* (6), 1009–1044.

Simmons, G.F., *Introduction to Topology and Modern Analysis*, Krieger Publishing Company, 2003.

Smets, F. and R. Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 2007, *97* (3), 586–606.

Staiger, D. and J.H. Stock, "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 1997, *65*, 557–586.

Steinbach, R., P. Mulhoe, and B. Smit, "An Open Economy New Keynesian DSGE Model of the South African Economy," 2009. South African Reserve Bank, April.

Stock, J.H., J.H. Wright, and M. Yogo, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economics Statistics*, 2002, *20* (4), 518–529.

Strang, G., *Linear Algebra and Its Applications,* 3^{rd} *ed.*, Hardcourt Brace Jovanovich, 1988.

— , "The Fundamental Theorem of Linear Algebra," *American Mathematical Monthly*, 1993, *100* (9), 848–855.

Vajda, S., H. Rabitz, E. Walter, and Y. Lecourtier, "Qualitative and Quantitative Identifiability Analysis of Nonlinear Chemical Kinetic Models," *Chem. Eng. Communications*, 1989, *83*, 191–219.

APPENDIX – Basic SOE Model

The model is a variant of Monacelli (2003) and Justiniano and Preston (2004) as described also in Steinbach et al. (2009). The only modification to Steinbach et al. (2009) is adding possibility of partial price indexation for imported prices $\pi_{f,t}$ and that we simplified the foreign block – inflation, output and interest rates– using a simple VAR specification – consequently, we do not estimate any of these autoregressive parameters or variances. The results for 'home parameters' are not affected by this change. This paper is not an empirical and this section is just an examples section, so we do not scrutinize the model into details, just indicate the use of the method.

Likelihood The parameter space then consists of 30 structural parameters, including standard deviations. In terms of estimation method we focus on inspecting the properties of the *log-likelihood* of the model, so we make use of full-information method and our results are thus relevant even for other method of estimations. We *do not estimate* the model, rather we carry out a-priori analysis of the identification, since identification is the feature of the model, not of data. For computing the Information matrix, that i.s. *expected value* of the Hessian of the log-likelihood we sample from the prior-mode of the model, so no real data are used.

We set identical sample length for time and frequency domain likelihood function. We mercifully choose $T = 200$ for our experiments. We start the evaluation of the Information matrix at the prior mode of the model.

Naturally the analysis –except of truly linear dependent combinations– is not scale invariant. We have thus also carried out the analysis for (i) unscaled Fisher information matrix (FIM), (ii) FIM scaled by the size of the parameters and (iii) correlation form of FIM – after singularity has been eliminated. We do not in this case normalize the FIM so all columns have unit norm, since we do not consider this plausible in case of identification analysis.

Identification Patterns After calculating the Information matrix, we find out a singular value decomposition to find subspaces associated with the operator implied by the matrix. There is no absolute zero – we didn't let any parameter just float in the air. However we evaluate the rank of the matrix –i.e. dimension of the column space– to be only $r = 26$, which implies four very problematic identification patterns. Fig. 2 depicts the evolution of singular values and log-singular values of the model. The 24-th singular value is still $9.4094e-5$. The shape of the singular values profile also suggest a significant portion weak identification resulting from the model structure.

Let's proceed to the inspection of the nullspace of the map and associated four identification patterns. Note that since there is no true zero, the patterns are, as advertised, sorted in order of their problemacy. After the nullspace we may inspect those weakly and those strongly identified portions of the parameter space – the *approximate nullspace*. We have checked that the parameters identified by the procedure above result into their posterior distribution equal to their prior distribution in Bayesian setting.

The first hopelessly unidentified pattern consist of ξ_w (`xi_empl`), denoting the CES parameter for labor packers that subsequently appears in the wage Phillips curve. The pattern shows it is not interacting with anything, but it does not contribute to anything. This should not come as a surprise. The corresponding vector includes 0.99 at the location of the parameter in θ and zeros or very little numbers (-0.02 for θ_w, the wage Calvo parameter).The second identification pattern points to α_w (`alpha_w`), the indexation parameter and the pattern is similar to previous, since the parameter is not interacting with others.

As an example we plot the v associated with the pattern in Fig. 3 to demonstrate the logic and the intuitive nature of our plots. The white square is associated with α_w, all

Fig. 2: Singular values – small open economy model

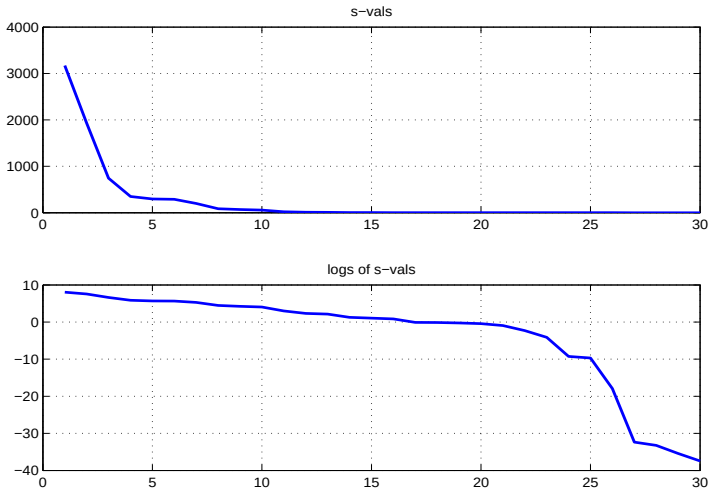
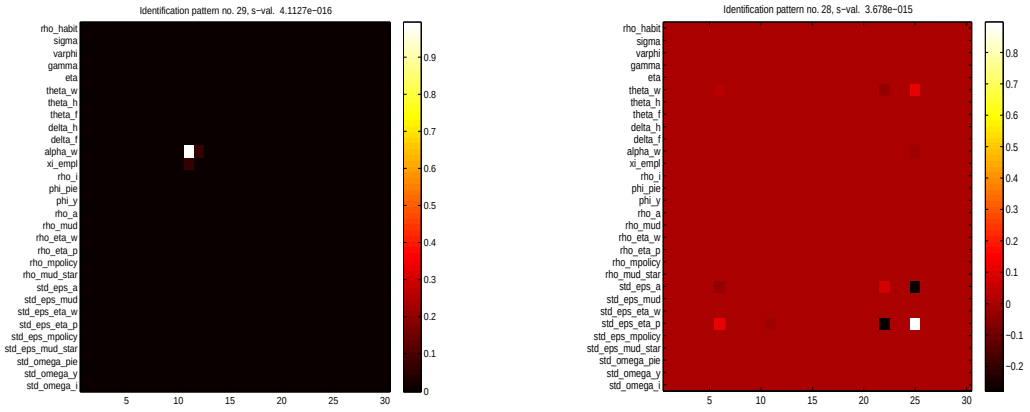


Fig. 3: Identification pattern – SOE model



other coefficients are corresponding to zero. The colors help to eye-ball patterns. Though all calculations are precise and numeric, the method of *interocular trauma* is well-known to work to understand patterns. This just confirms the structure of the vector as described above.

The third pattern is more interesting, since it points out an interaction of parameters, namely of standard deviation of innovations to cost-push shock σ_p into home prices Phillips curve. The standard deviation of this shock is not identified and displays small interaction with standard deviation of technology shock σ_a. Obviously, the main trouble is the σ_p and we can spot traces of σ_a, yet note the positioning of the zero-point. Black-white pattern does not necessarilly mean perfectly opposite relationship.

One could go on with the identification patterns to find out that the problematic parameters are – the indexation param. of wages α_w, the packers labor demand param. ξ_w, the standard deviation of cost-push shock σ_p, the Calvo parameter for wages θ_w, the standard deviation of wage cost-push shock σ_w, the autocorrelation of the cost-push shock to home prices ρ_p, the standard deviation of foreign risk-premium shock $\sigma^\star_\mu$, the labor supply Frisch

Tab. 2: Parameter subset selection – SOE model

order	A	B	order	A	B
1	rho_i	rho_i	*16*	delta_h	std_eps_a
2	theta_h	theta_h	*17*	rho_mpolicy	delta_h
3	rho_mud	std_omega_y	*18*	sigma	std_eps_mpolicy
4	std_omega_y	std_omega_i	*19*	rho_eta_w	rho_eta_w
5	std_omega_i	rho_mud	*20*	std_eps_mud	gamma
6	rho_a	rho_habit	*21*	std_eps_a	std_eps_mud
7	gamma	phi_pie	*22*	delta_f	delta_f
8	std_eps_mpolicy	rho_a	*23*	varphi	varphi
9	phi_y	std_omega_pie	*24*	std_eps_mud_star	rho_eta_p
10	rho_habit	rho_mpolicy	*25*	rho_eta_p	std_eps_mud_star
11	std_omega_pie	rho_mud_star	*26*	std_eps_eta_w	std_eps_eta_w
12	rho_mud_star	eta	*27*	std_eps_eta_p	std_eps_eta_p
13	eta	theta_f	*28*	alpha_w	theta_w
14	phi_pie	sigma	*29*	theta_w	alpha_w
15	theta_f	phi_y	*30*	xi_empl	xi_empl

elasticity φ, a typical weakly identified parameter, etc. On the other hand very well identified parameters, as determined by the structure of the column space, is the autocorrelation of the nominal interest rate in the interest rule, ρ_r or home prices Calvo parameter θ_h both with almost no interactions with other parameters.

Subset selection of parameters To have a clearer view on the selection of parameters, we calculate the ordering of parameters in terms of their strengh of influence and collineraity by use of the algorithm introduced above. The sorting algorithm relied on k successive column-partitioning and column intersections in each step. The results for unscaled FIM and scaling by the absolute size of the parameters produce very similar, though not identical, results. Importantly those hopelessly identified parameters are sorted always right.

We provide the *parameter ordering* in the Tab. 2, where the case A stands for FIM without any scaling and B is the FIM scaled by parameter size. From inspection of the table the method suggests that most reliably estimated parameters should be ρ_i (`rho_i`), Taylor rule smoothing parameter, home country Calvo parameter, (`theta_h`), autoregression parameter for home country interest risk-premium shock (`rho_mud`), for instance. Surprisingly highly positioned is the interest rate rule coefficient of inflation, ϕ_π (`phi_pi`), which in Steinbach et al. (2009) seems not to be identified too well and in general it is difficult to estimate in many models. The parameters denoting variance of measurement errors for output and interest rates $\sigma_{y,i}$ (`std_omega_y,i`) would be also very well identified, though they are not estimated. Next well identified parameters are habit formation, ρ_{habit} (`rho_habit`) and technology shock persistence ρ_a (`rho_a`).