NBP Narodowy Bank Polski

NBP Working Paper No. 170

Global identification of linearized DSGE models

Andrzej Kocięcki, Marcin Kolasa

NBP Working Paper No. 170

Global identification of linearized DSGE models

Andrzej Kocięcki, Marcin Kolasa

Economic Institute
Warsaw, 2013

Andrzej Kocięcki – Narodowy Bank Polski; email: andrzej.kociecki@nbp.pl
Marcin Kolasa – Narodowy Bank Polski, Warsaw School of Economics;
email: marcin.kolasa@nbp

Print:
NBP Printshop

Published by:
Narodowy Bank Polski
Education & Publishing Department
ul. Świętokrzyska 11/21
00-919 Warszawa, Poland
phone +48 22 653 23 35
www.nbp.pl

ISSN 2084-624X

Contents

1 Introduction 5

2 DSGE model 7

3 Equivalent state-space representations 9

4 Identication of deep parameters 12

5 Algorithm checking identication 15

6 Example: An-Schorfheide model 17

6.1 Model summary 17

6.2 Simple case 18

6.3 Full model 18

7 Concluding remarks 21

References 22

Tables and gures 25

Appendices 26

Abstract

This paper introduces a time domain framework to analyze global identification of stochastically nonsingular DSGE models. A formal identification condition is established that relies on the restrictions linking the observationally equivalent minimal state space representations and on the inherent constraints imposed by them on deep model parameters. We next develop an algorithm that checks global identification by searching for observationally equivalent model parametrizations. The algorithm is efficient as the identification conditions it employs shrink considerably the space of candidate deep parameter points and does not require solving the model at each of these points. We also derive two complementary necessary conditions for global identification. Their usefulness and the working of the algorithm are illustrated with an example.

JEL: C13, C51, E32
Keywords: global identification, DSGE models

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have developed into useful tools for macroeconomic analysis. A growing number of policy making institutions, and central banks in particular, use them not only for designing counter-factual experiments, but also for assessing the current stance of the economy and forecasting. The latter application has been supported by growing evidence that estimated medium-sized DSGE models can be competitive with time series models and expert judgment (Del Negro and Schorfheide, 2013).

It has been well understood that DSGE models can suffer from serious identification deficiencies, making estimation problematic at best. One of the first papers to examine this issue was Canova and Sala (2009) who developed simple diagnostic tools for detecting problems with identification in DSGE models estimated by impulse response matching. Their findings allowed them to state in this context that "observational equivalence, partial and weak identification problems are widespread". A more formal analysis, applicable also to likelihood-based methods, is offered by Iskrev (2010). He establishes conditions for local identification based on the rank of the Jacobian matrix that maps the deep parameters of a DSGE model to its implied first and second moments of observable variables. Komunjer and Ng (2011) draw on control theory and spectral analysis. They derive their local identification conditions based on the rank of the appropriately defined Jacobian matrix that uses the restrictions between the observationally equivalent state space systems. Another important theoretical contribution is by Qu and Tkachenko (2012) who establish their rank conditions for local identification using the spectral density matrix that maps from deep model parameters to functions defined in a Banach space.

Local identification is the necessary condition for existence of well-behaved estimators and hence is "to be or not to be" for econometricians. However, one might argue that what really matters for economists is whether there exist another point in the parameter space, possibly distinctly far from the original one, that results in the same autocovariances, impulse responses etc. - a question that relates to the problem of global identification.

In this paper we offer a theoretical analysis of global identification of DSGE models. Similarly to Komunjer and Ng (2011), our framework relies on conditions linking the observationally equivalent minimal state space representations, derived using classic linear system theory. This allows us to state the necessary (order) condition for global identification. We next use the inherent constraints that are imposed by the model solution on the structural parameters to establish the formal and operational condition for their global identification. This result has two applications. First, it allows us to state two useful necessary conditions for a special case of linear restrictions imposed on semi-structural

parameters. Second, we use it to develop an algorithm that checks global identification by searching for observationally equivalent model parametrizations. The algorithm is efficient as the identification conditions it employs shrink considerably the space of candidate deep parameter points and does not require solving the model at each of these points. This contrasts with Qu and Tkachenko (2013) - the only formal attempt to handle global identification known to us to date.

Our approach has several further important features that make it particularly attractive. As already mentioned, it has a global rather than local flavor. Also, the whole analysis is done in the time domain and hence can be considered more transparent than approaches employing spectral methods. Importantly, and unlike the previous literature on local identification (Iskrev, 2010; Komunjer and Ng, 2011; Qu and Tkachenko, 2012), our method does not rely on evaluating the ranks of the matrices that are obtained numerically. Problems that may arise with this approach are acknowledged by Iskrev (2010) and Komunjer and Ng (2011), see also Canova et al. (2013). In the last appendix we provide additional mathematical arguments on why identification based on differential calculus may be condemned to failure. In contrast, our method is based on solving the system of nonlinear equations and hence is less prone to the aforementioned problems.

The rest of this paper is structured as follows. Section two lays out the structure of a typical DSGE model and its solution. Section three discusses the equivalent state space representations. In section four we work out the formal condition for global identification. Section five presents our algorithm for checking global identification. Applications of the derived necessary conditions and algorithm to a widely analyzed DSGE model of An and Schorfheide (2007) are described in section six. Section seven concludes. The paper is supplemented with technical appendices containing the proofs of the derived theorems and propositions, as well as some additional discussion of the related literature.

2 DSGE model

A DSGE model is a system of non-linear equations involving expectations. While solving this type of models using global methods is in principle possible, it can be prohibitively time consuming unless the number of state variables is very small. In consequence, most studies use local approximations of the original models. In particular, likelihood-based estimation that requires calculating the model solution at each optimization step is usually done with linearized models.[1]

Once linearized, most DSGE models can be cast in the following form

$$\Gamma_1(\theta) \begin{bmatrix} s_{t+1} \\ E_t p_{t+1} \end{bmatrix} = \Gamma_0(\theta) \begin{bmatrix} s_t \\ p_t \end{bmatrix} + \Gamma_2(\theta)\varepsilon_{t+1} \tag{1}$$

where s_t is an $n \times 1$ vector of states, p_t is a $q \times 1$ vector of controls, matrices $\Gamma_0(\theta)$, $\Gamma_1(\theta)$ and $\Gamma_2(\theta)$ are explicit functions of deep model parameters collected in an $m \times 1$ vector $\theta \in \Theta \subseteq \mathbb{R}^m$, and $\varepsilon_t \sim i.i.d.\, N(0, \Sigma(\theta))$ is a $k \times 1$ vector of exogenous variables, where $\Sigma(\theta) : (k \times k)$ is assumed to be symmetric positive definite for every $\theta \in \Theta$.

Assuming that system (1) has a unique stable solution,[2] it can be written as

$$s_t = A(\theta)s_{t-1} + B(\theta)\varepsilon_t \tag{2}$$

$$p_t = F(\theta)s_t \tag{3}$$

where $A(\theta) : (n \times n)$, $B(\theta) : (n \times k)$ and $F(\theta) : (q \times n)$ are matrices that implicitly depend on deep model parameters θ.

Let y_t denote an $r \times 1$ vector of observable variables. Then the measurement equations, linking observables to model variables, can be written as

$$y_t = G(\theta) \begin{bmatrix} s_t \\ p_t \end{bmatrix} + J(\theta)\varepsilon_t \tag{4}$$

where $G(\theta) : r \times (n + q)$ and $J(\theta) : (r \times k)$ are matrices that may depend on θ. Note that equation (4) allows for measurement errors, collected with structural shocks in the already defined vector ε_t.

If we decompose $G(\theta)$ into blocks corresponding to states and controls $G(\theta) = [G^s(\theta)\ G^p(\theta)]$, then using the model solution (2) and (3) allows us to rewrite measurement equa-

[1] See Fernández-Villaverde et al. (2006) for a discussion on how second order approximation errors affect the likelihood function.

[2] See Blanchard and Kahn (1980). Other popular solution algorithms include Anderson and Moore (1985), Uhlig (1999), Klein (2000) or King and Watson (2002). Anderson (2008) offers a comparison of their accuracy and efficiency.

tion (4) as

$$y_t = C(\theta)s_{t-1} + D(\theta)\varepsilon_t \tag{5}$$

where matrices $C(\theta) : (r \times n)$ and $D(\theta) : (r \times k)$ are defined as

$$C(\theta) = (G^s(\theta) + G^p(\theta)F(\theta))A(\theta) \tag{6}$$

$$D(\theta) = (G^s(\theta) + G^p(\theta)F(\theta))B(\theta) + J(\theta) \tag{7}$$

Consequently, the law of motion for observable variables y_t has a representation in the state space form given by transition equation (2) and measurement equation (5). For future reference, such a representation will be called the ABCD-representation. The term is not accidental and indicates that we are in the world of the A, B, C and Ds explored by Fernández-Villaverde et al. (2007).

In what follows, we restrict our attention to the square case $r = k$, which means that the system is stochastically nonsingular. Moreover, from now on, we use the generic simplifying notation $X := X(\theta)$, where X is a matrix that explicitly or implicitly depends on θ. Analogously, when referring to other points in the deep parameter space, we write $\bar{X} := X(\bar{\theta})$.

3 Equivalent state-space representations

This section derives the conditions linking the observationally equivalent ABCD-representations. The obtained results will be used in the next section to establish the conditions for global identification of deep parameters in DSGE models.

Let $y = (y_1, \ldots, y_T) \in \mathcal{Y} \subseteq \mathbb{R}^{r \times T}$ denote data underlying a DSGE model and $p(y; s_0, A, B, C, D, \Sigma)$ be a probability density function (pdf) of its ABCD-representation with respect to Lebesgue measure on $\mathcal{Y}$, i.e. $\int_{\mathcal{Y}} p(y; s_0, A, B, C, D, \Sigma)dy = 1$.[3] In fact, we show in Appendix 2 that such a pdf exists if and only if the following holds

Assumption 1. *D is nonsingular for all $\theta \in \Theta$.*

This assumption is explicitly adopted in a similar context e.g. by Fernández-Villaverde et al. (2007) and Dupor and Han (2011). To state the next assumption we need to define the observability matrix $O = [C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C']'$ and the controllability matrix $K = [B \vdots AB \vdots A^2B \vdots \ldots \vdots A^{n-1}B]$.

Assumption 2. *Matrices O and K have full column and row rank, respectively, i.e. $rank(O) = rank(K) = n$.*

Assumption 2 is well known from the linear system theory and holds if and only if the ABCD-representation is written for minimal dimension state vector s_t (which always exists). It is not restrictive in the sense that usually we can rewrite the model so that it holds. In particular, it precludes an ABCD-representation in which there are columns of zeros in matrix A. See Komunjer and Ng (2011) and their supplementary material for more discussion on this point, with explicit examples on how to obey Assumption 2 in specific models. In Appendix 3 we show that in order to check if Assumptions 1 and 2 are satisfied, we just have to solve the model for any (e.g. randomly selected) $\theta \in \Theta$. If these two assumptions are valid for this θ, then in fact they hold for almost all $\theta \in \Theta$.

Our last assumption is

Assumption 3. *The initial state vector s_0 is a known constant.*

Naturally, in a stationary environment (i.e. when the system started long before time 0), Assumption 3 can be strengthened to $s_0 = 0$. In fact, the latter is implicitly done by Komunjer and Ng (2011) since they assume that all information concerning identification is included in the transfer function.

We use the standard definition of identification of a sampling model in terms of a pdf, see e.g. Haavelmo (1944) or Rothenberg (1971):

[3] The initial state appears as an argument of the ABCD-representation since we do not require it to be identically equal to zero.

Definition 1. A sampling model is globally identified at $\phi \in \Phi$ if and only if, for all $\bar{\phi} \in \Phi$, $[p(y;\phi) = p(y;\bar{\phi})$ for all $y \in \mathcal{Y}] \Rightarrow \phi = \bar{\phi}$

Definition 1 makes it explicit that identification is a matter of a well designed model. In contrast, the definition used in Komunjer and Ng (2011) refers to asymptotic properties of a model (see Komunjer and Ng (2011), p. 2000).

Our identification analysis is based on the following theorem

Theorem 1. *Let Assumptions 1 to 3 hold. Then* $p(y; s_0, A, B, C, D, \Sigma) = p(y; s_0, \bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma})$ *for all* $y \in \mathcal{Y}$ *implies: 1)* $\bar{A} = TAT^{-1}$*, 2)* $\bar{B} = TBU$*, 3)* $\bar{C} = CT^{-1}$*, 4)* $\bar{D} = DU$*, 5)* $\bar{\Sigma} = U^{-1}\Sigma U'^{-1}$*, 6)* $s_0 = Ts_0$*, for some nonsingular matrices* $T : (n \times n)$ *and* $U : (k \times k)$*.*

The conclusion of Theorem 1 is essentially the same as that of Proposition 1-S in Komunjer and Ng (2011) (modulo treatment of s_0). However, the hypotheses underlying our results differ. A more detailed discussion of our Theorem 1 in relation to Komunjer and Ng (2011) is available in Appendix 1. In particular, we demonstrate that our assumptions are weaker.

In Appendix 2 we show that $U = LH\bar{L}^{-1}$, where $H : (k \times k)$ is an orthogonal matrix while L and $\bar{L}$ are lower triangular matrices with positive diagonal elements that come from the Choleski decomposition of Σ and $\bar{\Sigma}$, respectively, i.e. $\Sigma = LL'$, $\bar{\Sigma} = \bar{L}\bar{L}'$. In particular, if we assume that shocks ε_t are independent, i.e. $\Sigma = \mathrm{I}_k$, then we have a useful corollary

Corollary 1. *Let Assumptions 1 to 3 hold. Moreover, assume* $\varepsilon_t \sim i.i.d.\, N(0, \mathrm{I}_k)$*. Then* $p(y; s_0, A, B, C, D) = p(y; s_0, \bar{A}, \bar{B}, \bar{C}, \bar{D})$ *for all* $y \in \mathcal{Y}$ *implies: 1)* $\bar{A} = TAT^{-1}$*, 2)* $\bar{B} = TBH$*, 3)* $\bar{C} = CT^{-1}$*, 4)* $\bar{D} = DH$*, 5)* $s_0 = Ts_0$*, for some nonsingular matrix* $T : (n \times n)$ *and orthogonal matrix* $H : (k \times k)$*.*

Remark 1. Note that unless the initial state s_t is assumed equal to zero, it affects identification of the DSGE model by putting restrictions on possible $T's$.[4] In general, and unlike by Assumption 3, we can treat the initial state vector as an unknown constant so that it essentially becomes an additional parameter. This may be natural if one contemplates Bayesian estimation for then a prior on s_0 can be imposed. In such a case, condition 6) from Theorem 1, and 5) from Corollary 1, should be replaced by $\bar{s}_0 = Ts_0$.

What Theorem 1 and Corollary 1 imply is that the ABCD-representation is not identified. We can only hope for identification of some subset of elements in A, B, C, D and Σ. A highly relevant question is therefore: what is the maximal number of functionally

[4]Since Komunjer and Ng (2011) implicitly assume $s_0 = 0$, the conditions 6) in Theorem 1 and 5) in Corollary 1 do not show up in their Proposition 1-S.

independent elements in these matrices that we can (in principle) uniquely retrieve? One can think of this problem as finding the number of functionally independent elements in the reduced form of the ABCD-representation. We have

Proposition 1. *Let Assumptions 1 to 3 hold. The maximal number of functionally independent elements in the ABCD-representation of a linearized DSGE model is* $2nk + \frac{1}{2}k(k+1) + n$ *if* $s_0 \neq 0$*, or* $2nk + \frac{1}{2}k(k+1)$ *if* $s_0 = 0$*. This holds irrespective of whether* $\varepsilon_t \sim i.i.d.\, N(0, \mathrm{I}_k)$ *or* $\varepsilon_t \sim i.i.d.\, N(0, \Sigma)$*.*

This proposition says that since the reduced form contains exactly $2nk + \frac{1}{2}k(k+1) + n$ functionally independent elements (for the case $s_0 \neq 0$), we cannot pin down uniquely more than $2nk + \frac{1}{2}k(k+1) + n$ deep parameters comprising vector θ.[5] This brings us to the first necessary condition for global identification of DSGE models

Proposition 2. *Let Assumptions 1 to 3 hold. The necessary (order) condition for global identification of a linearized DSGE model is that the number of deep parameters is less than or equal to* $2nk + \frac{1}{2}k(k+1) + n$ *if* $s_0 \neq 0$*, or* $2nk + \frac{1}{2}k(k+1)$ *if* $s_0 = 0$ *.*

Note that this order condition for global identification is exactly the necessary condition for local identification given in Komunjer and Ng (2011), Proposition 2-S, which accomodates the square case but implicitly assumes that $s_0 = 0$.

[5] This is provided that the particular point θ at which we check global identifiability belongs to the set of regular points, see Fisher (1966), p. 163.

4 Identification of deep parameters

We are now ready to establish the formal condition for global identification of deep parameters in linearized DSGE models and provide some further necessary conditions. Using the identification condition given in definition 1 and assuming that system (1) has a unique stable solution, we are in a position to state

Definition 2. A linearized DSGE model is globally identified at $\theta \in \Theta$ if and only if, for all $\bar{\theta} \in \Theta$, $[p(y; s_0, A, B, C, D, \Sigma) = p(y; s_0, \bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma})$ for all $y \in \mathcal{Y}] \Rightarrow (\bar{A} = A, \bar{B} = B, \bar{C} = C, \bar{D} = D, \bar{\Sigma} = \Sigma, \bar{\theta} = \theta)$

Our goal is to obtain an equivalent but more operational definition of identification, provided that Assumptions 1 to 3 hold. To this end, let us denote the conclusion from Theorem 1 as $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E}$, where $\mathcal{E} = \{\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 | \bar{A} = TAT^{-1}; \bar{B} = TBU; \bar{C} = CT^{-1}; \bar{D} = DU; \bar{\Sigma} = U^{-1}\Sigma U'^{-1}; s_0 = Ts_0$;nonsingular $T, U\}$. In Appendix 4 we show that Definition 2 can be equivalently stated as

Definition 3. Let Assumptions 1 to 3 hold. Then a linearized DSGE model is globally identified at $\theta \in \Theta$ if and only if, for all $\bar{\theta} \in \Theta$, $[\bar{A} = TAT^{-1}; \bar{B} = TBU; \bar{C} = CT^{-1}; \bar{D} = DU; \bar{\Sigma} = U^{-1}\Sigma U'^{-1}; s_0 = Ts_0] \Rightarrow (T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{\theta} = \theta)$

Unfortunately, Definition 3 is still hardly operational since its application would amount to searching over the whole deep parameter space to check whether $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E}$ holds only for $T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{\theta} = \theta$. Importantly, the model solution would need to be computed for every candidate $\bar{\theta}$. To circumvent this problem, we use the links between the model's semi-structural parameters Γ_0, Γ_1, Γ_2, G, J and its solution given by matrices A, B, C, D, F. Two of such links are already available as equations (6) and (7), which were obtained by plugging the model solution to measurement equation (5). However, they impose constraints only on G and J. To find inherent constraints on Γ_0, Γ_1 and Γ_2, let us rewrite system (1) by partitioning Γ_0 and Γ_1 into blocks corresponding to states and controls

$$\begin{bmatrix} \Gamma_1^s & \Gamma_1^p \end{bmatrix} \begin{bmatrix} s_{t+1} \\ E_t p_{t+1} \end{bmatrix} = \begin{bmatrix} \Gamma_0^s & \Gamma_0^p \end{bmatrix} \begin{bmatrix} s_t \\ p_t \end{bmatrix} + \Gamma_2 \varepsilon_{t+1} \tag{8}$$

Using (2) and (3) yields

$$\Gamma_1^s A s_t + \Gamma_1^s B \varepsilon_{t+1} = (\Gamma_0^s + \Gamma_0^p F - \Gamma_1^p F A) s_t + \Gamma_2 \varepsilon_{t+1} \tag{9}$$

Since equation (9) must hold for all (permissible) values of s_t and ε_{t+1}, it follows that

$$\Gamma_1^s A = \Gamma_0^s + \Gamma_0^p F - \Gamma_1^p F A \tag{10}$$

$$\Gamma_1^s B = \Gamma_2 \tag{11}$$

Naturally, equations (10), (11), and also (6) and (7) must hold for any $\bar{\theta} \in \Theta$. Let us collect these constraints in set $\mathcal{Z}$, defined as

$$\begin{aligned} \mathcal{Z} \;=\; & \{\bar{A},\bar{B},\bar{C},\bar{D} | \bar{\Gamma}_1^s \bar{A} = \bar{\Gamma}_0^s + \bar{\Gamma}_0^p \bar{F} - \bar{\Gamma}_1^p \bar{F}\bar{A}; \bar{\Gamma}_1^s \bar{B} = \bar{\Gamma}_2; \bar{C} = (\bar{G}^s + \bar{G}^p \bar{F})\bar{A}; \\ & \bar{D} = (\bar{G}^s + \bar{G}^p \bar{F})\bar{B} + \bar{J}\} \end{aligned} \tag{12}$$

One can think of $\mathcal{Z}$ as a counterpart of intrinsic constraints imposed on structural parameters by reduced form parameters in classic simultaneous equations models, even though one might argue that this analogy is too far-fetched since the ABCD-representation is not the reduced form of the underlying DSGE model. It is important to notice that $\mathcal{Z}$ is a tautology. This means that $\bar{A},\bar{B},\bar{C},\bar{D},\bar{\Sigma},s_0 \in \mathcal{E} \Leftrightarrow (\bar{A},\bar{B},\bar{C},\bar{D},\bar{\Sigma},s_0 \in \mathcal{E}) \wedge (\bar{A},\bar{B},\bar{C},\bar{D} \in \mathcal{Z})$. Hence, definition 3 may be equivalently stated as

Definition 4. Let Assumptions 1 to 3 hold. Then a linearized DSGE model is globally identified at $\theta \in \Theta$ if and only if, for all $\bar{\theta} \in \Theta$, $[\bar{\Gamma}_1^s TAT^{-1} = \bar{\Gamma}_0^s + \bar{\Gamma}_0^p \bar{F} - \bar{\Gamma}_1^p \bar{F} TAT^{-1}; \bar{\Gamma}_1^s TBU = \bar{\Gamma}_2; C = (\bar{G}^s + \bar{G}^p \bar{F})TA; DU = (\bar{G}^s + \bar{G}^p \bar{F})TBU + \bar{J}; \bar{\Sigma} = U^{-1}\Sigma U'^{-1}; s_0 = Ts_0] \Rightarrow (T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{\theta} = \theta)$

Definition 4 will prove essential for designing an efficient algorithm to check global identification of a DSGE model. The key ingredient of definition 4 is that it involves semi-structural parameters Γ_0, Γ_1, Γ_2, G and J . This fact is important since in many applications those parameters will be subject to the known linear restrictions. This opens up a possibility to give an analytical identification condition for semi-structural parameters, which of course will become a necessary identification condition for deep parameters. In fact, we state a relatively easily applicable necessary condition for global identification, which can be considered as a useful complementary tool to the order condition given in Proposition 2.

To this end, let us assume that all restrictions imposed on Γ_0, Γ_1, Γ_2, G and J are linear, which may be formally stated as $\Upsilon vec([\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]') = d$ and $\Psi vec([G \vdots J]') = h$. In this notation, $\Upsilon : (r_\Gamma \times (n+q)(2n+2q+k))$ is a known matrix, $d : (r_\Gamma \times 1)$ is a known vector and r_Γ is the number of independent restrictions imposed on Γ_0, Γ_1 and Γ_2. Similarly, $\Psi : (r_{GJ} \times r(n+q+k))$ is a known matrix, $h : (r_{GJ} \times 1)$ is a known vector and r_{GJ} is the number of independent restrictions imposed on G and J. Note that in this setup we allow for cross-restrictions between rows within $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$ and within $[G \vdots J]$.

Proposition 3. *Assume that all restrictions imposed on Γ_0, Γ_1, Γ_2, G and J are linear.*

Define matrices

$$Q = \begin{bmatrix} \mathrm{I}_{n+q} \otimes \begin{bmatrix} A' & A'F' & -\mathrm{I}_n & -F' & 0_{n\times k} \\ B' & 0_{k\times q} & 0_{k\times n} & 0_{k\times q} & -\mathrm{I}_k \end{bmatrix} \\ \Upsilon \end{bmatrix} \text{ and } R = \begin{bmatrix} \mathrm{I}_r \otimes \begin{bmatrix} A' & A'F' & 0_{n\times k} \\ B' & B'F' & \mathrm{I}_k \end{bmatrix} \\ \Psi \end{bmatrix}.$$

If a linearized DSGE model is globally identified at $\theta \in \Theta$, then Q and R are of full column rank (equivalently $Q'Q$ and $R'R$ are nonsingular). The latter implies that $r_\Gamma \geq (n+q)(n+2q)$ and $r_{GJ} \geq rq$.

If there are no restrictions involving parameters from different rows in $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$, Proposition 3 can be specialized. To this end, denote by $\Gamma_{(i)}$ the ith row of $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$. Let all linear restrictions imposed on the ith row of $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$ be written as $\Upsilon_i \Gamma'_{(i)} = d_i$, where $\Upsilon_i : (r_i \times (2n + 2q + k))$ is a known matrix, $d_i : (r_i \times 1)$ is a known vector and r_i is the number of independent restrictions imposed on $\Gamma_{(i)}$. The restrictions imposed on $[G \vdots J]$ are given as before. Then we have a useful corollary

Corollary 2. *Assume that there are no restrictions between rows of $[\Gamma_1 \vdots \Gamma_0 \vdots \Gamma_2]$ and all restrictions imposed on $[\Gamma_1 \vdots \Gamma_0 \vdots \Gamma_2]$ and $[G \vdots J]$ are linear. Define matrices*

$$Q_i = \begin{bmatrix} A' & A'F' & -\mathrm{I}_n & -F' & 0_{n\times k} \\ B' & 0_{k\times q} & 0_{k\times n} & 0_{k\times q} & -\mathrm{I}_k \\ & & \Upsilon_i & & \end{bmatrix} \text{ and } R = \begin{bmatrix} \mathrm{I}_r \otimes \begin{bmatrix} A' & A'F' & 0_{n\times k} \\ B' & B'F' & \mathrm{I}_k \end{bmatrix} \\ \Psi \end{bmatrix}.$$

If a linearized DSGE model is globally identified at $\theta \in \Theta$, then, for every $i = 1, ..., n+q$, Q_i and R are of full column rank (equivalently $Q_i'Q_i$ and $R'R$ are nonsingular). The latter implies that $r_i \geq n + 2q$ for all i, and $r_{GJ} \geq rq$.

Remark 2. The results in this section were derived for $\varepsilon_t \sim i.i.d.\, N(0, \Sigma)$. However, it is straightforward to obtain the analogous results when $\varepsilon_t \sim i.i.d.\, N(0, \mathrm{I}_k)$. In particular, Proposition 3 and Corollary 2 are still valid.

5 Algorithm checking identification

Having necessary conditions for global identification (such as Proposition 2 or 3) is useful. However, it would also be desirable to have an analytical sufficient condition for global identification. Unfortunately, we were unable to produce it. What we offer instead is the insight that immediately follows from our considerations presented in the previous section, which imply that a DSGE model is globally identified at $\theta \in \Theta$ if and only if the system of equations collected in the square bracket in Definition 4 has a unique solution $T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{\theta} = \theta$. It follows that if we can find a solution to the system

$$\bar{\Gamma}_1^s T A T^{-1} = \bar{\Gamma}_0^s + \bar{\Gamma}_0^p \bar{F} - \bar{\Gamma}_1^p \bar{F} T A T^{-1} \tag{13}$$

$$\bar{\Gamma}_1^s T B U = \bar{\Gamma}_2 \tag{14}$$

$$C = (\bar{G}^s + \bar{G}^p \bar{F}) T A \tag{15}$$

$$DU = (\bar{G}^s + \bar{G}^p \bar{F}) T B U + \bar{J} \tag{16}$$

$$\bar{\Sigma} = U^{-1} \Sigma U'^{-1} \tag{17}$$

$$s_0 = T s_0 \tag{18}$$

such that $\bar{\theta} \neq \theta$, the DSGE model is not globally identified at θ.

Note that, for given θ, matrices A, B, C, D and Σ are known while matrices $\bar{\Gamma}_0$, $\bar{\Gamma}_1$, $\bar{\Gamma}_2$, $\bar{G}$, $\bar{J}$ and $\bar{\Sigma}$ are analytically linked to $\bar{\theta}$. Hence, using some θ of interest, our algorithm will search for $\bar{\theta}$, as well as for some nonsingular T and U, and some $\bar{F}$, that solve the system of equations (13) to (18). The efficiency of our procedure when applied even to large DSGE models follows from two facts. First, unlike Qu and Tkachenko (2013), we do not have to solve the model for each candidate deep parameter point to check whether a model is globally identified. Instead, using intrinsic constraints (12) we automatically connect deep parameters with a model solution through the system of nonlinear equations (13) to (18). Second, by using the minimal ABCD-representation and Theorem 1, we reduce the number of additional "unknowns" from A, B, C and D in (12) to T and U in the ultimate system of nonlinear equations (13)-(18).

For a typical DSGE model, the system given by equations equations (13) to (18) is nonlinear in $\bar{\theta}$ and hence numerical methods have to be applied to solve them. Using the fact that, by construction, the system is satisfied for $\bar{\theta} = \theta$, $\bar{F} = F$, $T = \mathrm{I}_n$ and $U = \mathrm{I}_k$, we generate the starting values for the numerical algorithm by randomizing around this point.[6] Naturally, the more diffuse distribution we use in the randomization, the longer

[6] We use Matlab routine `fsolve`. If our model conforms to Corollary 1 rather than to Theorem 1, we

it takes to obtain a solution, but the more likely it is to find alternative parameter values that are far from a given θ.

Several remarks are in order. In a linearized model, stochastic properties of shocks collected in ε_t affect only the variance-covariance matrix Σ and do not show up in the model equations. Hence, while searching for observationally equivalent model parametrizations, we can abstract from the subset of $\bar{\theta}$ that define $\bar{\Sigma}$, and recover them afterwards using (17). Moreover, if we follow Komunjer and Ng (2011) and assume $s_0 = 0$, then we can also drop condition (18) as it does not impose any restrictions on the solution. Finally, it is instructive to briefly discuss two special model setups, in both of which we additionally assume that A is nonsingular.[7] The first case highlights the role of measurement errors. It is easy to verify that, if all shocks in our DSGE model are structural, and additionally matrix B is full row rank,[8] then equations (15) and (16) become identical, simplifying to

$$(G^s + G^p F) = (\bar{G}^s + \bar{G}^p \bar{F})T \tag{19}$$

The second special case concerns the choice of observable variables. If these are chosen such that $\bar{G}^p$ is nonsingular for any $\bar{\theta} \in \Theta$,[9] then (15) can be used to solve for $\bar{F}$ and hence the numerical algorithm needs to search only for $\bar{\theta}$, T and U .

suggest to parametrize an orthogonal H by $H = 2(\mathrm{I}_k + X)^{-1} - \mathrm{I}_k$, where $X : (k \times k)$ is a skew symmetric matrix, i.e. it satisfies $X + X' = 0$.

[7] This assumption seems to be "generic" or "typical" if the state space is minimal. It fact, using arguments from Appendix 3, if A is nonsingular for any $\theta \in \Theta$, then A will be nonsingular for almost all θ.

[8] A necessary condition for that is $n \leq k$, i.e. the number of states cannot be larger than the number of shocks (and observables in the square case).

[9] This condition is satisfied e.g. for the canonical New Keynesian model of Clarida et al. (1999), where the observable variables are output, inflation and the interest rate.

6 Example: An-Schorfheide model

We use the findings presented above to analyze the model in An and Schorfheide (2007). Identification of this model was also examined locally by Komunjer and Ng (2011) so it provides a natural object to highlight the main features of our approach. We first use a simplified version of the model to demonstrate usefulness of the necessary conditions for global identification given by Propositions 2 and 3, and next apply our algorithm to the full version.

6.1 Model summary

When written in log-linearized form, the model is given by the following equations

$$x_t = E_t x_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}(R_t - E_t \pi_{t+1} - E_t z_{t+1})$$

$$\pi_t = \beta E_t \pi_{t+1} + \frac{\tau(1-\nu)}{\nu \bar{\pi}^2 \phi}(x_t - g_t)$$

$$R_t = \rho_m R_{t-1} + (1-\rho_m)[\psi_1 \pi_t + \psi_2 (x_t - g_t)] + \sigma_m \varepsilon_{m,t}$$

$$z_t = \rho_z z_{t-1} + \sigma_z \varepsilon_{z,t}$$

$$g_t = \rho_g g_{t-1} + \sigma_g \varepsilon_{g,t}$$

There are five endogenous variables in the model: detrended output x_t, inflation π_t, interest rate R_t, productivity z_t and government spending g_t. They are driven by three mutually uncorrelated white noise shocks to productivity growth $\varepsilon_{z,t}$, government purchases $\varepsilon_{g,t}$ and monetary policy $\varepsilon_{m,t}$. The 13-dimensional vector of parameters is $\theta = [\ \tau \quad \nu \quad \psi_1 \quad \psi_2 \quad \bar{\pi} \quad \phi \quad \beta \quad \rho_z \quad \rho_g \quad \rho_m \quad \sigma_z \quad \sigma_g \quad \sigma_m\ ]'$. For further use, let us define the following two parameter groups: $\theta_{Phi} = [\ \nu \quad \bar{\pi} \quad \phi\ ]$, collecting those appearing only in the slope of the Phillips curve, and $\theta_{Tay} = [\ \psi_1 \quad \psi_2 \quad \rho_m \quad \sigma_m\ ]$, consisting of parameters showing up only in the Taylor rule. The baseline parameter values used by Komunjer and Ng (2011) are reproduced in Table 1 for convenience. We will use them as our benchmark θ.

It is easy to notice that three of the model parameters, namely ν, $\bar{\pi}$ and ϕ, cannot be separately identified since all of them show up only in the last term of the Phillips curve. Hence, while checking identification we always fix any two in θ_{Phi} at their benchmark values, which leaves us with at most eleven free parameters.

6.2 Simple case

We start with a simple case, in which government spending and productivity growth are white noise and the Taylor rule does not allow for interest rate smoothing, i.e. $\rho_z = \rho_g = \rho_m = 0$.[10] Note that, with these assumptions, shocks to productivity growth do not show up in the model so we can accommodate only two observables, which we chose to be x_t and π_t so that $y_t = [\ x_t \quad \pi_t\ ]'$. Reducing the model to just these two endogenous variables yields the following two-equation system

$$y_t = E_t y_{t+1} + \varepsilon_{g,t} - \frac{1}{\tau}(\psi_1 \pi_t + \psi_2 (y_t - \varepsilon_{g,t}) + \varepsilon_{m,t} - E_t \pi_{t+1})$$

$$\pi_t = \beta E_t \pi_{t+1} + \frac{\tau(1-\nu)}{\nu \bar{\pi}^2 \phi}(y_t - \varepsilon_{g,t})$$

and the matrices associated with (1) for a vector of controls $p_t = [\ y_t \quad \pi_t\ ]'$ are

$$\Gamma_0 = \begin{bmatrix} 1 + \frac{\psi_2}{\tau} & \frac{\psi_1}{\tau} \\ -\frac{\tau(1-\nu)}{\nu\bar{\pi}^2\phi} & 1 \end{bmatrix} \qquad \Gamma_1 = \begin{bmatrix} 1 & \frac{1}{\tau} \\ 0 & \beta \end{bmatrix} \qquad \Gamma_2 = \begin{bmatrix} -(1 + \frac{\psi_2}{\tau})\sigma_g & \frac{1}{\tau}\sigma_m \\ \frac{\tau(1-\nu)}{\nu\bar{\pi}^2\phi}\sigma_g & 0 \end{bmatrix}$$

Since in this model version there are no state variables ($n = 0$) and only two shocks ($k = 2$), the maximum number of parameters that can be identified according to the order condition established in Proposition 2 is only three. However, we can use Corollary 2 to figure out that leaving only three free parameters might not be enough to ensure identification. To see it, let us fix all parameters but ψ_1, ψ_2 and σ_m at their baseline values. This results in three linear restrictions on the first row of $[\Gamma_1 \vdots \Gamma_0 \vdots \Gamma_2]$,[11] while at least four (the number of controls $q = 2$) are needed for identification, so the model is still not identified.

6.3 Full model

We showed using a simple example how our necessary conditions can be easily used to immediately detect identification problems. We now move to the fully-fledged model variant. In this case, these necessary conditions are not powerful enough to give a negative verdict on identification, hence we use our algorithm described in section 5.

If we define an auxiliary variable $\tilde{R}_t = \rho_m R_{t-1} + \sigma_m \varepsilon_{m,t}$, the An and Schorfheide (2007) model can be cast in the form given by (1), with states $s_t = [\ z_t \quad g_t \quad \tilde{R}_t\ ]'$, controls $p_t = [\ y_t \quad \pi_t \quad R_t\ ]'$, exogenous shocks $\varepsilon_t = [\ \varepsilon_{z,t} \quad \varepsilon_{g,t} \quad \varepsilon_{m,t}\ ]'$ and matrices Γ_0,

[10] A similar setup has recently been analyzed by Le et al. (2013).

[11] These restrictions are: $\Gamma_0^{11}\sigma_g + \Gamma_2^{11} = 0$, $\Gamma_1^{11} = 1$ and $\Gamma_1^{12} = \frac{1}{\tau}$, where Γ_i^{lk} is the lth row, kth column element of Γ_i.

Γ_1 and Γ_2 defined as

$$\Gamma_0 = \begin{bmatrix} \rho_z & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho_g & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_m \\ -\frac{\rho_z}{\tau} & -(1-\rho_g) & 0 & 1 & 0 & \frac{1}{\tau} \\ 0 & \frac{\tau(1-\nu)}{\nu\bar{\pi}^2\phi} & 0 & -\frac{\tau(1-\nu)}{\nu\bar{\pi}^2\phi} & 1 & 0 \\ 0 & -(1-\rho_m)\psi_2 & 1 & (1-\rho_m)\psi_2 & (1-\rho_m)\psi_1 & -1 \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{\tau} & 0 \\ 0 & 0 & 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \Gamma_2 = \begin{bmatrix} \sigma_z & 0 & 0 \\ 0 & \sigma_g & 0 \\ 0 & 0 & \sigma_m \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The vector of observable variables is now $y_t = [\ x_t \quad \pi_t \quad R_t\ ]'$, which means that $G = [\ 0_{3\times 3} \quad \mathrm{I}_3\ ]$ for all θ.

We know from the earlier literature that this setup is not identified, even if we fix two of the three parameters in θ_{Phi}. As discussed by Komunjer and Ng (2011), there are further identification issues concerning the coefficients in the Taylor rule. The results of applying our algorithm that we document in Table 2 confirm these findings. For instance, reparametrizing the model such that $\psi_1 = 2$, $\psi_2 = -0.67$, $\rho_m = 0.688$ and $\sigma_m = 0.0018$ results in exactly the same likelihood (and hence impulse responses and moments) for observables as under our benchmark values for these parameters, holding the remaining ones fixed. Consistently with Komunjer and Ng (2011), our algorithm cannot find any observationally equivalent state space representation if we additionally fix either of the following three Taylor rule parameters $[\ \psi_1 \quad \psi_2 \quad \rho_m\ ]$. However, and in contrast to Komunjer and Ng (2011), we find that fixing the standard deviation of monetary shocks σ_m does the trick as well.

We supplement the identification analysis of the An and Schorfheide (2007) model with an examination of the effect of replacing interest rate smoothing by autoregression in the monetary policy shock (assumed white noise in the baseline). More specifically, the monetary policy rule is now given by

$$R_t = \psi_1\pi_t + \psi_2(x_t - g_t) + m_t$$

where

$$m_t = \rho_m m_{t-1} + \sigma_m \varepsilon_{m,t}$$

As can be seen from Table 3, fixing ρ_m at a non-zero level is now not enough to ensure identification of the remaining Taylor rule parameters. On the other hand, restricting any other parameter in θ_{Tay} is sufficient for identification of the rest in this set, including ρ_m.

The following two considerations highlight the attraction of our algorithm compared to the previous literature. First, Komunjer and Ng (2011) show that if they reduce the degree of inertia in shocks ε_z and ε_g to 0.1 and 0.15, respectively, their method fails to confirm identification of the model, even if any of the Taylor rule parameters is restricted. More generally, their method is sensitive to the tolerance level they set while performing the matrix rank tests. Our algorithm does not suffer from such numerical problems. In particular, if we consider low (but non-zero) shock inertia, our identification analysis leads to the same conclusions as for the benchmark parameter set.

The second remark is related to the global flavor of our algorithm. The qualitative results presented above can be obtained with a minimum degree of randomization of the initial values in our algorithm. To search through more distant areas of the parameter space, this parametrization of the algorithm needs to be increased, which can be easily done. However, doing so does not lead to a discovery of any observationally equivalent set of parameters if two of the θ_{Phi} subset and one of the θ_{Tay} subset are fixed. This suggests that the restrictions discussed above are sufficient globally - a result that Komunjer and Ng (2011) could not establish with their local approach. Importantly, unless we restrict $\bar{\pi}$ to be positive, our algorithm does find an observationally equivalent representation of the model with restricted ν, ϕ and either of θ_{Tay}, which is obtained for $\bar{\pi} = -1.008$, i.e. the negative of $\bar{\pi}$ in our original parametrization. While this result is trivial and not economically meaningful, it builds our confidence in the algorithm as having a potential to go beyond a local analysis.

7 Concluding remarks

In this paper we developed a time domain framework for analyzing global identification of DSGE models. In particular, we produced an operational condition for global identification and showed how to apply it to design the algorithm that efficiently searches for observationally equivalent deep parameter sets. We also demonstrated how the algorithm can be used to analyze global identification of a standard small-scale DSGE model that is widely used in the literature. To balance our contribution in relation to Komunjer and Ng (2011), we should mention that the price of using the time domain approach is that we have to confine ourselves to stochastically nonsingular models.

Our framework can be easily applied to more sophisticated models. We did (though not reported) it for the medium-sized small open economy model described in Justiniano and Preston (2010). The reason why this setup might be of interest is related to some identification problems encountered by these authors while they estimated the model. However, applying our algorithm to this example did not result in finding parameter values that would be observationally equivalent to the baseline parametrization used by Justiniano and Preston (2010). This is consistent with Iskrev (2010) or Komunjer and Ng (2011), who show that identification problems in richer models, like the canonical Smets and Wouters (2007) setup, are less widespread than one might expect given the evidence for small models. Our mathematical insight into these problems presented in Appendix 7 may be some form of explanation. It follows that some of the identification problems plaguing estimation of medium-sized DSGE models are of a weak rather than strong nature, e.g. they might result from low curvature of the likelihood function for a given dataset rather than from intrinsic features of the model structure.

References

Adachi, Masahisa (1993) *Embeddings and immersions* (American Mathematical Society, Providence, Rhode Island)

An, Sungbae, and Frank Schorfheide (2007) 'Bayesian analysis of DSGE models.' *Econometric Reviews* 26(2-4), 113–172

Anderson, Gary (2008) 'Solving linear rational expectations models: A horse race.' *Computational Economics* 31(2), 95–113

Anderson, Gary, and George Moore (1985) 'A linear algebraic procedure for solving linear perfect foresight models.' *Economics Letters* 17(3), 247–252

Blanchard, Olivier J., Jean-Paul L'Huillier, and Guido Lorenzoni (2012) 'News, noise, and fluctuations: An empirical exploration.' Development Research Working Paper Series 09/2012, Institute for Advanced Development Studies, September

Blanchard, Olivier Jean, and Charles M. Kahn (1980) 'The solution of linear difference models under rational expectations.' *Econometrica* 48(5), 1305–11

Canova, Fabio, and Luca Sala (2009) 'Back to square one: Identification issues in DSGE models.' *Journal of Monetary Economics* 56(4), 431–449

Canova, Fabio, Filippo Ferroni, and Christian Matthes (2013) 'Choosing the variables to estimate singular DSGE models.' CEPR Discussion Papers 9381, CEPR, March

Clarida, Richard, Mark Gertler, and Jordi Gali (1999) 'The science of monetary policy: A New Keynesian perspective.' *Journal of Economic Literature* 37(4), 1661–1707

Del Negro, Marco, and Frank Schorfheide (2013) 'DSGE model-based forecasting.' In *Handbook of Economic Forecasting,* ed. G. Elliott and A. Timmermann, vol. 2 of *Handbook of Economic Forecasting* (Elsevier) pp. 57–10

Dupor, Bill, and Jing Han (2011) 'Handling non-invertibility: theory and applications.' working paper, Ohio State University

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, and Manuel S. Santos (2006) 'Convergence properties of the likelihood of computed dynamic models.' *Econometrica* 74(1), 93–119

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson (2007) 'ABCs (and Ds) of understanding VARs.' *American Economic Review* 97(3), 1021–1026

Fisher, Franklin M. (1966) *The identification problem in econometrics* (McGraw-Hill, New York)

Goodwin, Graham C., and Robert L. Payne (1977) *Dynamic system identification: Experiment design and data analysis* (Academic Press, New York)

Guidorzi, Roberto (1975) 'Canonical structures in the identification of multivariable systems.' *Automatica* 11, 361–374

Haavelmo, Trygve (1944) 'The probability approach in econometrics.' *Econometrica* 12, 1–115

Hansen, Lars Peter, and Thomas J. Sargent (1991) 'Two difficulties in interpreting vector autoregressions.' In *Rational Expectations Econometrics,* ed. L. P. Hansen and T. J. Sargent (Westview Press, Boulder) pp. 77–119

Hirsch, Morris W. (1976) *Differential topology* (Springer-Verlag, New York)

Hunt, Brian R., Tim Sauer, and James A. Yorke (1992) 'Prevalence: A translation-invariant "almost every" on infinite-dimensional spaces.' *Bulletin of the American Mathematical Society* 27(2), 217–238

Iskrev, Nikolay (2010) 'Local identification in DSGE models.' *Journal of Monetary Economics* 57(2), 189–202

Justiniano, Alejandro, and Bruce Preston (2010) 'Can structural small open-economy models account for the influence of foreign disturbances?' *Journal of International Economics* 81(1), 61–74

King, Robert G, and Mark W Watson (2002) 'System reduction and solution algorithms for singular linear difference systems under rational expectations.' *Computational Economics* 20(1-2), 57–86

Klein, Paul (2000) 'Using the generalized Schur form to solve a multivariate linear rational expectations model.' *Journal of Economic Dynamics and Control* 24(10), 1405–1423

Komunjer, Ivana, and Serena Ng (2011) 'Dynamic identification of dynamic stochastic general equilibrium models.' *Econometrica* 79(6), 1995–2032

Kurmann, André, and Christopher Otrok (2013) 'News shocks and the term structure of interest rates: Lessons for DSGE models .' working paper, Federal Reserve Board, March

Le, Vo Phuong Mai, Patrick Minford, and Michael R. Wickens (2013) 'A Monte Carlo procedure for checking identification in DSGE models.' CEPR Discussion Papers 9411, CEPR, March

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang (2013) 'Fiscal foresight and information flows.' *Econometrica* 81(3), 1115–1145

Olver, Peter J. (1993) *Applications of Lie groups to differential equations* (Springer-Verlag, New York)

Qu, Zhongjun, and Denis Tkachenko (2012) 'Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models.' *Quantitative Economics* 3(1), 95–132

_ (2013) 'Local and global parameter identification in DSGE models allowing for indeterminacy.' working paper, Boston University, May

Rothenberg, Thomas J. (1971) 'Identification in parametric models.' *Econometrica* 39(3), 577–591

Schmitt-Grohé, Stephanie, and Martín Uribe (2012) 'What's news in business cycles.' *Econometrica* 80(6), 2733–2764

Sims, Eric R. (2012) 'News, Non-Invertibility, and Structural VARs.' In *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments (Advances in Econometrics, Volume 28),* ed. N. Balke, F. Canova, F. Milani, and M. A. Wynne (Emerald Group Publishing Limited) pp. 81–135

Smets, Frank, and Rafael Wouters (2007) 'Shocks and frictions in US business cycles: A Bayesian DSGE approach.' *American Economic Review* 97(3), 586–606

Sternberg, Shlomo (1964) *Lectures on differential geometry* (Prentice-Hall, Inc., New Jersey)

Uhlig, Harald (1999) 'A toolkit for analyzing nonlinear dynamic stochastic models easily.' In *Computational Methods for the Study of Dynamic Economies,* ed. R. Marimon and A. Scott (Oxford University Press) pp. 30–61

Tables and figures

Table 1: Baseline parametrization of An-Schorfheide model

Parameter	Value	Parameter	Value
τ	2.0	ρ_z	0.9
ν	0.1	ρ_g	0.95
ψ_1	1.5	ρ_m	0.75
ψ_2	0.125	σ_z	0.003
$\bar{\pi}$	1.008	σ_g	0.006
ϕ	53.68	σ_m	0.002
β	0.9975		

Table 2: Identification of An-Schorfheide model - baseline

Restricted parameters	Non-identified parameters
none	θ_{Phi}; θ_{Tay}
any one in θ_{Phi}	other two in θ_{Phi}; θ_{Tay}
any two in θ_{Phi}	θ_{Tay}
any two in θ_{Phi}; ψ_1	none
any two in θ_{Phi}; ψ_2	none
any two in θ_{Phi}; ρ_m	none
any two in θ_{Phi}; σ_m	none

Note: $\theta_{Phi} \equiv [\ \nu \quad \bar{\pi} \quad \phi\ ]$ and $\theta_{Tay} \equiv [\ \psi_1 \quad \psi_2 \quad \rho_m \quad \sigma_m\ ]$.

Table 3: Identification of An-Schorfheide model - alternative Taylor rule

Restricted parameters	Non-identified parameters
A. Baseline	
any two in θ_{Phi}; any one in θ_{Tay}	none
B. Alternative	
any two in θ_{Phi}; ψ_1	none
any two in θ_{Phi}; ψ_2	none
any two in θ_{Phi}; ρ_m	ψ_1; ψ_2; σ_m
any two in θ_{Phi}; σ_m	none

Note: $\theta_{Phi} \equiv [\ \nu \quad \bar{\pi} \quad \phi\ ]$ and $\theta_{Tay} \equiv [\ \psi_1 \quad \psi_2 \quad \rho_m \quad \sigma_m\ ]$.

Appendices

Appendix 1: Discussion of Theorem 1 in relation to Komunjer and Ng (2011)

Although Komunjer and Ng (2011), KN henceforth, work in the frequency domain, we are able to compare our assumptions to prove Theorem 1 with those adopted by them. First of all, in contrast to KN, we do not need to assume that A is a stable matrix for all $\theta \in \Theta$ (Assumption 2 in KN). That is, our ABCD-representation does not have to imply a stationary moving average solution. The second difference is not so transparent and to fully appreciate it we have to define the concept of transfer function

$$H(z) = D + C(z\mathrm{I}_n - A)^{-1}B \tag{20}$$

where $z \in \mathbb{C}$ (a set of complex numbers). KN assume that the transfer function is left-invertible, which holds if and only if $|z| > 1 \Rightarrow rank(H(z)) = k$ (i.e. transfer function has full column rank). Our counterpart assumption is that D is nonsingular. The natural question is about the relationship between these two assumptions. To answer it, the following result is important[12]

Proposition. *If the transfer function is left-invertible, then $rank(D) = k$, i.e. D is of full column rank.*

Proof. Assume that $k \leq r$. The transfer function is said to be left-invertible if and only if $|z| > 1 \Rightarrow rank(H(z)) = k$. This condition is equivalent to $|z| > 1 \Rightarrow rank(D + C(\mathrm{I}_n - Az^{-1})^{-1}Bz^{-1}) = k$ if and only if $|s| < 1 \Rightarrow rank(D + C(\mathrm{I}_n - As)^{-1}Bs) = k$ (where $z^{-1} = s$), if and only if $|s| < 1 \Rightarrow det((D + C(\mathrm{I}_n - As)^{-1}Bs)'(D + C(\mathrm{I}_n - As)^{-1}Bs)) \neq 0$, if and only if $det((D + C(\mathrm{I}_n - As)^{-1}Bs)'(D + C(\mathrm{I}_n - As)^{-1}Bs)) = 0 \Rightarrow |s| \geq 1$. Using the last equivalent definition, we prove contrapositive of the proposition, i.e. $rank(D) < k$ implies $det((D + C(\mathrm{I}_n - As)^{-1}Bs)'(D + C(\mathrm{I}_n - As)^{-1}Bs)) = 0$ and $|s| < 1$. To this end note that $rank(D) < k$ if and only if $det(D'D) = 0$. However, $0 = det(D'D) = det((D + C(\mathrm{I}_n - A \cdot 0)^{-1}B \cdot 0)'(D + C(\mathrm{I}_n - A \cdot 0)^{-1}B \cdot 0)) = 0$. Hence there exists $s = 0$ (with $|s| < 1$) such that $det((D + C(\mathrm{I}_n - As)^{-1}Bs)'(D + C(\mathrm{I}_n - As)^{-1}Bs)) = 0$. This proves the proposition. □

Since in the square case left-invertibility reads $|z| > 1 \Rightarrow det(H(z)) \neq 0$, we immediately have

Corollary. *Assume $k = r$. If the transfer function is left-invertible then D is nonsingular.*

[12] Though our approach is strictly confined to the square case, the following proposition is proved for $k \leq r$.

There are two perspectives from which we can comment on the above proposition and corollary. The first one is mathematical and the second one methodological. As for the former, our contribution in comparison to KN is that we prove an analogous theorem under weaker conditions, which is always welcomed. Moreover, we use the time domain framework that does not have a clause "with T infinitely large". The other (methodological) perspective is related to non-invertible and/or non-fundamental solutions to (log)linearized DSGE models.[13] Although, in principle, the KN framework accommodates identification problems in the non-invertible case, it does so only in the context of the innovations representation of the original ABCD-representation. In our framework we judge global identification in a model as it stands (which may be non-invertible or even have a divergent moving average component), whereas using the approach of KN one must first take a stand on whether the (stationary) solution is invertible or not, choose the appropriate representation and then check (local) identifiability. However, the main point from our perspective is that the innovations representation's parameters are highly nonlinear functions of A, B, C, D and Σ. Hence, the straightforward relationship with Γ_0, Γ_1 and Γ_2 from model (1) breaks down, which makes the problem of global identification more difficult to tackle, at least in our approach.

Appendix 2: Proof of Theorem 1

Using equations (2) and (5), the whole data sampling process may be compactly written as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix} s_0 + \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{T-2}B & CA^{T-3}B & CA^{T-4}B & \cdots & D \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

Since $\varepsilon_t \sim i.i.d.\, N(0, \Sigma)$, where Σ is positive definite, we can decompose $\Sigma = LL'$, where L is lower triangular with positive diagonal elements. Then we can rewrite the system above as

[13] There is a rapidly growing literature on the problem of invertibility and/or fundamentalness of solutions to DSGE models. This theoretical problem is not new and was subject of debate in 90's, see e.g. Hansen and Sargent (1991). However, the recent "news shocks" literature confirms resurgence of interests in this problem, see e.g. Blanchard et al. (2012), Kurmann and Otrok (2013), Sims (2012), Schmitt-Grohé and Uribe (2012), Leeper et al. (2013).

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix} s_0 + \begin{bmatrix} DL & 0 & 0 & \cdots & 0 \\ CBL & DL & 0 & \cdots & 0 \\ CABL & CBL & DL & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{T-2}BL & CA^{T-3}BL & CA^{T-4}BL & \cdots & DL \end{bmatrix} \begin{bmatrix} \varepsilon_1^\star \\ \varepsilon_2^\star \\ \varepsilon_3^\star \\ \vdots \\ \varepsilon_T^\star \end{bmatrix}$$

where $\varepsilon_t^\star \sim i.i.d.\, N(0, \mathrm{I}_k)$.

Let us define $y = (y_1', y_2', \ldots, y_T')'$ and

$$\mu = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix} s_0; \quad R = \begin{bmatrix} DL & 0 & 0 & \cdots & 0 \\ CBL & DL & 0 & \cdots & 0 \\ CABL & CBL & DL & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{T-2}BL & CA^{T-3}BL & CA^{T-4}BL & \cdots & DL \end{bmatrix}$$

It follows that $y \sim N(\mu, \Omega) := p(y; \mu, \Omega)$, where $\Omega = RR'$. We have the following sequence of equivalences: D is nonsingular (Assumption 1), if and only if DL is nonsingular (since L is the Choleski "square root", which is nonsingular by assumption that all diagonal elements in L are positive), if and only if R is nonsingular, if and only if Ω is positive definite. But note that $p(y; \mu, \Omega) = p(y; \bar{\mu}, \bar{\Omega})$ for all $y \in \mathcal{Y}$ implies $\Omega = \bar{\Omega}$ (see e.g. Theorem 4 in Rothenberg (1971)). On the other hand, $\Omega = RR' = \bar{R}\bar{R}' = \bar{\Omega}$ if and only if $\bar{R} = RV$, where V is orthogonal (by Vinograd's theorem). Since R is nonsigular, we have $R^{-1}\bar{R} = V$. But since $R^{-1}\bar{R}$ is block lower triangular, it follows that V is block diagonal and each diagonal block is itself an orthogonal matrix. Denote the first diagonal block of V as $H : (k \times k)$, with $(HH' = H'H = \mathrm{I}_k)$. Since $\bar{R} = RV$, then writing the first k columns in this relationship we have

$$\begin{bmatrix} \bar{D}\bar{L} \\ \bar{C}\bar{B}\bar{L} \\ \bar{C}\bar{A}\bar{B}\bar{L} \\ \vdots \\ \bar{C}\bar{A}^{T-2}\bar{B}\bar{L} \end{bmatrix} = \begin{bmatrix} DLH \\ CBLH \\ CABLH \\ \vdots \\ CA^{T-2}BLH \end{bmatrix} \Leftrightarrow \begin{bmatrix} \bar{D} \\ \bar{C}\bar{B} \\ \bar{C}\bar{A}\bar{B} \\ \vdots \\ \bar{C}\bar{A}^{T-2}\bar{B} \end{bmatrix} = \begin{bmatrix} DU \\ CBU \\ CABU \\ \vdots \\ CA^{T-2}BU \end{bmatrix}$$

where $U = LH\bar{L}^{-1}$. Using a part of the above relations we get

$$\begin{bmatrix} \bar{C} \\ \bar{C}\bar{A} \\ \vdots \\ \bar{C}\bar{A}^{n-1} \end{bmatrix} \bar{B} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} BU$$

or equivalently $\bar{O}\bar{B} = OBU$. Using Assumption 2 implies $\bar{B} = \bar{O}^{-}OBU$, where $\bar{O}^{-}$ denotes any left inverse of $\bar{O}$. Further, it is easy to see that

$$\bar{C}\bar{K} = [\bar{C}\bar{B}\vdots\bar{C}\bar{A}\bar{B}\vdots\bar{C}\bar{A}^{2}\bar{B}\vdots\ldots\vdots\bar{C}\bar{A}^{n-1}\bar{B}] = CK(\mathrm{I}_n \otimes U)$$

which implies by Assumption 2 that $\bar{C} = CK(\mathrm{I}_n \otimes U)\bar{K}^{-}$, where $\bar{K}^{-}$ is any right inverse of $\bar{K}$. On the other hand it is evident that

$$\bar{O}\bar{K} = OK(\mathrm{I}_n \otimes U)$$

which implies by Assumption 2 that $\bar{O} = OK(\mathrm{I}_n \otimes U)\bar{K}^{-}$ and then $\mathrm{I}_n = \bar{O}^{-}OK(\mathrm{I}_n \otimes U)\bar{K}^{-}$. Since $\mathrm{I}_n \otimes U$ is nonsingular and K has full row rank, it follows that $K(\mathrm{I}_n \otimes U)\bar{K}^{-}$ is nonsingular and hence $\bar{O}^{-}O = (K(\mathrm{I}_n \otimes U)\bar{K}^{-})^{-1}$. Note that this implies that $\bar{O}^{-}O$ is nonsingular, too. Moreover, it is easy to realize that

$$\bar{O}\bar{A}\bar{K} = OAK(\mathrm{I}_n \otimes U)$$

which implies by Assumption 2 that $\bar{A} = \bar{O}^{-}OAK(\mathrm{I}_n \otimes U)\bar{K}^{-}$. In addition, $\bar{D}\bar{L} = DLH \Rightarrow \bar{D}\bar{L}\bar{L}'\bar{D}' = DLL'D' \Rightarrow \bar{D}\bar{\Sigma}\bar{D}' = D\Sigma D' \Rightarrow \bar{\Sigma} = \bar{D}^{-1}D\Sigma(\bar{D}^{-1}D)'$ (by Assumption 1). But $\bar{D}^{-1}D = U^{-1}$, thus $\bar{\Sigma} = U^{-1}\Sigma U'^{-1}$. Since $p(y;\mu,\Omega) = p(y;\bar{\mu},\bar{\Omega})$ for all $y \in \mathcal{Y}$ also implies $\mu = \bar{\mu}$, we have $Os_0 = \bar{O}s_0 \Rightarrow \bar{O}^{-}Os_0 = s_0$. Ultimately, putting all these results together, we obtain $\bar{D} = DU$, $\bar{B} = \bar{O}^{-}OBU$, $\bar{C} = CK(\mathrm{I}_n \otimes U)\bar{K}^{-}$, $\bar{A} = \bar{O}^{-}OAK(\mathrm{I}_n \otimes U)\bar{K}^{-}$, $\bar{\Sigma} = U^{-1}\Sigma U'^{-1}$ and $\bar{O}^{-}Os_0 = s_0$. Denoting $T = \bar{O}^{-}O$ and bearing in mind that $\bar{O}^{-}O = (K(\mathrm{I}_n \otimes U)\bar{K}^{-})^{-1}$, the theorem follows.

Appendix 3: Proof of Proposition 1

If M is a manifold, let $dim(M)$ denote the number of local coordinates in the Euclidean space required to describe the points in M. Denote the set $\mathcal{S} = \{A, B, C, D, \Sigma, s_0 \in \mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n} |$ Assumptions 1 and 2 hold$\} = \mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_3$, where $\mathcal{S}_1 = \{A, B, C, D, \Sigma, s_0 \in \mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n} | \, rank(O) = n\} = \{A, B, C, D, \Sigma \in \mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n} | \, det(O'O) \neq 0\}$, $\mathcal{S}_2 = \{A, B, C, D, \Sigma \in \mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n} | \, rank(K) = n\} = \{A, B, C, D, \Sigma \in \mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n} | \, det(KK') \neq 0\}$ and $\mathcal{S}_3 = \{A, B, C, D, \Sigma \in \mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n} | \, det(D) \neq 0\}$. Evidently, each $\mathcal{S}_i$ for $i = 1, 2, 3$ is an open subset of $\mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n}$

since it is the inverse image of the open set $\mathbb{R}\setminus\{0\}$. As a finite intersection of open subsets is open, we conclude that $\mathcal{S}$ is the manifold and $dim(\mathcal{S}) = n^2 + 2nk + k^2 + \frac{1}{2}k(k+1) + n$. In fact, since the determinant is a polynomial that is an analytic function of its elements, it implies that $\mathcal{S}$ is dense in $\mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n}$ (for analytic functions such as the determinant, the latter cannot be equal to 0 on an open subset of $\mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n}$, unless it is identically equal to zero). Hence, to check if Assumptions 1 and 2 are valid we have to solve the model for any $\theta \in \Theta$ (which may be randomly selected or a common calibration). If Assumptions 1 and 2 hold at any $\theta \in \Theta$, then, since $\mathcal{S}$ is an open and dense subset of $\mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n}$, we conclude that Assumptions 1 and 2 hold for almost all $\theta \in \Theta$. Equivalently, all $\theta \in \Theta$ such that Assumptions 1 and 2 are violated form a nowhere dense subset of $\mathbb{R}^{n^2+2nk+k^2+\frac{1}{2}k(k+1)+n}$ of measure zero.

Let G_1 and G_2 be groups. Let us define the direct product of groups as their Cartesian product $G = G_1 \times G_2$ with the group composition $g\bar{g} := (g_1\bar{g}_1,\ g_2\bar{g}_2)$, for any $g = (g_1, g_2), \bar{g} = (\bar{g}_1, \bar{g}_2) \in G$. Such defined G is also a group. As long as G_1 and G_2 are manifolds, we get $dim(G) = dim(G_1) + dim(G_2)$. In fact, in our case G_1 and G_2 will be manifolds being open subsets of the Euclidean space or its submanifold.

To proceed further, we need a few basic notions from group theory. Let X be any set. Consider the mapping $G \times X \to X$ that sends (g, x) into $g \circ x$, where "$\circ$" is a binary operation. We say that a group G acts on X if: 1) $e \circ x = x$ for all $x \in X$, where "e" denotes the identity element in group G; 2) $\bar{g} \circ (g \circ x) = (\bar{g}g) \circ x$ for all $g, \bar{g} \in G$ and $x \in X$. Note that X may be the Cartesian product i.e. $X = X_1 \times X_2 \times \ldots \times X_n$. If this is the case, the action of G on X will be defined as $g \circ (x_1, x_2, \ldots, x_n) := (g \circ_1 x_1,\ g \circ_2 x_2,\ ...,\ g \circ_n x_n)$. Note that a binary operation may be distinct for every X_i. Since the case when X is the Cartesian product is what we need in our proof, we confine ourselves to this case. For any given $x_1, \ldots, x_n \in X_1 \times \ldots \times X_n$, let us define $Stab_{x_1,\ldots,x_n} = \{g \in G | (x_1, x_2, \ldots, x_n) = g \circ (x_1, x_2, \ldots, x_n)\}$ and call it the stabilizer of $x_1, \ldots, x_n$. We say that G acts freely on X if $Stab_{x_1,\ldots,x_n} = \{e\}$ for all $x_1, \ldots, x_n \in X_1 \times \ldots \times X_n$ (recall that e denotes the identity element in G). Lastly, define the orbit of $x_1, \ldots, x_n$ as $Orb_{x_1,\ldots,x_n} = \{g \circ (x_1, x_2, \ldots, x_n) | g \in G\} \equiv \{g \circ_1 x_1,\ g \circ_2 x_2,\ \ldots,\ g \circ_n x_n |\ g \in G\}$.

Now we are in a position to prove the proposition. First, consider the case $\varepsilon_t \sim i.i.d.\, N(0, \Sigma)$. Let us define the set $\mathcal{E} = \{TAT^{-1},\ TBU',\ CT^{-1},\ DU',\ U'^{-1}\Sigma U^{-1},\ Ts_0 |\ T \in GL_n, U \in GL_k\}$,[14] where GL_q denotes the general linear group, i.e. $GL_q = \{g \in \mathbb{R}^{q\times q} |\ det(g) \neq 0\}$. In fact, it is easy to show that $\mathcal{E}$ is the orbit of A, B, C, D, Σ, s_0 for we can write $\mathcal{E} = Orb_{A,B,C,D,\Sigma,s_0} = \{(T, U) \circ (A, B, C, D, \Sigma, s_0) |\ (T, U) \in GL_n \times GL_k\}$. Indeed, to prove that the direct product of groups $(GL_n \times GL_k)$ acts on $\mathcal{S}$ we have to check

[14] With some abuse to definition of $\mathcal{E}$ introduced in section 4, we replaced U by U'. This is a matter of convention and has no consequences.

two conditions that define the group action. Since the identity element in the direct product of groups $GL_n \times GL_k$ is $(\mathrm{I}_n, \mathrm{I}_k)$, putting $T = \mathrm{I}_n$ and $U = \mathrm{I}_k$ in $\mathcal{E}$ we easily get $(\mathrm{I}_n, \mathrm{I}_k) \circ (A, B, C, D, \Sigma, s_0) = (A, B, C, D, \Sigma, s_0)$. To check the second defining property of the group action, we have to show that $(\bar{T}, \bar{U}) \circ ((T, U) \circ (A, B, C, D, \Sigma, s_0)) = ((\bar{T}, \bar{U})(T, U)) \circ (A, B, C, D, \Sigma, s_0) := (\bar{T}T, \bar{U}U) \circ (A, B, C, D, \Sigma, s_0)$, where the last equality follows form the rule of group composition in the group direct product. We have $(\bar{T}, \bar{U}) \circ ((T, U) \circ (A, B, C, D, \Sigma, s_0)) = (\bar{T}, \bar{U}) \circ (TAT^{-1}, TBU', CT^{-1}, DU', U'^{-1}\Sigma U^{-1}, Ts_0) = (\bar{T}TAT^{-1}\bar{T}^{-1}, \bar{T}TBU'\bar{U}', CT^{-1}\bar{T}^{-1}, DU'\bar{U}', \bar{U}'^{-1}U'^{-1}\Sigma U^{-1}\bar{U}^{-1}, \bar{T}Ts_0) = (\bar{T}TA(\bar{T}T)^{-1}, \bar{T}TB(\bar{U}U)', C(\bar{T}T)^{-1}, D(\bar{U}U)', (\bar{U}U)'^{-1}\Sigma(\bar{U}U)^{-1}, \bar{T}Ts_0) = (\bar{T}T, \bar{U}U) \circ (A, B, C, D, \Sigma, s_0)$. Hence $(GL_n \times GL_k)$ acts on $\mathcal{S}$ and $\mathcal{E} = Orb_{A,B,C,D,\Sigma,s_0}$.

Now we demonstrate that $Stab_{A,B,C,D,\Sigma,s_0} = \{(T, U) \in GL_n \times GL_k |\ A = TAT^{-1}, B = TBU', C = CT^{-1}, D = DU', \Sigma = U'^{-1}\Sigma U^{-1}, s_0 = Ts_0\} = \{\mathrm{I}_n, \mathrm{I}_k\}$, the identity element in $GL_n \times GL_k$. If $A = TAT^{-1}$ and $C = CT^{-1}$ then $[C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C']' = [C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C']'T^{-1}$. But by Assumption 2, $[C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C']'$ possesses a left inverse, hence it follows that $T^{-1} = \mathrm{I}_n \Leftrightarrow T = \mathrm{I}_n$. Moreover, by Assumption 1, $D = DU' \Leftrightarrow U' = D^{-1}D = \mathrm{I}_k = U$. Hence, under our assumptions, $GL_n \times GL_k$ acts freely on $\mathcal{S}$.

In order to proceed further we need to impose some regularity condition

Assumption. *(Regularity condition, RC): Each point $A, B, C, D, \Sigma, s_0 \in \mathcal{S}$ has arbitrarily small neighborhood V whose intersection with each orbit Orb_{A,B,C,D,Σ,s_0} is a pathwise connected subset $V \cap Orb_{A,B,C,D,\Sigma,s_0}$ of the orbit (a subset is pathwise connected if any two points can be joined by a smooth curve).*

Next, let us introduce the following

Definition. Let a group G act on $X_1 \times \ldots \times X_n$. A real valued function $I : X_1 \times \ldots \times X_n \to \mathbb{R}$ is called invariant if for all $x_1, \ldots, x_n \in X_1 \times \ldots \times X_n$ and all $g \in G$ we have $I(g \circ (x_1, x_2, \ldots, x_n)) = I(x_1, x_2, \ldots, x_n)$.

Under the free action and when the RC holds, each orbit is a submanifold of $\mathcal{S}$ (see e.g. Olver, 1993, p. 22) that is completely characterized by its (functionally independent) invariants. In fact, the total number of these invariants (so-called the complete set of functionally independent invariants) gives the dimension of each orbit (being a manifold), see e.g. Olver, 1993, p. 210-214. In particular, by Theorem 2.17 in Olver (1993), at each $A, B, C, D, \Sigma, s_0 \in \mathcal{S}$ there are precisely $dim(\mathcal{S}) - dim(GL_n \times GL_k) = n^2 + 2nk + k^2 + \frac{1}{2}k(k+1) + n - (n^2 + k^2) = 2nk + \frac{1}{2}k(k+1) + n$ functionally independent invariants in a neighborhood of A, B, C, D, Σ, s_0 (we used the fact that since GL_q may be identified with an open subset of $\mathbb{R}^{q^2}$, both GL_n and GL_k are manifolds and $dim(GL_n \times GL_k) = dim(GL_n)$

$+ dim(GL_k) = n^2 + k^2$). Hence $dim(Orb_{A,B,C,D,\Sigma,s_0}) = dim(\mathcal{E}) = 2nk + \frac{1}{2}k(k+1) + n$, for each $A, B, C, D, \Sigma, s_0 \in \mathcal{S}$.

Now assume that $\varepsilon_t \sim i.i.d.\, N(0, \mathrm{I}_k)$. Since the covariance is the identity matrix, we work with the set $\mathcal{S}_H = \{A, B, C, D, s_0 \in \mathbb{R}^{n^2+2nk+k^2+n} |$ Assumptions 1 and 2 hold$\}$. Evidently, $\mathcal{S}_H$ is the manifold with $dim(\mathcal{S}_H) = n^2 + 2nk + k^2 + n$. Define the set $\mathcal{E}_H = \{TAT^{-1}, TBH', CT^{-1}, DH', Ts_0 |\ T \in GL_n, H \in O_k\}$, where O_k denotes the orthogonal group, i.e. $O_k = \{g \in \mathbb{R}^{k\times k} |\ g'g = gg' = \mathrm{I}_k\}$. Using the same arguments as above, it is easy to show that $\mathcal{E}_H$ is the orbit of A, B, C, D, s_0 for we can write $\mathcal{E}_H = Orb_{A,B,C,D,s_0} = \{(T, H) \circ (A, B, C, D, s_0) |\ (T, H) \in GL_n \times O_k\}$, where $GL_n \times O_k$ denotes the direct product of two groups (with component-wise group composition as before). Evidently, $Stab_{A,B,C,D,s_0} = \{\mathrm{I}_n, \mathrm{I}_k\}$, the identity element in $GL_n \times O_k$ (the proof is literally the same replacing H with U). Hence, under the RC, using the same theory from Olver (1993) we have: $dim(Orb_{A,B,C,D,s_0}) = dim(\mathcal{E}_H) = dim(\mathcal{S}_H) - dim(GL_n \times O_k) = dim(\mathcal{S}_H) - dim(GL_n) - dim(O_k) = n^2 + 2nk + k^2 + n - n^2 - \frac{1}{2}k(k-1) = 2nk + \frac{1}{2}k(k+1) + n$, for each $A, B, C, D, s_0 \in \mathcal{S}_H$, where we used the fact that O_k is the manifold of dimension $\frac{1}{2}k(k-1)$.

Appendix 4: Equivalence between Definitions 2 and 3

Note that Theorem 1 may be trivially strengthened to $p(y; s_0, A, B, C, D, \Sigma) = p(y; s_0, \bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}); \forall y \in \mathcal{Y} \Leftrightarrow \bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E}$ (see section 4 for definition of $\mathcal{E}$). Then definition 2 may be rewritten as $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow (\bar{A} = A, \bar{B} = B, \bar{C} = C, \bar{D} = D, \bar{\Sigma} = \Sigma, \bar{\theta} = \theta)$. But this is equivalent to $[\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow (\bar{A} = A, \bar{B} = B, \bar{C} = C, \bar{D} = D, \bar{\Sigma} = \Sigma)] \wedge [\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow \bar{\theta} = \theta]$. To state an alternative definition of identification we prove

Lemma. *Under Assumptions 1 and 2 we have* $[\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow (\bar{A} = A, \bar{B} = B, \bar{C} = C, \bar{D} = D, \bar{\Sigma} = \Sigma)] \Leftrightarrow [\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow (T = \mathrm{I}_n, U = \mathrm{I}_k)]$.

Proof. If the solution of the system $\bar{A} = TAT^{-1}; \bar{B} = TBU; \bar{C} = CT^{-1}; \bar{D} = DU; \bar{\Sigma} = U^{-1}\Sigma U'^{-1}; s_0 = Ts_0$ (with respect to T, U) is $T = \mathrm{I}_n, U = \mathrm{I}_k$, then the system simplifies to $\bar{A} = A, \bar{B} = B, \bar{C} = C, \bar{D} = D, \bar{\Sigma} = \Sigma$. On the other hand, if $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E}$ holds, it follows that $\bar{D} = DU$ and $[\bar{C}' \vdots \bar{A}'\bar{C}' \vdots \bar{A}'^2\bar{C}' \vdots \ldots \vdots \bar{A}'^{n-1}\bar{C}'] = [C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C']T^{-1}$. If those equations have a unique solution $\bar{D} = D$, then by Assumption 1 $\bar{D} = DU \Rightarrow D = DU \Leftrightarrow U = D^{-1}D = \mathrm{I}_k$. Moreover, if $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E}$ implies $\bar{A} = A, \bar{C} = C$, then $[C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C'] = [C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C']T^{-1}$. But by Assumption 2, $[C' \vdots A'C' \vdots A'^2C' \vdots \ldots \vdots A'^{n-1}C']$ possesses a left inverse, hence it follows that $T^{-1} = \mathrm{I}_n \Leftrightarrow T = \mathrm{I}_n$. □

Using the above Lemma we have $p(y; s_0, A, B, C, D, \Sigma) = p(y; s_0, \bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}); \forall y \in \mathcal{Y} \Rightarrow (\bar{A} = A, \bar{B} = B, \bar{C} = C, \bar{D} = D, \bar{\Sigma} = \Sigma, \bar{\theta} = \theta)$ if and only if $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow (\bar{A} = A, \bar{B} = B, \bar{C} = C, \bar{D} = D, \bar{\Sigma} = \Sigma, \bar{\theta} = \theta)$, if and only if $[\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow T = \mathrm{I}_n, U = \mathrm{I}_k] \wedge [\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow \bar{\theta} = \theta]$, if and only if $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, s_0 \in \mathcal{E} \Rightarrow T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{\theta} = \theta$.

Appendix 5: Proof of Proposition 3

Using Definition 4, a DSGE model is globally identified at $\theta \in \Theta$ if and only if $\mathcal{E} \wedge \mathcal{Z} = \{\bar{\Gamma}_1^s T A T^{-1} = \bar{\Gamma}_0^s + \bar{\Gamma}_0^p \bar{F} - \bar{\Gamma}_1^p \bar{F} T A T^{-1}, \bar{\Gamma}_1^s T B U = \bar{\Gamma}_2, C = (\bar{G}^s + \bar{G}^p \bar{F}) T A, DU = (\bar{G}^s + \bar{G}^p \bar{F}) T B U + \bar{J}, \bar{\Sigma} = U^{-1} \Sigma U'^{-1}, s_0 = T s_0\}$ implies $T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{\theta} = \theta$, implies $T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{F} = F, \bar{\Gamma}_1^s = \Gamma_1^s, \bar{\Gamma}_0^s = \Gamma_0^s, \bar{\Gamma}_0^p = \Gamma_0^p, \bar{\Gamma}_1^p = \Gamma_1^p, \bar{\Gamma}_2 = \Gamma_2, \bar{G}^s = G^s, \bar{G}^p = G^p, \bar{J} = J$. Applying the tautology $[\alpha \Rightarrow (\beta \wedge \gamma)] \Rightarrow [(\alpha \wedge \beta) \Rightarrow \gamma]$ that is true for any statements α, β, γ, we have $[(\mathcal{E} \wedge \mathcal{Z}) \Rightarrow T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{\theta} = \theta] \Rightarrow [(\mathcal{E} \wedge \mathcal{Z}, T = \mathrm{I}_n, U = \mathrm{I}_k, \bar{F} = F) \Rightarrow \bar{\Gamma}_1^s = \Gamma_1^s, \bar{\Gamma}_0^s = \Gamma_0^s, \bar{\Gamma}_0^p = \Gamma_0^p, \bar{\Gamma}_1^p = \Gamma_1^p, \bar{\Gamma}_2 = \Gamma_2, \bar{G}^s = G^s, \bar{G}^p = G^p, \bar{J} = J]$. The last conclusion holds if and only if $\{\bar{\Gamma}_1^s A = \bar{\Gamma}_0^s + \bar{\Gamma}_0^p F - \bar{\Gamma}_1^p F A, \bar{\Gamma}_1^s B = \bar{\Gamma}_2, C = (\bar{G}^s + \bar{G}^p F) A, D = (\bar{G}^s + \bar{G}^p F) B + \bar{J}\}$ has one and only one solution with respect to $\bar{\Gamma}_1, \bar{\Gamma}_2, \bar{\Gamma}_0, \bar{G}$ and $\bar{J}$. To simplify the notation, let us drop the bar symbols connected with semi-structural parameters. Let all linear restrictions imposed on $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$ be written as $\Upsilon vec([\Gamma_1 \vdots \Gamma_0 \vdots \Gamma_2]') = d$, where $\Upsilon : r_\Gamma \times (n+q)(2n+2q+k)$ is a known matrix, $d : (r_\Gamma \times 1)$ is a known vector and r_Γ is the number of independent restrictions imposed on $\Gamma_1, \Gamma_0, \Gamma_2$. Similarly, let all linear restrictions on G and J be written as $\Psi vec([G \vdots J]') = h$, where $\Psi : r_{GJ} \times r(n+q+k)$ is a known matrix, $h : (r_{GJ} \times 1)$ is a known vector and r_{GJ} is the number of all linear restrictions imposed on G and J. All equations together with the restrictions can be compactly written

$$\begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} vec([\Gamma_1 \vdots \Gamma_0 \vdots \Gamma_2]') \\ vec([G \vdots J]') \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

where $Q = \begin{bmatrix} \mathrm{I}_{n+q} \otimes \begin{bmatrix} A' & A'F' & -\mathrm{I}_n & -F' & 0_{n\times k} \\ B' & 0_{k\times q} & 0_{k\times n} & 0_{k\times q} & -\mathrm{I}_k \end{bmatrix} \\ \Upsilon \end{bmatrix}$, $R = \begin{bmatrix} \mathrm{I}_r \otimes \begin{bmatrix} A' & A'F' & 0_{n\times k} \\ B' & B'F' & \mathrm{I}_k \end{bmatrix} \\ \Psi \end{bmatrix}$,

$c_1 = \begin{bmatrix} 0_{(n+k)(n+q)\times 1} \\ d \end{bmatrix}$, $c_2 = \begin{bmatrix} vec([C \vdots D]') \\ h \end{bmatrix}$. Since the system is linear in $\Gamma_1, \Gamma_0, \Gamma_2, G$ and J (and consistent by construction), the solution will be unique if and only if $\begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}$ has full column rank, which holds if and only if Q is of full column rank and R is of full

column rank. The last conclusion of the proposition follows from the fact that for any matrix to have a full column rank it is necessary that the number of rows must at least equal to the number of columns. This imposes a requirement on the minimal number of rows in Υ and Ψ.

Appendix 6: Proof of Corollary 2

Assume there are no restrictions involving parameters from different rows in $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$. Denote by $\Gamma_{(i)}$ the ith row of $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$. Let all linear restrictions imposed on the ith row of $[\Gamma_1^s \vdots \Gamma_1^p \vdots \Gamma_0^s \vdots \Gamma_0^p \vdots \Gamma_2]$ be written as $\Upsilon_i \Gamma'_{(i)} = d_i$, where $\Upsilon_i : r_i \times (2n + 2q + k)$ is a known matrix, $d_i : (r_i \times 1)$ is a known vector and r_i is the number of independent restrictions imposed on $\Gamma_{(i)}$. In addition, assume that the linear restrictions imposed on $[G \vdots J]$ are given precisely as in the previous appendix. In this case, matrix Q from Proposition 3 (i.e. previous appendix) takes the following form

$$Q = \begin{bmatrix} Q^\star & 0 & 0 & 0 \\ 0 & Q^\star & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & Q^\star \\ \Upsilon_1 & 0 & 0 & 0 \\ 0 & \Upsilon_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \Upsilon_{n+q} \end{bmatrix}$$

where $Q^\star = \begin{bmatrix} A' & A'F' & -\mathrm{I}_n & -F' & 0_{n\times k} \\ B' & 0_{k\times q} & 0_{k\times n} & 0_{k\times q} & -\mathrm{I}_k \end{bmatrix}$. By Proposition 3, if a DSGE model is globally identified at $\theta \in \Theta$, then Q is of full column rank and R is of full column rank. But in the present case Q is of full column rank iff $Q_i = \begin{bmatrix} A' & A'F' & -\mathrm{I}_n & -F' & 0_{n\times k} \\ B' & 0_{k\times q} & 0_{k\times n} & 0_{k\times q} & -\mathrm{I}_k \\ & & \Upsilon_i & & \end{bmatrix}$ is of full column rank for each $i = 1, ..., n + q$. The last conclusion is proved by arguments used in the previous appendix.

Appendix 7: Mathematical considerations on local and global identification

In this appendix we indicate some possible problems in establishing local and global identification in DSGE modeling. To fix ideas, we assume $\varepsilon_t \sim i.i.d.\, N(0, \Sigma)$ and $s_0 \neq 0$

throughout this section (though see the remark at the end of this appendix). What Proposition 1 says formally (see Appendix 3) is that the equivalence class from the conclusion of Theorem 1, i.e. $\mathcal{E} = \{TAT^{-1}, TBU, CT^{-1}, DU, U^{-1}\Sigma U'^{-1}, Ts_0 |\ T, U \text{ nonsingular}\}$, is the manifold of dimension $2nk + \frac{1}{2}k(k+1) + n$, for every A, B, C, D, Σ, s_0. Hence, $\mathcal{E}$ may be fully parametrized with $2nk + \frac{1}{2}k(k+1) + n$ elements. In fact, the classic theory of canonical forms in linear system theory is all about finding such a parametrization. Although we can readily obtain one canonical form for the general ABCD-representation,[15] finding the canonical form for the ABCD-representation that is explicitly derived from a given DSGE model is not an easy task. The problem is that it is difficult to arrive at the canonical form that obeys the restrictions explicitly imposed on semi-structural parameters Γ_0, Γ_1, Γ_2, G and J. What we know for sure is that in every $\mathcal{E}$ (for each A, B, C, D, Σ, s_0) we can find at least one canonical form that obeys the restrictions imposed on semi-structural parameters (put $T = \mathrm{I}_n$ and $U = \mathrm{I}_k$). This is sufficient for our further mathematical considerations.

The usual situation in DSGE modeling is that the number of deep parameters is relatively small in comparison with the number of functionally independent elements in the ABCD-representation. Interestingly, this fact has far reaching consequences for global (and local) identification, which will be explained below. To this end, we first reproduce the relevant definitions and theorems from differential topology and next discuss them in the context of our problem.

In general, we are interested in the behavior of some function $f : V \to \mathcal{E}$, where V is an open subset of $\mathbb{R}^p$, for some p (think of V as a space of all deep or semi-structural parameters). To keep our presentation as simple as possible, we suppose that the canonical form is given. Note that the canonical form is parametrized with $n_{\mathcal{E}} = 2nk + \frac{1}{2}k(k+1) + n$ functionally independent elements. This amounts to considering $f : V \to \mathbb{R}^{n_\mathcal{E}}$, $x = (x_1, x_2, \ldots, x_p) \mapsto f(x) = (f_1(x), f_2(x), \ldots, f_{n_\mathcal{E}}(x))$. If each component f_i possesses all its partial derivatives of order $r \geq 1$, which are continuous at all points $x \in V$, we say that f is a C^r-map. A $C^\infty-$ map is called a smooth function. An $(n_\mathcal{E} \times p)$ Jacobian matrix

[15] Consider the function $g(A, B, C, D, \Sigma, s_0) = (T_0AT_0^{-1}, T_0BD^{-1}, CT_0^{-1}, \mathrm{I}_k, D\Sigma D', T_0s_0)$, where T_0 is a nonsigular matrix comprising the first n independent rows from the observability matrix O (which must exist by Assumption 2). Then it is easy to show that $\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, \bar{s}_0 \in \mathcal{E}$ if and only if $g(\bar{A}, \bar{B}, \bar{C}, \bar{D}, \bar{\Sigma}, \bar{s}_0) = g(A, B, C, D, \Sigma, s_0)$. Hence, setting $A^\star = T_0AT_0^{-1}$, $B^\star = T_0BD^{-1}$, $C^\star = CT_0^{-1}$, $D^\star = \mathrm{I}_k$, $\Sigma^\star = D\Sigma D'$, $s_0^\star = T_0s_0$, the pdf of the general ABCD-representation may be parametrized with uniformly globally identified parameter $s_0^\star$, $A^\star$, $B^\star$, $C^\star$, $D^\star$, $\Sigma^\star$ i.e. $p(y; s_0, A, B, C, D, \Sigma) = p(y; s_0^\star, A^\star, B^\star, C^\star, D^\star, \Sigma^\star)$ for all $y \in \mathcal{Y}$ and all s_0, A, B, C, D, Σ. You can convince yourself that $s_0^\star$, $A^\star$, $B^\star$, $C^\star$, $D^\star$, $\Sigma^\star$ contain in total $2nk + \frac{1}{2}k(k+1) + n$ functionally independent elements, as suggested by Proposition 1. This fact will be made evident if we analytically derive $T_0AT_0^{-1}$ and CT_0^{-1}. It turns out that those two matrices contain nk functionally independent elements (in particular, many 0's and 1's). For example, if C has full row rank then CT_0^{-1} comprises only 0's and 1's, otherwise it also consists of unrestricted elements, see e.g. Guidorzi (1975) or Goodwin and Payne (1977), pp. 66-68.

is defined as $J(x) := \partial f'(x)/\partial x$. If the rank of $J(x)$ is p at every point $x \in V \subset \mathbb{R}^p$, we call $f : V \to \mathbb{R}^{n_\mathcal{E}}$ an immersion. In other words, an immersion is locally injective (i.e. locally 1-1) at every $x \in V$ (by the Implicit Function Theorem). An immersion that is an injective (i.e. 1-1) map will be called a 1-1 immersion. If, as in our application, V is the space of deep (or semi-structural) parameters, then immersion is simply a name for a model that is locally identifiable at every point of deep (or semi-structural) parameters space, and 1-1 immersion is a model that is globally identified at every point of deep (or semi-structural) parameters space.

Theorem. *(Sternberg, 1964, pp. 58-59) Let V be an open set in $\mathbb{R}^p$. Let $f : V \to \mathbb{R}^{n_\mathcal{E}}$ be a smooth function, where $n_\mathcal{E} \geq 2p$. Given any $\epsilon > 0$, there is a $p \times n_\mathcal{E}$ matrix $R = (r_{ij})$ such that for each i, j, $|r_{ij}| < \epsilon$ and such that the map $g : V \to \mathbb{R}^{n_\mathcal{E}}$ defined as $g(x) = f(x) + xR$ is an immersion, where $x = (x_1, x_2, \ldots, x_p)$ is a row vector.*

Definition. Let $f : \mathbb{R}^p \to \mathbb{R}^{n_\mathcal{E}}$ and $\delta : \mathbb{R}^p \to \mathbb{R}^+$be a positive continuous function. W say that $g : \mathbb{R}^p \to \mathbb{R}^{n_\mathcal{E}}$ is a $\delta-$approximation of f if $d(f(x), g(x)) < \delta(x)$ for all $x \in \mathbb{R}^p$, where d is the metric in $\mathbb{R}^{n_\mathcal{E}}$.

Theorem. *(Adachi, 1993, p. 49 and 54) Given a smooth function $f : \mathbb{R}^p \to \mathbb{R}^{n_\mathcal{E}}$ and the positive continuous function $\delta : \mathbb{R}^p \to \mathbb{R}^+$we have: a) if $n_\mathcal{E} \geq 2p$, there exists an immersion $g : \mathbb{R}^p \to \mathbb{R}^{n_\mathcal{E}}$ which is a $\delta-$approximation of f ; b) if $n_\mathcal{E} \geq 2p + 1$, there exists a 1-1 immersion $g : \mathbb{R}^p \to \mathbb{R}^{n_\mathcal{E}}$ which is a $\delta-$approximation of f .*

To state the last theorem in this section, let $C^r(V, \mathbb{R}^{n_\mathcal{E}})$ denote the space of all C^r-maps from V to $\mathbb{R}^{n_\mathcal{E}}$. The elements of $C^\infty(V, \mathbb{R}^{n_\mathcal{E}})$ are smooth functions. Let us define the norm $\|f\|_{r,L} := \{|f|_r | x \in L, r \geq 1\}$, where $|f|_r = \underset{i,x}{sup}|f_i(x)| + \underset{i,j,x}{sup}\left|\frac{\partial f_i(x)}{\partial x_j}\right| + \cdots + \underset{i,\alpha,x}{sup}\left|\frac{\partial^r f_i(x)}{(\partial x_1)^{\alpha_1}(\partial x_2)^{\alpha_2}\ldots(\partial x_p)^{\alpha_p}}\right|$, where $\alpha_1 + \alpha_2 + \ldots + \alpha_p = r$. In other words, g will be close to f in $\|\cdot\|_{r,L}$ norm if g and all its partial derivatives up to order r will be uniformly (on L) close to those of f. This norm makes $C^\infty(V, \mathbb{R}^{n_\mathcal{E}})$ into topological space $C^\infty(V, \mathbb{R}^{n_\mathcal{E}})_{r,L}$, where neighborhoods are defined in terms of $\|\cdot\|_{r,L}$.

Theorem. *(Hirsch, 1976, ch. 2) Let V be an open set in $\mathbb{R}^p$ and $L \subset V$ compact. Then a) if $n_\mathcal{E} \geq 2p$, the set of all smooth immersions $f : \mathbb{R}^p \to \mathbb{R}^{n_\mathcal{E}}$ is open and dense in $C^\infty(V, \mathbb{R}^{n_\mathcal{E}})_{r,L}$; and b) if $n_\mathcal{E} \geq 2p + 1$, the set of all smooth 1-1 immersions is open and dense in $C^\infty(V, \mathbb{R}^{n_\mathcal{E}})_{r,L}$.*

Let us interpret the above theorems in the context of DSGE models. To this end, think of the domain $f : V \to \mathbb{R}^{n_\mathcal{E}}$ as the space of all deep parameters Θ (recall that the range is the space of functionally independent elements in an equivalence class), hence $p = dim(\Theta)$

and $n_{\mathcal{E}} = 2nk + \frac{1}{2}k(k+1) + n$.[16] First, the first theorem of this appendix is suggestive in that it undermines the sense of checking local identifiability using the derivatives since $\partial g'(\theta)/\partial\theta = \partial f'(\theta)/\partial\theta + R'$, where all entries in matrix R are arbitrarily close to 0. Hence, as long as $n_{\mathcal{E}} \geq 2 \cdot dim(\Theta)$, such a defined g has derivatives arbitrarily close to that of f at all $\theta \in \Theta$, but at the same time is locally identified at all $\theta \in \Theta$. Second, the second theorem says that as long as $n_{\mathcal{E}} \geq 2 \cdot dim(\Theta) + 1$, any smooth map $f : \Theta \to \mathbb{R}^{n_{\mathcal{E}}}$ can be approximated arbitrarily closely by smooth 1-1 immersions, hence by a model that is globally identified at all $\theta \in \Theta$ (i.e. uniformly).

There is an obvious consequence of this result. Bearing in mind that in most applications the mapping from θ to A, B, C, D, Σ is only numerically given, establishing both local and global identification using differential calculus is not possible if $n_{\mathcal{E}} \geq 2 \cdot dim(\Theta) + 1$. The last theorem suggests that if $n_{\mathcal{E}} \geq 2 \cdot dim(\Theta)$ ($n_{\mathcal{E}} \geq 2 \cdot dim(\Theta) + 1$), the property of a DSGE model to be uniformly locally (globally) identified is generic (i.e. typical). Although it is known that there are generic properties that at the same time can have arbitrarily small Lebesgue measure, in our case the notions of "genericity" and "almost everywhere" are equivalent in some formal sense, see e.g. Hunt et al. (1992). Hence, the last theorem roughly means that if $n_{\mathcal{E}} \geq 2 \cdot dim(\Theta)$ ($n_{\mathcal{E}} \geq 2 \cdot dim(\Theta) + 1$), almost all smooth functions are locally identified (globally identified) at all $\theta \in \Theta$. In other words, if $n_{\mathcal{E}} \geq 2 \cdot dim(\Theta) + 1$ and numerical solution to a DSGE model is used, you cannot arrive at the model that is not uniformly globally identified at all $\theta \in \Theta$ since an infinitely small perturbation to the model solution algorithm results in a uniformly globally identified model.[17] Moreover, even if your linearized DSGE model is not (uniformly) globally or locally identified, taking into account higher order terms in the approximation (or just analyzing the nonlinear model as it is) will almost surely identify your model. Another implication of the above theorems may be stated informally as: if you tell me that condition $n_{\mathcal{E}} \geq 2 \cdot dim(\Theta) + 1$ is met in your particular DSGE model (keep in mind however footnote 16), I will say that there is *a priori* zero probability that your DSGE model will not be globally identified at all $\theta \in \Theta$.

Although we reproduced the theorems involving smooth functions, there are versions

[16] This makes sense provided that there is no evident non-identification at the level of the mapping from deep parameters to semi-structural parameters. Otherwise, you should think of the domain of $f : V \to \mathbb{R}^{n_{\mathcal{E}}}$ as the space of semi-structural parameters. In fact, the mapping from deep parameters to functionally independent elements may be symbolically defined as a composite map $f_2 \circ f_1$, where f_1 is a map $\theta \mapsto \Gamma_1, \Gamma_0, \Gamma_2, G, J$ and f_2 is a map $\Gamma_1, \Gamma_0, \Gamma_2, G, J \mapsto u$ (where u denotes an element of $2nk + \frac{1}{2}k(k+1) + n$ dimensional manifold). Clearly, identification (i.e. injection) of f_1 is the necessary condition for identification (i.e. injection) of the whole composite map $f_2 \circ f_1$.

[17] A good analogue is computation of the determinant of a square (say $n \times n$) matrix, in which each entry is generated independently from some continuous distribution. Since the set of nonsingular matrices forms an open and dense subset of $\mathbb{R}^{n^2}$, the probability that your randomly selected matrix is singular has Lebesgue measure 0.

that require only C^r-maps for $r \geq 1$, see e.g. Hirsch (1976). Moreover, since $\|f\|_{r,L}$ norm does not control the closeness "at infinity" (since L is compact), it may not be always fully satisfying. However, there is a modified norm that results in the so-called strong topology. See e.g. Hirsch (1976) for this more general treatment. Further, the material in this appendix is literally the same if instead of $\varepsilon_t \sim i.i.d.\, N(0, \Sigma)$ we assume $\varepsilon_t \sim i.i.d.\, N(0, \mathrm{I}_k)$ since the dimension of manifold $\mathcal{E}$ remains the same (see Proposition 1). On the other hand, if $s_0 = 0$, then all results in this appendix are perfectly valid given that we adjust the dimension of $\mathcal{E}$ so as $n_{\mathcal{E}} = 2nk + \frac{1}{2}k(k+1)$.

www.nbp.pl