Andrle, Michal

Working Paper

A note on identification patterns in DSGE models

ECB Working Paper, No. 1235

Provided in Cooperation with:
European Central Bank (ECB)

Suggested Citation: Andrle, Michal (2010) : A note on identification patterns in DSGE models, ECB Working Paper, No. 1235, European Central Bank (ECB), Frankfurt a. M.

This Version is available at:
https://hdl.handle.net/10419/153669

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1235 / AUGUST 2010

A NOTE ON IDENTIFICATION PATTERNS IN DSGE MODELS

by Michal Andrle

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1235 / AUGUST 2010

A NOTE ON IDENTIFICATION PATTERNS IN DSGE MODELS[1]

by Michal Andrle[2]

In 2010 all ECB publications feature a motif taken from the €500 banknote.

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=1656963.

1 The main part of the paper was written while I was with European Central Bank, DG-Research, Econometric Modelling Division in 2009. I would like to thank Jan Brůha, Antti Ripatti and Anders Warne for useful comments and discussions. I thank Marco Ratto and participants of the 6th Dynare conference for comments. I am very 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 IMF.

2 Czech National Bank, Monetary and Statistics Dept., Macroeconomic Forecasting Division; email: michal.andrle@cnb.cz

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract

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

Keywords: DSGE, identification, information matrix, rank, singular value decomposition,

J.E.L. Classification: F31, F41

Non-technical summary

The paper discusses identification issues in Dynamic Stochastic General Equilibrium (DSGE) models and suggest a simple and intuitive method to detect and interpret identification problems, based on singular value decomposition.

The use of formal econometric techniques makes sence only if the model is identified, to ensure that different set of structural parameters do not result in observationally equivalent outcomes. Weak or non- identified models make both frequentist and Bayesian estimation and policy analysis more challenging. For unidentified parameters the criterion function used for estimation is non-responsive to changes in these parameters.

The weak or non-identification of a parameter is caused its either by lack of influence of the parameter value on the criterion function (likelihood in our case) or because it affects the criterion function in the same way as some other parameters. In case of DSGE models the identification problem may arise due to non-unique mapping from reduced form parameters of the model to deep structural parameters having economic interpretation.

The paper proposes a particular matrix transformation, singular value decompostion (SVD), as an extremely useful tool for detecting and interpreting local weak and non- identification. SVD may be used to analyze the Hessian of the estimation criterion function, the Fisher information matrix in case of likelihood-based estimation, or the linearized mapping from structural to reduced form parameters. SVD reveals the unidentified parameters as the members of the nullspace of the analyzed mapping. The components of the nullspace are labelled as 'identification patterns'. The notion of the nullspace is explained in the paper.

SVD is suggested for checking the nullspace of the Fisher information matrix or the Jacobian of the mapping from deep structural parameters to reduced ones. The number of non-zero singular values determines the rank of the matrix and the number of identifiable parameters. If a rank-defficiency of the matrix –a non-identification sign– is found, SVD is the answer to the question what is the cause of the reduced rank. Further, by sequentially eliminating most weakly identified parameters it is possible to sort the parameters in terms of their identifiability.

The singular value decomposition is also linked to measures of collinearity and variance inflation factors. It is demonstrated that a single collinear relationship among parameters inflates both bivariate and multiple correlation coefficients based on the Fisher information matrix. The correlation analysis thus may not be fully revealing about the true causes of identification problems.

The identification method analyzed is demonstrated using two well-known DSGE models to explore weakly and non- identified parameters as well as to sort the parameters in terms of their 'identification strength'.

Introduction

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

DSGE models are typically estimated using formal econometric full- or limited information methods instead of less formal parameters calibration. In order to obtain meaningful results, parameters must be identified, to ensure that different set of structural parameters do not result in observationally equivalent outcomes. It is important to remember that identification is relevant in both classical (frequentist) and Bayesian approaches to statistical inference.

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

Our goal is to find non-identified and weakly identified parameterizations, their structures, and suggest plausible restrictions on unidentified parameters. To determine what parameters, or combinations of structural parameters, are causing the problem, we propose inspecting basic subspaces of linear maps which allow us to locate identifiable and non-identifiable subspaces in the parameter space. We suggest that *singular value decomposition* (SVD) is natural choice for locating unidentified parameter subspace and provides also insight into the 'strength' of the identification. Further, we demonstrate that bivariate and multiple correlation measures may provide misleading view about identification and we propose a simple method for detecting best identified parameters based on rank-revealing factorizations.

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

The DSGE model identification problem was under-researched for a while and still seem to be neglected by some economists in the field. The importance of the issue was reminded by Canova and Sala (2006) and Canova and Sala (2009), insightful paper by Cochrane (2007) or investigations by Iskrev (2008) and Iskrev (2009b) with a focus on estimation. Cochrane (2007) and Beyer and Farmer (2007) raise the important issue of observationally equivalent structures and types of equilibria in connection with –often arbitrary– lag length restrictions; a point previously raised by Pesaran (1987) and Sargent (1978). These papers alert economists to the importance of inspecting border-line lag specifications, since absence or presence of sufficient lag length may prevent or deliver identifiability for

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

important structural parameters.

The recent contribution of Komunjer and Ng (2009) provides necessary and sufficient order and rank conditions for local identifiability of DSGE models, so we will not restate the results in this paper. 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 builds on 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 an auto-covariance generating function (ACGF).

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 found interesting works 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 purpose of the paper is to suggest a simple and workable method for identification checks.

The structure of the paper is as follows. The first section 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, suggests an 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, 'identification' of the model is often understood as an estimation of the model, although structural identification is also dealt with. Another important difference is that in engineering both the system's output and input are usually readily available, which is not the case in economics.

1 Identification and Observational Equivalence

1.1 Identification Defined

Identification is related to observational equivalence. 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. 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 an observationally equivalent structure. This paper only deals with local identification of DSGE models, meaning identification near the specific value of θ in the parameter space Θ.

In a brilliant 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, Rothemberg (1971) stated the results for two important cases – (i) the case of an Information matrix without the existence of the *reduced form* structure and (ii) the case of existence of the reduced form, where reduced form parameters $\tau \in \mathcal{T} \subset \mathbb{R}^m$ help to establish the identification of the structural parameters. We assume existence of reduced form parameters and the mapping $T(\theta, \tau) = 0$ since we focus on DSGE models.

In the case where $T(\theta, \tau) = 0$ exists and, importantly, if the 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 We briefly comment on a Bayesian view of the identification. The problem does not seem as clearcut under the Bayesian paradigm as under the Classical one. See Aldrich (2002) for an enlightening review and discussion of 'how likelihood and identification went Bayesian'. The issue may 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. There were requests for broadening the concept of identification for the Bayesians, view that "underidentifiability causes no real difficulty in the Bayesian approach" (Dréze 1972). This was followed by the 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'" (Dréze 1975).

The uninformativeness of 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 marginal prior distribution demonstrates a lack of identification. The likelihood is weakly or non- responsive to changes in the parameter and the prior information dominates regardless of the sample size. On the other hand, with explicit or implicit dependence among parameters data may be marginally informative even for conditionally unidentified parameter. This is way difference of marginal posterior from prior density is not a sufficient sign of identification.

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

We are limiting ourselves to identification in terms of *data informativeness* and thus we explore the properties of likelihood or other criterion functions.

1.2 DSGE Models Case

We write 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 a particular θ so we explore linear map $\tau = T\theta$. Note that the non-uniqueness of the solution of this map is sufficient to cause non-dentification. Its uniqueness is only a necessary condition for identificiation since τ may not be well identified.[3]

[3]Recall that state-space models can be related by similarity transformations, 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). 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).

We take the model under inspection as given including 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 of structural parameters into reduced form parameters.

We will focus on the Information matrix in particular 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).[4] All exercises are done without the use of data and we do not estimate the model. Identification can be evaluated in selected areas of parameter space. However, the number of time periods T is an important as it enters the likelihood function as a parameter and determines what frequencies of data are considered.

We shall thus explore properties the information matrix

$$R(\theta) \equiv \mathbb{E}\left\{\left(\frac{\partial L}{\partial \theta'}\right)'\left(\frac{\partial L}{\partial \theta'}\right)\right\} = \mathbb{E}\left\{\left(\frac{\partial \tau}{\partial \theta'}\right)'\left[\left(\frac{\partial L}{\partial \tau'}\right)'\left(\frac{\partial L}{\partial \tau'}\right)\right]\left(\frac{\partial \tau}{\partial \theta'}\right)\right\} \quad (3)$$

$$= \mathbb{E}\left\{\left(\frac{\partial \tau}{\partial \theta'}\right)' R(\tau)\left(\frac{\partial \tau}{\partial \theta'}\right)\right\}, \quad (4)$$

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.

2 Identification and Linear Maps

The identification issue generally boils down to inspection of a matrix rank. The object of interest may be a 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 $\tau = T\theta$. We need to inspect the rank of the matrix and determine the causes of matrix rank-deficiency. In case of reduced form parameters, the question may also be posed about the invertibility of the linear map (its matrix) T.

Let us 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 we employ standard results from linear algebra.

[4]The likelihood function can be calculated both in time and frequency domain. The frequency domain al to perform identification analysis across frequencies.

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* – nullspace of T, null(T), range of T and nullspace and range of T'. Most importantly,

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

We define the rank of T as rank$(T) = \dim \text{range}(T)$. A non-obvious 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 a linear transform is a subspace of its domain and that it consists of all vectors which are the solution to $T\theta = 0$. Importantly, when the range is of dimension r, the nullspace 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, can we 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 the *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 we require the linear map, determined by matrix T, to be of a full rank. We demonstrate that the inspection of T's subspaces clarifies the structure of the problem. The null space of T is orthogonal to range of T', while the range of A is orthogonal to left-nullspace, null(T'), which has a dimension of $m - r$.

An absence of null(T') implies that there is a solution and an absence of null(T) indicates that the solution is *unique*. The absence of nullspace of T is easy to check using the rank conditions.

When there is no unique solution, meaning the nullspace is not empty, it is crucial to explore the nullspace structure, which determines the subspace of unidentifiable parameters. We may also inspect the row Echelon form to find out free elements in case of singular T. Consequently, we need to find 'suitable' basis for the subspace. The linear map is expressed by a matrix T, yet linear maps are defined with respect to two bases – domain and target. When the bases are not explicitly stated standard basis can be assumed. Operators (maps from a space to itself) require only one basis. Most matrix factorisations consider square matrices only. It is useful to have a more general tool in order to analyze the matrix of a linear mapping. One such useful tool is *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 SVD, we can find a very 'nice' basis for the needed subspaces. We obtain orthonormal basis vectors; the matrix of the linear map is diagonal with respect to both bases. It is important to note that we can find SVD factorisation for any $(m \times n)$ matrix. After applying SVD to real matrices no complex numbers are involved. We are left, however, with two orthogonal bases, which do not cancel mutually. While this is potentially problematic, in our case it is not. 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 related to SVD output.

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. Furthermore 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.[5]

SVD is a standard way of determining a *rank* of the matrix. 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 the number of nonzero singular values. This is why the diagonal form under the new basis is so useful – the inputs from the subspaces associated with the zero singular values are eliminated. 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 $r = 1$. Notably, 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 it is clear whether the identification problem becomes a concern. If the matrix is rank-deficient, i.e. $r < \min\{m, n\}$, then it means that the

[5] 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'.

nullspace of A is not empty. If it is not empty, its structure will indicate the cause of the problem. We know 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 problematic components form a basis for the null space of A. The unidentified parameter combinations can be found in V_2', and as a bonus they are sorted according to degree of distance to singularity and in the form of normalized unit vectors.

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

2.2 Identification Patterns, Nullspace and Restrictions

As we have previously demonstrated, inspecting several subspaces associated with a linear map suggests insights into the roots of (near) singularity. Let us assume we have already decomposed the Information matrix $R(\theta)$ or map T and found it to be of rank r, $r < \min\{m, n\}$ and thus rank-deficient. Now we have matrices $V = [V_1\ V_2]$ and $U = [U_1\ U_2]$ which capture the necessary information about the map. In the case of a square symmetric matrix we have $U = V$.

The rows of V_2' provide an orthonormal basis of the nullspace 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 constitute a mapping from the original parameter space Θ to the lower dimensional parameter space $\mathcal{K}$ of dimension $(n - r)$.

Rank-deficiency implies a need for a restriction of the parameter space of a particular order. Inspecting the nullspace and the range of the map suggests what these restrictions should be in order to achieve local identification. Basically, it is necessary to get rid of linear combinations identified by 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 a $(r \times n)$ 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 of rank r, as required. This rank condition is related to the rank conditions of stacked $[R(\theta);\ \phi(.)]$ as in Rothemberg (1971).

To better illustrate the idea, imagine a very special case of $R(\tau) = 1$ with perfectly identified reduced form parameters, which gives us $R(\theta) = \mathbb{E}\{T'T\}$. Using SVD to analyze $T = U\Sigma V'$ yields

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

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

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 left the $(n \times 1)$ vectors provide the basis for the spaces associated with the associated eigenvalue – zero or numerically close to zero. As indicated, $\|v_k\| = 1$ and $v_i v_j' = 0$ for $i \neq j$. Since for each v_i in the nullspace we know that $Tv_i = 0$, using the 'column view' it is clear that elements of v_i determine linear dependent columns.

Exploring 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 further illustrate the idea, let us demonstrate three interesting cases which can help identification patterns comprehension. Assume for the moment that in θ parameter vector the element θ_1 is redundant from the system, with its value not affecting the criterion function and thus completely unidentified. Furthermore, assume that parameters θ_2 and θ_3 substantially affect the criterion function, but are perfectly positively collinear and enter the system as $(\theta_2 + \theta_3)$. Finally, let θ_4 and θ_5 enter the system as $(\theta_4 - \theta_5)$, while θ_i for $i = \{6, \ldots, n\}$ is well identified and unrelated to $\theta_{1,\ldots,5}$. The nullspace given by V_2 is thereby the following:

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

We have three identification patterns. First, for θ_1, we see that the first column of V_2 suggests that the parameter is strictly unrelated to any other parameters and that it is completely unidentified, since it is fully mapped to the nullspace. The second column of V_2 a implies perfect dependency between the two parameters. An identical increase in one

parameter together with identical decrease in the other does not affect anything, trivially. The third pattern is the reverse.

We therefore need to adopt three restrictions, respecting the structure of V_2, in order to reduce the system identified in the lower-dimensional space $\mathcal{K}$ by three dimensions. We must restrict θ_1 to a single particular value and at least one element from the (θ_2, θ_3), (θ_4, θ_5) pairs. The identification problem is not solved by fixing $\theta_{1,4,5}$, for instance.

Often the problem is to choose a subset of parameters and to select those which are close to being orthogonal with others, so that the conditioning number of the reduced sub-problem is maximized. In sections bellow a sub-set selection method is suggested in order to select or sort the parameters by their strength of identification.

More complex patterns For truly rank-deficient problems the properties of the T's nullspace can be traced into great 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 free parameters are determined by inspecting 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 will demonstrate in the next section. Having a simple program that a produces list of unidentified and weakly identified parameters proved to be useful.

3 Identification Strength

This section examines the case in which all parameters are identified. By identified parameters we understnand also those that may be only weakly identified(in the sense that their individual impact on the criterion function is very small or there is hi collinearity among parameters). SVD 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 the rank of the linear mapping T is determined. 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. Fortunaterly SVD is the right tool for determining 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 associated with a set of singular values considered 'small' by the researcher. Judging what the 'small' means may be often difficult. We will now show how to explore approximate nullspace, how to interpret results from SVD in terms of weak identification and, finally, we suggest a subset selection

approach to sorting parameters by their 'degree of identification'.

3.1 Near collinearity

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

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

SVD makes similar things more similar and dissimilar things dissimilar and so it decorrelates the identification patterns, without using a 'correlation' analysis. As collinearity can be among multiple columns, bivariate relationships or correlations are only necessary, not sufficient signs of rank deficiency or weak identification. Recall also that dim range(T) = dim range(T'), so rows must be considered as well.

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

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

that is – the linear combinations of columns are zero for $\{\gamma_i \neq 0\}_{i=1}^{n}$. Often, we do not have perfect collinearity, but only strong collinearity. The point is that it is useful to know the linear dependencies having them sorted according to their distance from zero.

That is, basically, what SVD delivers. Note that since $T = U\Sigma V'$ we have for $j \in [1, p]$, $p = \min\{m, n\}$. 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 need to adopt the column-view of the operation again 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 significant patterns of multicollinearity, most γ_k coefficients quickly approach zero. Since we have $\|Tv_i\| = \sigma_i$, we can re-normalize coefficients in the linear form and express this as a norm of the error, 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 the selected vectors is unity in that case.

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

Scaling of the Problem When analyzing the matrix T or $R(\theta)$ using SVD, the scaling of the matrix matters for the identification analysis. Clearly scaling does not affect the detection of perfect linear combinations since these are immune to rescaling. SVD is dependent on scaling, however, and thus rescaling columns is often advisable if parameter units differ a lot.

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

The condition number is invariant when matrix is multiplied by a constant – note that in case of a determinant the opposite is the truth.[6] The condition number is not, however, invariant to the rescaling of individual columns. If one partitions 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 then is what is the most 'natural' scaling for the problem. Our view is that it is best not to scale or to scale by the absolute value of the parameters used in the estimation. The reason is that in when searching for criterion function extreme during estimation a *gradient-based method* relies on the score vector and its relation to SVD of the Hessian. The information embodied in SVD can also be 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 is an issue for non-finite precision calculations. It may not always be sufficient to inspect the structure of V_2, since for any nonsingular matrix A we have $XV_2A = 0$ with an altered zero structure of linear dependency, so the problem should be put into a linear transformation invariant setup, see Belsley and Klema (1974), where inspection of $G \equiv -V_{22}V_{21}^{-1}$ is analyzed.

Conditioning of the Problem It is a standard result in computation to check for the *conditioning* of a matrix, which, simply put, determines the sensitivity of the problem $T\theta = \tau$ to small variations in T and τ. Ill-conditioned problems display extreme sensitivity. The condition number κ designates a distance to singularity and is the ratio of largest to smallest

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

singular values.[7] The ratio of the first and the i-th singular value is called the condition index. The steepness of sorted singular values conveys a lot of information about the decay of identifiability, as is the case with the 'scree plot' in principal components analysis. For singular problems $\kappa \to \infty$, see e.g. Golub and van Loan (1996) or a detailed treatment in Higham (1996), inter alia.

3.2 Parameter subset selection

It is 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?

Let us assume that the nullity of the map T or Fisher Information matrix is one – hence, we need to eliminate at least 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 of the problem – given a particular scaling.

The problem differs for a general $(m \times n)$ mapping T and for a square symmetric Fisher information matrix. In case of the mapping from structural to reduced form parameters it is important to leave independent parameters, leading to a column subset-selection problem, i.e. a permutation matrix P_1 which reorders T as $TP_1 = [T_1 \; T_2]$. In case of a symmetric square matrix an elimination of a parameter is creating a submatrix R_- using a permutation matrix P_2 and obtaining the matrix R_2 with the same condition number

$$R_2 = P_2 R P_2' = \begin{bmatrix} R_- & \times \\ \times & \times \end{bmatrix} = P_2 U \Sigma U' P_2'. \tag{19}$$

To determine a 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. This approach is difficult, unless $k = 1$ which we use in order to sort the parameters, excluding always a parameter whose elimination most improves condition number in each round.

For the general problem of $k \neq 1$ 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.[8]

[7] Some details on implementation are important. First, the conditioning of the problem is norm-dependent, our statements relate to 2-norm. Calculating SVD in `Matlab` by Mathworks, for instance, uses the command `[U S V] = svd(A)`. Further, if you check the rank of the matrix, then SVD is used again and the smallest non-zero singular value is checked with entered tolerance, default tolerance being related to machine ε.

[8] The problem of subset selection is 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.

The procedure is as follows: 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 $k \leq r = \text{rank}(T)$. Calculate rank-revealing QR factorization with column-pivoting[9] $V_1'P = QR$ where V_1' is $(k \times m)$ matrix in $V' = [V_1'\ V_2]'$ and P is the *permutation matrix*. Choose the subset of k components of the parameter set θ as $\widehat{\theta} = P'\theta$.

Note that the subset selection algorithm suggests what parameters to retain and what parameters not to retain, yet these groups are not necessarily sorted. We suggest a simple procedure for parameter sorting so 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 by one dimension and proceed with the selected columns in the same way until the number of columns is reduced to the last one. In the case of symmetric matrices one can eliminate each time one row and one column, corresponding to a particular parameter.

For a more detailed exposition of the column-subset selection algorithm via rank-revealing transformations, see Golub et al. (1976) or Golub and van Loan (1996).

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 with correlation measures

Correlations measures may not always be 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 provides a numerical example and also use 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 angle of these vectors in n dimensional space. As the bivariate correlation $\rho_{i,j}$ gets closer to ± 1, the vectors become more and more collinear.

Clearly, singularity may result from more than just two vectors, forming a linear combination. In this case one may find very low bivariate correlations between vectors, while the

A^+ denotes Moore-Penrose generalized inverse. There are both *deterministic* and *randomized* algorithms for this general difficult problem.

[9] In `Matlab` the command `[q r p] = qr(A)` delivers QR with column pivoting with the permutation matrix P. There are, however, more robust algorithms available.

matrix is effectively singular, which is easy to detected using SVD. An intuitive solution to the problem then seems to be 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 a 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 can be expressed using formula

$$1 - R_i^2 = \frac{|S|}{s_{11}|S_{22}|}, \tag{20}$$

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

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

$$\text{VIF}_k = \sum_{j=1}^{n} \frac{v_{kj}^2}{\sigma_j^2} = \sum_{j=1}^{n} \left(\frac{v_{kj}}{\sigma_j}\right)^2, \tag{21}$$

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 them may be small. Note that for each VIF_k all singular values are used – including those potentially very small in the case of near collinearity. Then, for an r-th element, we have

$$\sigma_r \to 0 \text{ and } v_{kr} \neq 0 \qquad \text{VIF}_k \to \infty \qquad R_k \to 1. \tag{22}$$

A single near-collinear relationship can thus make all variance inflation factors infinite or very large. All multiple correlation coefficients among vectors in the matrix are then close to unity. Further, Belsley et al. (1980) demonstrate that this behavior is shared by all bivariate correlations and thus partial correlation coefficients. Having many multiple correlation coefficients close to unity then does not imply that all parameters of the model are very dependent or weakly identified.

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

$$B = [B_1 \; b_5] = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.20 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.40 \\ 0.00 & 0.00 & 0.00 & 1.00 & -0.70 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}, \tag{23}$$

where the first four columns are formed as $B_1 = I_4 + \mathcal{E}$, where I_4 is identity and $\mathcal{E}$ is a draw from Gaussian distribution with very small 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.

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

$$R = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.24 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.48 \\ 0.00 & 0.00 & 0.00 & 1.00 & -0.84 \\ 0.00 & 0.24 & 0.48 & -0.84 & 1.00 \end{bmatrix}, \; \text{VIF} = 10^{12} \begin{bmatrix} 0.00 \\ 0.40 \\ 1.60 \\ 4.92 \\ 6.93 \end{bmatrix} \quad R^2_{1:5} = \begin{bmatrix} 0.9679 \\ 0.8501 \\ 0.9923 \\ 0.8290 \\ 0.9941 \end{bmatrix}. \tag{24}$$

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 the 39 multiple correlation coefficients in for the Smets and Wouters (2007) are all, except for three, 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).

The procedures shown were also used for other DSGE models, for instance in Andrle, Hlédik, Kameník and Vlček (2007–2008) or Steinbach, Mulhoe and Smit (2009)[10] to identify the strength of identification and identification patterns.[11]

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

[11]The solution of all models was obtained using the IRIS-Toolbox for Matlab by Jaromír Beneš, an objected-oriented toolbox for developing and using DSGE models – `www.iris-toolbox.com`.

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 a smoothing interest rate rule with contemporaneous effect of inflation and output growth. The model features three measured 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 specific coefficient values using the first data generating process (DGP1). Following An and Schorfheide (2006) we define the reduced form parameter κ for the slope of the Phillips curve, since its individual components are not identified and we would trivially obtain a rank-deficient Information matrix with right-singular vectors pointing to these components, associated with zero singular values. Contrary to An and Schorfheide (2006) we do not analyze the intercept parameters γ, r^A, π present in a 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 risk aversion coefficient, 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. To calculate the 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[12], (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 true parametrization. On the other hand, the curvature of $\sigma_R, \sigma_g, \sigma_z$ or ρ_R is reasonable. The authors provide the sensitivity analysis 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 provide 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.

We analyze the Information matrix without any scaling. It is of full rank and identification patterns (i.e. right-singular vectors), see plots in Fig. 1. Since the Information matrix is not scaled, the condition number is large and the profile of condition indices drops sharply

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

after the sixth dimension which 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 the 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; this is not due to collinearity but it is 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 due to both its small impact on the likelihood and its partial confounding with ϕ_1, the interest rate rule weight on inflation as can be viewed from the identification patterns 9 and 10. From an 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 out repeated subset-selection problem. 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

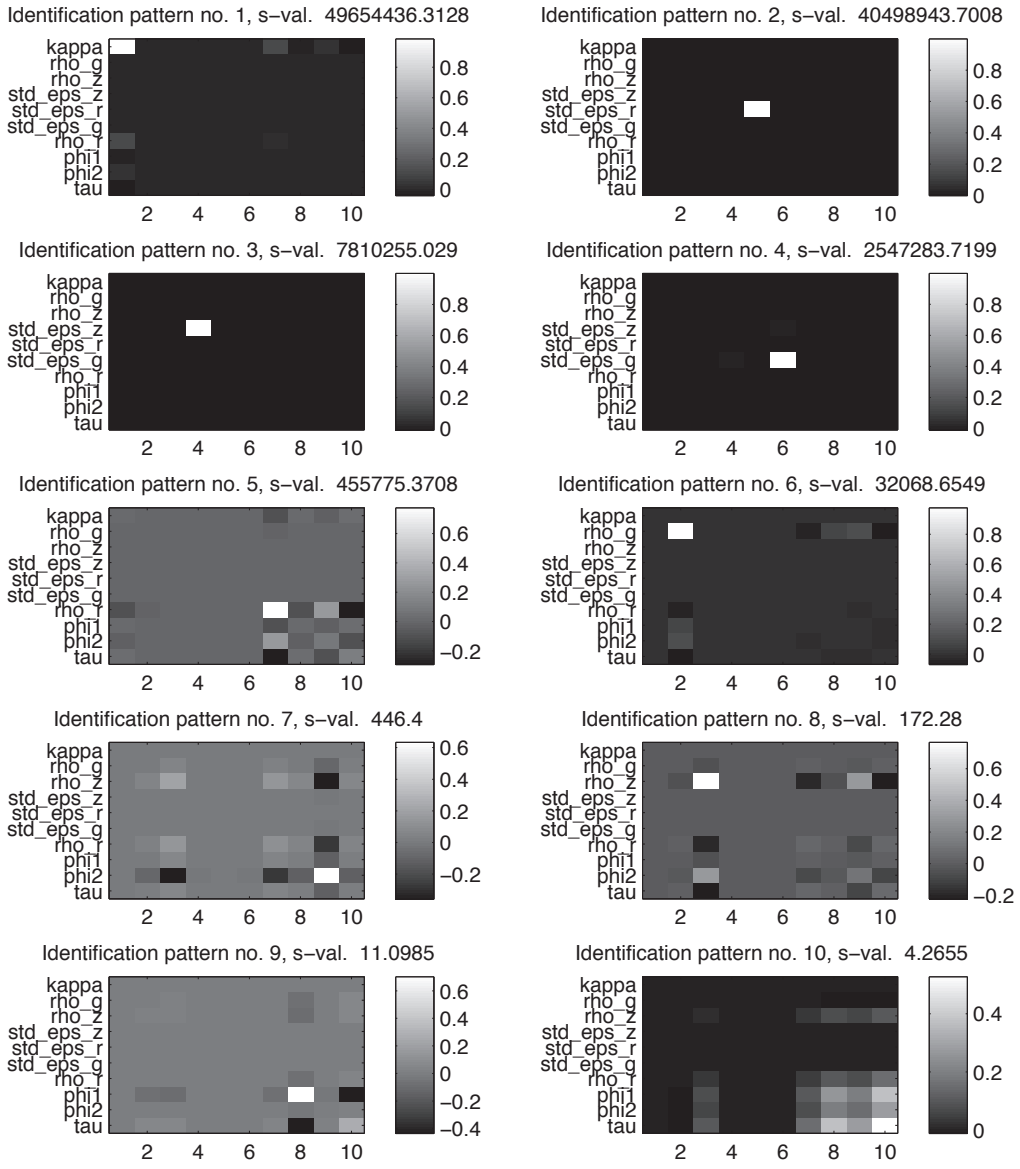
To test the method on the model by Smets and Wouters (2007) we use the model code and prior and posterior mode made public by the authors.

We inspect the Fisher Information matrix using $T = 200$. We also provide a correlation-version of the FIM –after eliminating singularity– and calculate variance inflation factors (VIFs) and corresponding right-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 parameters is larger than for the standard deviation of monetary policy shock.

We analyze only the parameters 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 would otherwise lead to rank-defficiency due to their collinearity with ξ_w and ξ_p, the Calvo parameters.

By inspecting the identification patterns we can see very complex interactions among virtually all parameters. This implies that it would be difficult and insufficient to solely

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		

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 the steady state level of hours worked, inflation and trend growth rate, and others are ρ_g, σ_b, σ_r or ρ, determining the auto-regression of exogenous government spending, variance of preference and monetary policy shock or the interest rate smoothing parameter, to a smaller degree. With few interactions, the first two parameters are poorly identified, whereas the second group belongs to better identified parameters, as indicated by the analysis of singular values.

The identification patterns of most other parameters are rather complex and would require lengthy analysis, which sidestep by applying the subset selection algorithm to both the unscaled and the parameter size scaled Information matrix. The results of this parameter ranking are listed in Table 1 for the case scaled by the relative size of parameters.

It seems that the 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 the best identified parameters, but seem to be identified better than indexation parameters ι_p and ι_w. The parameter affecting intertemporal elasticity of substitution σ_c is also left from the best identified parameters.

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

the analysis.[13]

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

The identification analysis of the model by means of correlation analysis will be severely affected by presence of parameters associated with very small singular values, λ_w, which affect all correlation results as explained above.

[13] 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 a simple and useful method for exploring the 'identification patterns' of DSGE models. The identification patterns are associated with the nullspace of the particular linear map being investigated. The problem of identification or the identification strength is shown 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 both strongly and weakly identified patterns of the parameter space, while also suggesting whether the result is due to lack of influence of the parameter or due to its interactions with other parameters. The method seems very flexible and can be used to carry out a-priori investigations of the model identification before a model has been estimated at all. The local nature of the method can be used for a global identification analysis using simple pseudo- or random simulation schemes or investigating particular border parametrisations. The presented analysis is not dependent on a particular estimation method being used.

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 is an extremely helpful tool. In contrast to an eigenvalue decomposition the singular value decomposition may be applied to any matrix, not necessarily square and symmetric. This is useful for an analysis of a mapping from deep to reduced-form parameters of a model.

We suggested a heuristic method for ordering the parameter vector in terms of individual element's 'identifiability' by carrying out a repeated sub-set selection problem, which takes into account both the 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

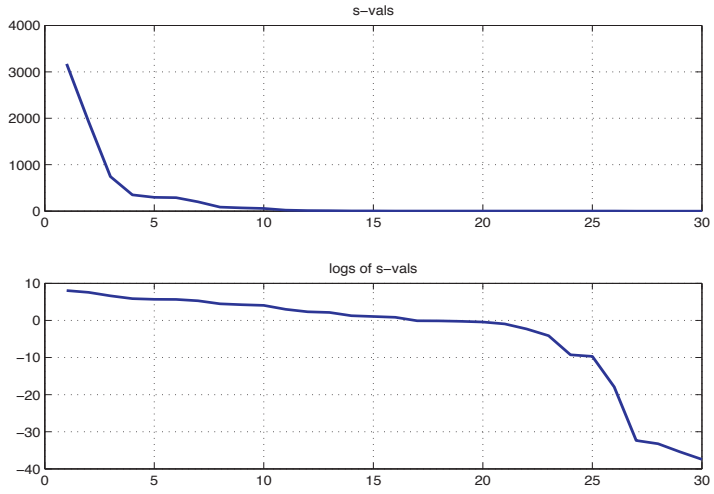
The appendix briefly discusses identification properties of a variant of Monacelli (2003) and Justiniano and Preston (2004) as described also in Steinbach et al. (2009). The modification to Steinbach et al. (2009) is adding possibility of partial price indexation for imported prices $\pi_{f,t}$ and that we simplified the foreign block using a simple VAR specification – consequently, we do not estimate any of these autoregressive parameters or variances.

Likelihood The parameter space consists of 30 structural parameters, including standard deviations. We *do not estimate* the model, rather we carry out an a-priori analysis of the identification. For computing the Information matrix, we sample shocks using the prior-mode parameters of the model, so no real data are used. The assumed sample size is $T = 200$.

The analysis –except of truly linear dependent combinations– is not scale invariant. We have 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.

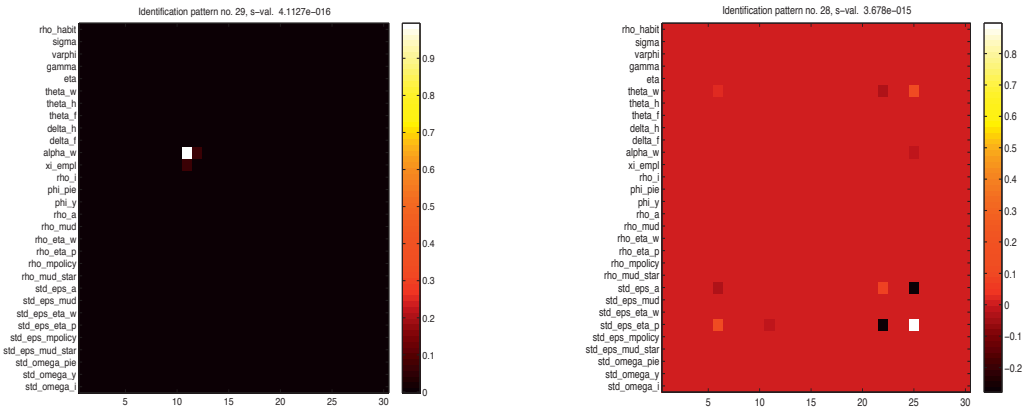
Identification Patterns The Information matrix is analyzed using SVD as suggested in the main text of the paper. There is no true zero – there are no parameter 'floating in the air'. The rank of the matrix –i.e. dimension of the column space– is evaluated as $r = 26$, suggesting at least four problematic 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 suggest a significant portion weak identification resulting from the model structure.

Fig. 2: Singular values – small open economy model



The nullspace has four dimensions. The approximate nullspace associated with 'small' singular values, indicates weakly identified parameters. The first poorly unidentified pattern consist of ξ_w (`xi_empl`), denoting the CES production function elasticity for labor packers that subsequently appears in the wage Phillips curve. The pattern shows it does not interact with anything, nor does it contribute to anything. This should not come as a surprise. The corresponding right-singular vector includes 0.99 at the location of the parameter in θ and zeros or very small 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 the previous one, since the parameter is not interacting with others.

Fig. 3: Identification pattern – SOE model



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

As an example we plot the v associated with the pattern in Fig. 3 to demonstrate the logic of our plots. The white square is associated with α_w, all other coefficients are corresponding to zero. The colors help to eye-ball patterns. The analysis by an *interocular trauma* is known to work well for understanding patterns.

The third pattern is more interesting, since it points out an interaction of parameters, namely of standard deviation of a 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. The main problem is the σ_p and we can spot traces of σ_a, yet note the positioning of the zero-point.

Proceeding with the analysis the results show 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 $\sigma_\mu^\star$, the labor supply Frisch elasticity φ, a typical weakly identified parameter. On the other hand, 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.

To have a clearer view on the selection of parameters, the parameter ordering, in terms of their

strength of influence and collinearity, is calculated by repeated subset-selection problem. The results for unscaled FIM and scaling by the absolute size of the parameters produce similar, though not identical, results. Importantly the least identified parameters are sorted always right.

The parameter ordering is provided in the Tab. 2, for the case of the FIM without any scaling (A) and for the FIM scaled by parameter sizes (B). The most reliably estimated parameters seem to 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). 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 often a parameter difficult to estimate. The parameters denoting variance of measurement errors for output and interest rates $\sigma_{y,i}$ (std_omega_y,i) are well identified, though they are not estimated. Other well identified parameters are habit formation, ρ_{habit} (rho_habit) and technology shock persistence ρ_a (rho_a).