Evaluating the strength of identification in DSGE models. An a priori approach

Nikolay Iskrev
Banco de Portugal

This draft: 26 May 2009

PRELIMINARY. COMMENTS WELCOME

Abstract

Reliable estimation of economic models is possible only when the parameters are well identified. This paper proposes a new framework for evaluating the strength of identification in linearized dynamic stochastic general equilibrium models (DSGE) models prior to their estimation. Identification fails or is weak when the empirical implications of the parameters are not sufficiently strong or distinct. By following the steps and applying the tools described in the paper, researchers can assess how well identified the parameters are, and determine the causes for identification problems when they occur.

Keywords: DSGE models, Weak identification, Information matrix

JEL classification: C32, C51, C52, E32

Contact Information: Banco de Portugal, Av. Almirante Reis 71, 1150-012, Lisboa, Portugal
e-mail: Nikolay.Iskrev@bportugal.pt; phone:+351 213130006

1 Introduction

There is a considerable consensus among academic economists and economic policy makers that modern macroeconomic models are rich enough to be useful as tools for policy analysis. It is also well understood that when structural models are used for quantitative analysis, it is crucial to use parameter values that are empirically relevant. The best way of obtaining such values is to estimate and evaluate the models in a formal and internally consistent manner. This is what the new empirical dynamic stochastic general equilibrium (DSGE) literature attempts to do.

The estimation of DSGE models exploits the restrictions they impose on the joint probability distribution of observed macroeconomic variables. A fundamental question that arises is whether these restrictions are sufficient to make possible the reliable estimation of the parameters. This is known as the identification problem in econometrics, and to answer it econometricians study the relationship between the true probability distribution of the data and the parameters of the underlying economic model (Koopmans (1949)). Such identification analysis should precede the statistical estimation of economic models (Manski (1995)).[1]

Although the importance of parameter identification has been recognized, the issue is rarely discussed when DSGE are estimated. Examples of models with unidentifiable parameters can be found in Kim (2003), Beyer and Farmer (2004) and Cochrane (2007). That DSGE models may be poorly identified has been pointed out by Sargent (1976) and Pesaran (1989). More recently, Canova and Sala (2009) summarize their study of identification issues in DSGE models with the conclusion: "it appears that a large class of popular DSGE structures can only be weakly identified".

This paper develops a new framework for studying identification issues in linearized DSGE models. It takes as a starting point a description of the economic model - structural equations and parameter space, and a description of the statistical model - a Gaussian likelihood for a set of observable endogenous variables, and a sample size.

[1]This is not always possible, however. If the distribution of the data involves unknown nuisance parameters, the identification of the parameters of interest would be conditional on those parameters. In such models the identification problem is no longer separate from that of statistical inference. DSGE models typically do not have this problem since all parameters are deemed structural and are either estimated or assumed known.

Then it provides an answer to each one of the following questions: (1) which parameters are locally identified and which are not; (2) how well identified are the identifiable parameters; (3) if identification fails for some parameters or is weak, is this due to data limitations, or is it intrinsic to the structure of the model; (4) how the answers to (1)-(3) vary across different regions in the parameter space.

The central tool employed in this framework is the expected Fisher information matrix, the use of which for identification analysis was first suggested by Rothenberg (1971). The information matrix can be derived in closed-form for linear Gaussian state space models, which explains why the identification analysis is restricted to linearized Gaussian DSGE models. This covers most, but not all estimated models in the current empirical DSGE literature.[2] Although the computation of the expected information matrix does not involve information about a particular sample of data, it does depend on which of the model variables are included in the analysis. This is one feature of the data that matters for identification. Furthermore, the strength of identification is an issue only in finite samples. Thus, having a meaningful measure of how well identified different parameters are, requires specifying the number of available observations. Consequently, this is a second feature of the data that matters for identification. However, if one can choose among different sets of observables and different sample sizes, it is straightforward to investigate what effect that would have on identification. Thus, we may add another question to those listed above, namely, (5) how the answers to (1)-(4) vary across different sets of observables and with the sample size. Finally, the information matrix is parameter-dependent, and, therefore, the conclusions regarding identifiability and identification strength may change across regions in the parameters space. The parameter space, defined as the set of theoretically admissible parameter values, should be specified as a part of the economic model under study.

It is a common misconception to regard identification as relevant to empirical work only, and to think of identification problems as caused by either deficiencies of the data or of the statistical methodology. While it is true that there is a purely statistical dimension of the problem, its economic modeling aspect is often far more important. Parameters are unidentifiable or weakly identified if the economic features they repre-

[2]In general, showing identification of a linearized Gaussian model is sufficient but not necessary for identification in the original non-linear model under more general distributional assumptions. More on the role of the Gaussian assumption can be found in Section 3.2.

sent have no empirical relevance at all, or very little of it. This may occur either because those features are unimportant on their own, or because they are redundant given the other features represented in the model. These issues are particularly relevant to DSGE models, which are sometimes criticized of being too rich in features, and possibly over-parameterized (Chari, Kehoe, and McGrattan (2009)). This paper shows how one can distinguish between the statistical and economic modeling aspects of identification problems, and provides tools for determining the causes leading to them.

Papers related to this one are Iskrev (2009), which addresses the parameter identifiability question, and Canova and Sala (2009), which is focused on the weak identification problem. In Iskrev (2009), I present an identifiability condition that is easier to use and more general than the one developed here. The condition is based on the Jacobian matrix of the mapping from theoretical first and second order moments of the observable variables to the deep parameters of the model. The condition is necessary and sufficient for identification with likelihood-based methods under normality, or with limited information methods that utilize only first and second order moments of the data. However, that paper does not deal with the weak identification issue, which is the main theme of this paper. The paper of Canova and Sala (2009) was the first one to draw attention to the problem of weak identification in DSGE models, and to discuss different strategies for detecting it. Those include: one and two dimensional plots of the estimation objective function, estimation with simulated data, and checking numerically the conditioning of matrices characterizing the mapping from parameters to the objective function. The paper of Canova and Sala (2009) differs form the present paper in several ways. First, they approach parameter identification from the perspective of a particular limited information estimation method, namely, equally weighted impulse response matching. In addition to the model and data deficiencies discussed above, weak identification in that setting may be caused by the failure to use some model-implied restrictions on the distribution of the data, and by the inefficient weighing of the utilized restrictions. Consequently, it is very difficult to disentangle the causes and quantify their separate contribution to the identification problem. Second, it is very common in DSGE models to have identification problems that stem from a near observational equivalence involving a large number of parameters. This means that the objective function is flat with respect to all of the parameters as a set. The plots used in Canova and Sala (2009) are limited to only two parameters at a time, and it is far from straightforward to select

the appropriate pairs from a large number of free parameters. Third, Canova and Sala (2009) do not discuss the role of the set of observables for identification. The effect of using different observables for the estimation of a DSGE model is investigated in Guerron (2007), who finds that the parameter estimates and the economic and forecasting implications of the model vary substantially with the choice of included variables. The last and perhaps most important difference is in the approach itself. While it is possible in principle to address all identification questions discussed here, by conducting Monte Carlo simulations, this is hardly a viable strategy for an a priori identification analysis of most DSGE models. Estimating a multidimensional and highly non-linear model even once is a numerically challenging and time consuming exercise. Doing that many times and for a large number of parameter values is completely impractical. In contrast, the tools used in this paper are simple, easy to apply, and general.

The remainder of the paper is organized as follows. Section 2 provides some background on identification and the Fisher information matrix, and explains the role of the latter for detecting and measuring identification problems in general parametric models. Section 3 starts with an introduction of the class of linearized DSGE models, followed by an outline of the derivation of the log-likelihood function and the Fisher information matrix for Gaussian models. Then it explains the various aspects and tools involved in the a priory approach to identification. The methodology is illustrated, in Section 4, with the help of a medium-scale DSGE model estimated in Smets and Wouters (2007). Concluding comments are given in Section 5.

2 Identification and the Information Matrix

Let a model be parameterized in terms of a vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$, and suppose that inference about $\boldsymbol{\theta}$ is made on the basis of T observations of a random vector $\boldsymbol{x}$ with a known joint probability density function $f(\boldsymbol{X};\boldsymbol{\theta})$, where $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\mathrm{T}}]$. When considered as a function of $\boldsymbol{\theta}$, $f(\boldsymbol{X};\boldsymbol{\theta})$ contains all available sample information about the value of $\boldsymbol{\theta}$ associated with the observed data. Thus, a basic prerequisite for making inference about $\boldsymbol{\theta}$ is that distinct values of $\boldsymbol{\theta}$ imply distinct values of the density function. Formally, we say that a point $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is identified if

$$f(\boldsymbol{X};\boldsymbol{\theta}) = f(\boldsymbol{X};\boldsymbol{\theta}_0) \text{ with probability } 1 \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0 \tag{2.1}$$

This definition is made operational by using the following property of the log-likelihood function[3] $\ell_T(\boldsymbol{\theta}) := \log f(\boldsymbol{X}; \boldsymbol{\theta})$

$$\mathrm{E}_0\, \ell_T(\boldsymbol{\theta}_0) \geq \mathrm{E}_0\, \ell_T(\boldsymbol{\theta}), \text{ for any } \boldsymbol{\theta} \tag{2.2}$$

It follows that the function $H(\boldsymbol{\theta}_0, \boldsymbol{\theta}) := \mathrm{E}_0\, (\ell_{\mathrm{T}}(\boldsymbol{\theta}) - \ell_T(\boldsymbol{\theta}_0))$ achieves a maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and $\boldsymbol{\theta}_0$ is identified if and only if that maximum is unique. While conditions for global uniqueness are difficult to find in general, local uniqueness of the maximum at $\boldsymbol{\theta}_0$ may be established by verifying the usual first and second order conditions, namely: (a) $\frac{\partial H(\boldsymbol{\theta}_0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$, (b) $\frac{\partial^2 H(\boldsymbol{\theta}_0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is negative definite. If the maximum at $\boldsymbol{\theta}_0$ is locally unique we say that $\boldsymbol{\theta}_0$ is locally identified. This means that there exists an open neighborhood of $\boldsymbol{\theta}_0$ where (2.1) holds for all $\boldsymbol{\theta}$. Global identification, on the other hand, extends the uniqueness of $f(\boldsymbol{X}; \boldsymbol{\theta}_0)$ to the whole parameter space. One can show that (see Bowden (1973)) the condition in (a) is always true, and the Hessian matrix in (b) is equal to the negative of the Fisher information matrix, defined as

$$\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}) = \mathrm{E}_0 \left[\left\{ \frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right\}' \left\{ \frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right\} \right]$$

Thus, we have the following result of Rothenberg (1971),

Theorem 1. *Let $\boldsymbol{\theta}_0$ be a regular point of the information matrix $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta})$ Then $\boldsymbol{\theta}_0$ is locally identifiable if and only if $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}_0)$ is non-singular.*

A point is called regular if it belongs to an open neighborhood where the rank of the matrix does not change. Without this assumption the condition is only sufficient for local identification. Although it is possible to construct examples where regularity does not hold (see Shapiro and Browne (1983)), typically the set of irregular points is of measure zero (see Bekker and Pollock (1986)). Thus, for most models the non-singularity of the information matrix is both necessary and sufficient for local identification. By definition, a model is (locally) identified if all points in the parameter space are (locally) identified. This can be checked by examining the rank of the information matrix at all points in $\boldsymbol{\Theta}$.

Verifying that the model is identified, at least locally, is important since identifiability is a prerequisite for the consistent estimation of the parameters. Intuitively, singularity

[3]This follows from the Jensen's inequality (see Rao (1973)) and the fact that the logarithm is a concave function.

of the information matrix means that likelihood function is flat at $\boldsymbol{\theta}_0$ and one has no hope of finding the true values of some the parameters even with an infinite number of observations. There are two possible reasons why a parameter θ_i may be locally unidentifiable:

(a) Changing θ_i does not change the likelihood, i.e.

$$\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i} = 0, \text{ for all } \boldsymbol{X} \tag{2.3}$$

(b) The effect on the likelihood of changing θ_i can be offset by changing other parameters in $\boldsymbol{\theta}$, i.e.

$$\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i} = \sum_{j \neq i} a_j \frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_j}, \text{ for all } \boldsymbol{X} \tag{2.4}$$

where a_j, $j \neq i$ are scalars.

In the first case row i and column i of the information matrix are vectors of zeros; in the second they are equal to linear combinations of the other rows/columns of the information matrix. The likelihood is flat with respect to θ_i in (a), and with respect to a linear combination of several elements of $\boldsymbol{\theta}$ in (b).

The rank condition ensures that the likelihood function is not flat, and that the expected log-likelihood function achieves a locally unique maximum at the true parameter values. However, it provides no information about the finite sample properties of a likelihood-based estimator of $\boldsymbol{\theta}$, which depend greatly on the degree of curvature of log-likelihood surface in the neighborhood of $\boldsymbol{\theta}_0$. Nearly flat likelihood means that small changes in the value of $\ell_T(\boldsymbol{\theta}_0)$, due to random variations in the data, result in relatively large changes in the value of $\boldsymbol{\theta}$ that maximizes the observed likelihood function. In this situations parameter identification is said to be weak in the sense that the estimates are prone to be very inaccurate even when the number of observations is large.

There is now a substantial literature on weak identification in econometrics, and in particular on the weak instruments problem in linear models. Yet, unlike identification in the strict sense, there does not exist a general definition of weakness, which one can apply to determine if a parameter or a model is weakly identified. Intuitively, it is clear the flatter the likelihood function, the less precise the parameter estimates will

be. This suggests that measures of estimation precision may be used to quantify the strength of parameter identification. Adopting this notion of identification strength, a parameter is considered weakly identified when the degree of precision with which it can be estimated is unacceptably low. In that sense what weak means depends on what is considered unacceptable, and is therefore a relative not an absolute concept.

The relationship between the degree of curvature of the expected likelihood function and the precision of a ML estimator $\hat{\boldsymbol{\theta}}_T$ is obtained from the asymptotic distribution of the latter. Assuming that $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0$, we have

$$\sqrt{\mathrm{T}}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}, \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}_0)\right) \tag{2.5}$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0) = \operatorname*{plim}_{\mathrm{T}\to\infty} \frac{1}{\mathrm{T}} \boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}_0)$ is the asymptotic expected Fisher information matrix. Therefore $\operatorname{cov}(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\mathrm{T}} \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}_0)$ provides an approximation of the sampling covariance matrix of $\hat{\boldsymbol{\theta}}_\mathrm{T}$, and $\operatorname{var}(\hat{\theta}_i) = \frac{1}{\mathrm{T}} \boldsymbol{\mathcal{I}}^{(ii)}(\boldsymbol{\theta}_0)$ approximates the sampling variance of $\hat{\theta}_i$, where $\boldsymbol{\mathcal{I}}^{(ii)}$ is the i-th diagonal element of the inverse of information matrix. Also, the asymptotic normality of $\hat{\boldsymbol{\theta}}_T$ implies that

$$T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)' \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(k) \tag{2.6}$$

Using (2.6), we can compute asymptotic confidence intervals for each θ_i, and joint confidence sets for $\boldsymbol{\theta}$ as a whole. In particular, a joint $(1-\alpha)$ confidence set contains all points $\boldsymbol{\theta}$ that satisfy

$$T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})' \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \leq c_\alpha \tag{2.7}$$

where c_α is the $(1-\alpha)$ quantile of the $\chi^2(k)$ distribution. Individual confidence intervals for each parameter can be constructed by projecting the k-dimensional ellipsoid defined by (2.7) onto the parameter axes. This leads to intervals of the form

$$\theta_i - \sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)} \leq \hat{\theta}_i \leq \theta_i + \sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)} \tag{2.8}$$

This interval has the following interpretation: in repeated samples the ML estimate of θ_i will fall within the interval $[\theta_i - \sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)}; \theta_i + \sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)}]$ in $(1-\alpha)\%$ of the time. Several caveats should be added to this interpretation, however. First, the joint confidence sets are constructed on the basis of a quadratic approximation of the pop-

ulation objective function - the expected log-likelihood, and may give an inaccurate assessment of the sampling uncertainty in finite samples. Second, the individual confidence intervals are conservative in the sense that the multidimensional rectangle defined by (2.8) includes points that lie outside the hyperellipsoid defined by (2.7). Third, the confidence intervals (2.8) are solely based on the curvature of the likelihood and do not incorporate any prior knowledge that may be available about the elements of $\boldsymbol{\theta}$. In DSGE models the parameters have clear economic meaning and are usually restricted to a range of possible values. The confidence intervals do not reflect such restrictions and may include values which are theoretically inadmissible.

Constructing confidence intervals as in (2.8) is a convenient way of assessing how well identified are the individual parameters on the basis of the likelihood function. A parameter is considered to be weakly identified if the interval is too wide at the desired level of confidence. However, this approach is not convenient for comparing the identification strength across different parameters or across different values of the same parameter. For that I will use the ratio of the half-length of the confidence interval to the absolute value of θ_i, i.e.

$$r(\theta_i) = \frac{\sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)}}{|\theta_i|} \tag{2.9}$$

Note that $r(\theta_i)$ is not defined when $\theta_i = 0$. When this is the case, the endpoints of the confidence interval should be reported instead of the ratio $r(\theta_i)$. In addition to the individual measures, I will also need a measure of the strength of identification of $\boldsymbol{\theta}$ as a whole. For that I will use the geometric average of the individual measures, i.e.

$$\bar{r}(\boldsymbol{\theta}) = \left(\prod_{i=1}^{k} r(\theta_i) \right)^{\frac{1}{k}} \tag{2.10}$$

Here the individual measures are weighted equally, which means that the parameters are considered to be equally important. When this is not the case, the weighted geometric average should be used instead. Denoting the weights by $w_1, w_2, \ldots, w_k$, the weighted measure of overall strength of identification is

$$\bar{r}_w(\boldsymbol{\theta}) = \exp\left(\frac{\sum_i^k w_i r(\theta_i)}{\sum_i^k w_i} \right) \tag{2.11}$$

Note that these quantities are independent of the units of measurement. Thus, they can be compared for different elements or values of $\boldsymbol{\theta}$.

The causes of weak identification may be expressed as in (2.3) and (2.4) by replacing the "=" sign with "≈". This means that the likelihood is almost, though not completely, flat with respect to one component of $\boldsymbol{\theta}$ in the first case, or with respect to a set of parameters - in the second. The weak identification version of (a) is equivalent to having a very small variance of i-th component of the score vector $\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$; the one in (b) is equivalent to having strong linear dependence, or near-collinearity, among the components of the score. To isolate the effects of these two causes we can factorize the information matrix as follows:

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0) = \boldsymbol{D}^{\frac{1}{2}} \tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\theta}_0) \boldsymbol{D}^{\frac{1}{2}} \tag{2.12}$$

where $\boldsymbol{D} = \text{diag}(\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}_0))$ is a diagonal matrix containing the variances of the elements of the score vector, and $\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\theta}_0)$ is the correlation matrix of the score vector. It is straightforward to show that the asymptotic variance of $\hat{\theta}_i$ can be expressed as[4]

$$\text{var}(\hat{\theta}_i) = \frac{1}{TD_i} \left(\frac{1}{1 - \boldsymbol{\varrho}_i^2} \right) \tag{2.13}$$

where D_i is the i-th element of $\boldsymbol{D}$, and $\boldsymbol{\varrho}_i$ is the multiple correlation coefficient between the i-th element and all other elements of the score. From (2.13) the variance of $\hat{\theta}_i$ is large if either D_i - the variance of $\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i}$, is small, or if $\boldsymbol{\varrho}_i$ - the degree of collinearity between $\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i}$ and the other elements of the score, is strong.

The decomposition in (2.13) is useful to researchers who wish to find out what features of the model lead to identification problems for some parameters. Consider the first term. A very small variance of $\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i}$ means that the likelihood is very insensitive to θ_i, or, in other words, that the statistical implications of that parameter are hard to detect. Hence, the economic feature represented by θ_i is not very important empirically. Note that $\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i}$ and its variance are not scale invariant, i.e. it matters in what units the parameters in $\boldsymbol{\theta}$ are measured. A unit-free measure is $\theta_i \frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \theta_i}$ - the elasticity of the likelihood with respect to θ_i. Thus, instead of D_i we should compare $\theta_i^2 D_i$ to asses

[4]This follows from the fact that the diagonal of the inverse of the correlation matrix contains the squared multiple correlation coefficients, see e.g. Tucker, Cooper, and Meredith (1972)

the relative importance of the elements of $\boldsymbol{\theta}$. The second term in (2.13) captures the fact that there may be some degree of overlap between the statistical implications of different parameters. The closer is $\boldsymbol{\varrho}_i$ to one, the more difficult it is to distinguish θ_i from the other elements of $\boldsymbol{\theta}$. In that sense we may say that, from an empirical point of view, the economic feature represented by θ_i is nearly redundant, given the other features of the model.

It is important to remember that the conclusions from such analysis are conditional on the statistical model one uses to evaluate the underlying economic model. A particular statistical model, be it the joint density of the variables in $\boldsymbol{x}$, or a set of moments of these variables, may not capture all empirical implications of the economic model. For instance, DSGE models are typically estimated using a relatively small subset of the endogenous variables in the model. Parameters that are nearly irrelevant with respect to those variables may achieve a much sharper identification from variables that are not used in the analysis. Moreover, DSGE models are usually estimated using only first and second moments of the data. Unless the structural shocks are truly Gaussian, higher moments would provide additional information about the parameters that may make identification stronger. Yet another caveat to making conclusions about the economic relevance of weakly identified parameters, is that we usually study an approximation of a DSGE model, typically a log-linearized version, and not the original system. Consequently, parameters which play a very distinct economic role in the model before approximation, may become redundant after that. I will show how to separate the statistical from the the purely model-related aspects of identification in the next section.

3 DSGE Models

3.1 Structural model and reduced form

A DSGE model is summarized by a system of non-linear equations. Currently, most studies involving either simulation or estimation of DSGE models use linear approximations of the original models. That is, the model is first expressed in terms of stationary variables, and then linearized around the steady-state values of these variables. Once

linearized, most DSGE models can be written in the following form:

$$\boldsymbol{\Gamma}_0(\boldsymbol{\theta})\boldsymbol{z}_t = \boldsymbol{\Gamma}_1(\boldsymbol{\theta})\,\mathrm{E}_t\,\boldsymbol{z}_{t+1} + \boldsymbol{\Gamma}_2(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_3(\boldsymbol{\theta})\boldsymbol{u}_t \tag{3.1}$$

where $\boldsymbol{z}_t$ is a $m-$dimensional vector of endogenous and exogenous state variables, and the structural shocks $\boldsymbol{u}_t$ are independent and identically distributed n-dimensional random vectors with $\mathrm{E}\,\boldsymbol{u}_t = \boldsymbol{0}$, $\mathrm{E}\,\boldsymbol{u}_t\boldsymbol{u}_t' = I_n$. The elements of the matrices $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and $\boldsymbol{\Gamma}_3$ are functions of a $k-$dimensional vector of deep parameters $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a point in $\boldsymbol{\Theta} \subset \mathbb{R}^k$. The parameter space $\boldsymbol{\Theta}$ is defined as the set of all theoretically admissible values of $\boldsymbol{\theta}$.

There are several algorithms for solving linear rational expectations models (see for instance Blanchard and Kahn (1980), Anderson and Moore (1985), Klein (2000), Christiano (2002), Sims (2002)). Depending on the value of $\boldsymbol{\theta}$, there may exist zero, one, or many stable solutions. Assuming that a unique solution exists, it can be cast in the following form

$$\boldsymbol{z}_t = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{u}_t \tag{3.2}$$

where the $m \times m$ matrix $\boldsymbol{A}$ and the $m \times n$ matrix $\boldsymbol{B}$ are unique for each value of $\boldsymbol{\theta}$.

The model in (3.2) cannot be taken to the data directly since some of the variables in $\boldsymbol{z}_t$ are not observed. Instead, the solution of the model is expressed in a state space form, with a transition equation given by (3.2), and a measurement equation

$$\boldsymbol{x}_t = \boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{z}_t \tag{3.3}$$

where $\boldsymbol{x}_t$ is a l-dimensional vector of observed state variables, $\boldsymbol{s}$ is a l-dimensional vector, and $\boldsymbol{C}$ is a $l \times m$ matrix. I will use $\boldsymbol{\tau}$ to denote the vector of all parameters characterizing the reduced-form of the model, i.e. $\boldsymbol{\tau} = [\boldsymbol{s}', \mathrm{vec}(\boldsymbol{A})', \mathrm{vec}(\boldsymbol{C})', \mathrm{vech}(\boldsymbol{\Omega})']'$, where $\boldsymbol{\Omega}(\boldsymbol{\theta}) = \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{B}(\boldsymbol{\theta})'$.

3.2 Log-likelihood function and the Information matrix

The log-likelihood function of the data $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\mathrm{T}]$ is derived using the prediction error method whereby a sequence of one-step ahead prediction errors $\boldsymbol{e}_{t|t-1} = \boldsymbol{x}_t - \boldsymbol{s} - \boldsymbol{C}\hat{\boldsymbol{z}}_{t|t-1}$ is constructed by applying the Kalman filter to the obtain one-step ahead forecasts of the state vector $\hat{\boldsymbol{z}}_{t|t-1}$. Assuming that the structural shocks

$\boldsymbol{u}_t$ are jointly Gaussian, it follows that the conditional distribution of $\boldsymbol{e}_{t|t-1}$ is also Gaussian with zero mean and covariance matrix given by $\boldsymbol{S}_{t|t-1} = \boldsymbol{C}\boldsymbol{P}_{t|t-1}\boldsymbol{C}'$, where $\boldsymbol{P}_{t|t-1} = \mathrm{E}\left(\boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1}\right)\left(\boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1}\right)'$ is the conditional covariance matrix of the one-step ahead forecast, and is also obtained from the Kalman filter recursion. This implies that the log-likelihood function of the sample is given by

$$\ell_T(\boldsymbol{\theta}) = \text{const.} - \frac{1}{2}\sum_{t=1}^{\mathrm{T}} \log\det(\boldsymbol{S}_{t|t-1}) - \frac{1}{2}\sum_{t=1}^{\mathrm{T}} \boldsymbol{e}_{t|t-1}'\boldsymbol{S}_{t|t-1}^{-1}\boldsymbol{e}_{t|t-1} \tag{3.4}$$

The ML estimator $\hat{\boldsymbol{\theta}}_T$ is the value of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ which maximizes (3.4). Following the discussion in Section 2, we can measure the precision of $\hat{\boldsymbol{\theta}}_T$ using the inverse of the Fisher information matrix. The next result, due to Klein and Neudecker (2000), provides an explicit expression for the Fisher information matrix for Gaussian models.

Theorem 2. *The expected Fisher information matrix is given by*

$$\begin{aligned}\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \mathrm{E}\left[\left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)' \boldsymbol{S}_t^{-1}\left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)\right] + \\ \frac{1}{2}\sum_{t=1}^{T}\left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_t)}{\partial \boldsymbol{\theta}'}\right)'(\boldsymbol{S}_t \otimes \boldsymbol{S}_t)^{-1}\left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_t)}{\partial \boldsymbol{\theta}'}\right)\end{aligned} \tag{3.5}$$

The asymptotic information matrix, defined as the limit of (3.5), can be computed using the following result (see Ljung (1999))

Theorem 3. *Let $\boldsymbol{S}_\infty = \boldsymbol{C}\boldsymbol{P}_\infty\boldsymbol{C}'$, where $\boldsymbol{P}_\infty = \lim_{\mathrm{T}\to\infty} \boldsymbol{P}_{t|t-1}$ is the steady state covariance matrix of the one-step ahead forecast vector $\hat{\boldsymbol{z}}_{t|t-1}$. Then*

$$\begin{aligned}\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \mathrm{E}\left[\left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)' \boldsymbol{S}_\infty^{-1}\left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)\right] + \\ \frac{1}{2}\left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_\infty)}{\partial \boldsymbol{\theta}'}\right)'(\boldsymbol{S}_\infty \otimes \boldsymbol{S}_\infty)^{-1}\left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_\infty)}{\partial \boldsymbol{\theta}'}\right)\end{aligned} \tag{3.6}$$

To evaluate either (3.5) or (3.6), one needs the derivatives of the reduced-form matrices $\boldsymbol{A}, \boldsymbol{\Omega}$ and $\boldsymbol{C}$ with respect to $\boldsymbol{\theta}$. Explicit formulas for computing these derivatives can be found in Iskrev (2009). Therefore, the full information matrix and all measures of identification strength discussed earlier can be evaluated analytically.

Since the Gaussian assumption is sometimes difficult to justify, it is important to understand the role it plays here. It has two important consequences. First, the likelihood function involves only first and second-order moments of the data. Therefore, for an efficient estimation of the parameters it is sufficient to use the model-implied restrictions on these moments only. Second, the Gaussian assumption facilitates the computation of the optimal weights one should place on the restrictions to achieve efficiency. To see this, note that the ML estimator can be interpreted as a generalized method of moments (GMM) estimator, where the moment function is the score vector. The optimal weighting matrix, given by the inverse of the variance of the score, is not available in closed-form unless Gaussianity is assumed. It can be shown that the inverse of the information matrix (3.6) is smaller than the asymptotic covariance matrix of an efficient GMM estimator for a general distribution. Thus, the confidence intervals computed using the information matrix provide an upper bound on the strength of identification for general statistical models that utilize only first and second moments.

3.3 A priori analysis of identification

From the earlier discussion it follows that the information matrix is all one needs to check the identifiability and to evaluate the strength of identification of the parameters in a model. Therefore, the result presented in the previous section provides us with the necessary tool to study identification in DSGE models. Consider what is involved in the computation of information matrix in (3.5) and (3.6). Taking the linearized structural model in (3.1) together with the assumption about the distribution of $\boldsymbol{u}$ as given, the expected Fisher information matrix depends on: (1) the true value of $\boldsymbol{\theta}$, (2) the set of observed variables in $\boldsymbol{x}$, and, in the case of (3.5), on (3) the number of observations T.

That identification is parameter-dependent is a property of all non-linear models, and implies that $\boldsymbol{\theta}$ may be identifiable in some regions of the parameter space, and unidentified in others. Similarly, identification may be strong in some regions and weak in others. Unless one has an a priori knowledge of the exact true value of $\boldsymbol{\theta}$, one has to study the properties of the information matrix at all theoretically plausible values, i.e. everywhere in $\boldsymbol{\Theta}$. I will return to this point later. The set of observed variables may be considered as a part of the econometric model, and in that sense as given. The practice in the empirical DSGE literature, however, shows that it is to some extent a matter of choice how many and which macroeconomic variables to include in the estimation.

The relevance of this for identification is that some parameters may be well identified if certain endogenous variables are included in $\boldsymbol{x}$, and poorly identified or unidentified if these variables are treated as unobserved. Finally, the value of T enters directly in the computation of $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta})$, and therefore may affect the rank of that matrix. Having more observations may help identify parameters which are otherwise unidentifiable. Naturally, the sample size also matters for the strength of identification of $\boldsymbol{\theta}$. This is seen from (2.7), where the volume of the joint confidence sets is inversely proportional to T.

The effect on identification of having different sets of observables can be investigated by making the appropriate changes in $\boldsymbol{C}$, the matrix which selects the observed among all model variables (see equation (3.3)). Similarly, the effect of having data sets of different sample size is straightforward to find by changing the value of T. Fixing these two dimensions of the statistical model, one can study how identification varies with the value of $\boldsymbol{\theta}$ by evaluating the information matrix at all points in the parameter space. There are two problems with implementing this in practice. First, it is usually impossible to know, before solving the model, for which values of $\boldsymbol{\theta}$ there are either zero or many solutions. Such points are typically deemed as inadmissible, and have to be excluded from $\boldsymbol{\Theta}$. A second problem arises from the fact that there are infinitely many points in $\boldsymbol{\Theta}$, and it is not feasible to evaluate the information matrix at all of them. In view of these difficulties, one approach is to start by specifying a larger set $\boldsymbol{\Theta}'$, such that the parameter space $\boldsymbol{\Theta}$ is a subset of $\boldsymbol{\Theta}'$, and evaluate the information matrix at a large number of randomly drawn points from $\boldsymbol{\Theta}'$, discarding values of $\boldsymbol{\theta}$ that do not imply a unique solution. The set $\boldsymbol{\Theta}'$ may be constructed by specifying a lower and an upper bound for each element of $\boldsymbol{\theta}$. Such bounds are usually easy to come by from the economic meaning of the deep parameters. An alternative approach is to define $\boldsymbol{\Theta}'$ by specifying some univariate probability distribution for each parameter θ_i. The benefit of this approach is that, by choosing the shape and parameters of the distribution, one can achieve a better coverage of the parts of the space that are believed to be more plausible. In practice the choice of distributions may follow the logic of specifying a prior distribution for a Bayesian estimation of DSGE models (see e.g. Del Negro and Schorfheide (2008)).

It should be stressed that the information matrix approach for identification analysis applies only to full information methods. Identification with full information is neces-

sary but not sufficient for identification with limited information. The same applies to the strength of identification - a well identified model may still suffer from weak identification problems if the statistical model is a limited information one. Thus, if a DSGE model is to be estimated with methods, such as impulse response matching, that do not utilize all model-implied restrictions on the distribution of the data, identification should be studied differently. A general rank condition for local identification in DSGE models, which applies to any estimation approach that utilizes only second moments of the data, is developed in Iskrev (2009). Applying that result, one can determine if $\boldsymbol{\theta}$ is identifiable from, for instance, the covariance and first-order autocovariance matrix of some observable endogenous variables. This is useful to know even in a full information setting since identification with limited information is sufficient, though not necessary, for identification with full information methods. Thus, finding that the rank condition in Iskrev (2009) is satisfied for some small number of second moments obviates the need to compute the information matrix, which is generally more computationally expensive. A second necessary condition for identification from Iskrev (2009), that does not depend on statistical model and the distributional assumptions in particular, concerns the invertibility of the mapping from $\boldsymbol{\tau}$ - the reduced-form parameters, to $\boldsymbol{\theta}$. Note that by the chain rule we have:

$$\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} = \frac{\partial \ell_T}{\partial \boldsymbol{\tau}'} \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'} \tag{3.7}$$

and therefore the information matrix may be written as

$$\boldsymbol{\mathcal{I}}_T = \mathrm{E}\left[\left(\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'} \right)' \left\{ \left(\frac{\partial \ell_T}{\partial \boldsymbol{\tau}'} \right)' \left(\frac{\partial \ell_T}{\partial \boldsymbol{\tau}'} \right) \right\} \left(\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'} \right) \right] \tag{3.8}$$

Thus, the Jacobian matrix $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'}$ must have full column rank in order for $\boldsymbol{\mathcal{I}}_T$ and its limit $\boldsymbol{\mathcal{I}}$ to be non-singular. If this condition does not hold some deep parameters are unidentifiable for purely model-related reasons, not because of deficiencies of the statistical model or lack of observations for some model variables. Furthermore, the conditioning of the Jacobian matrix, when it has full column rank, has implications for the strength of identification of $\boldsymbol{\theta}$. From (3.7) it is clear that the two types of weak identification problems discussed in Section 2 may be due to either one of the following two transformations - from $\boldsymbol{\theta}$ to $\boldsymbol{\tau}$, or from $\boldsymbol{\tau}$ to ℓ_T, or to both. The second transformation is partially determined by data limitations - how many and which of the

model variables are included, and the number of observations. The first one depends only on the model, and the Jacobian matrix measures how sensitive are the elements of $\boldsymbol{\tau}$ to those of $\boldsymbol{\theta}$. A very low sensitivity means that relatively large changes in some deep parameters have a very small impact on the value of $\boldsymbol{\tau}$. Consequently, these parameters would be difficult to pin down even if one had data for all endogenous variables in the model, instead of only some of them. In that sense we may say that such deep parameters are poorly identified in the model. To find out what parameters are poorly identified, as well as what features of the model are causing the problem, one may proceed as in Section 2. Specifically, θ_i is weakly identified if either one of the following two conditions holds:

(a) $\boldsymbol{\tau}$ is insensitive to changes in θ_i, i.e.

$$\left\| \frac{\partial \boldsymbol{\tau}}{\partial \theta_i} \theta_i \right\| \approx 0 \tag{3.9}$$

(b) The effect on $\boldsymbol{\tau}$ of changing θ_i can be well approximated by changing other parameters in $\boldsymbol{\theta}$, i.e.

$$\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}} \right) \approx 1 \tag{3.10}$$

If (a) is true, changing θ_i while keeping the other deep parameters fixed has almost no effect on $\boldsymbol{\tau}$. We need to multiply by θ_i in order to make the measure unit-free. If (b) is true, we can alter several elements of $\boldsymbol{\theta}$ simultaneously and have almost the same value of the reduced-form parameters. Note that this is equivalent to having strong collinearity among the columns of the Jacobian matrix $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_i}$. We can quantify the degree of collinearity with the cosine of the angle between the vector $\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}$, and the space spanned by the other columns of $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}}$. I will refer to (3.10) as the coefficient of multiple collinearity. Note that we can similarly compute the degree of linear dependence between $\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}$ and any number of other columns of the Jacobian matrix, and thus quantify the similarity between θ_i and a selected set of other deep parameters. As a special case we have the coefficient of pairwise collinearity, defined as the cosine of the angle between only two columns of $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}}$.

4 Application: Identification in the Smets and Wouters (2007) model

In this section I illustrate the identification analysis tools discussed above using a medium-scale DSGE model estimated in Smets and Wouters (2007) (SW07 henceforth). The model, based on the work of Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005), is an extension of the standard RBC model featuring a number of nominal frictions and real rigidities. I start with an outline of the main features of the model, and then turn to the identification of the parameters.

4.1 The model

The model, based on the work of Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005), is an extension of the standard RBC model featuring a number of nominal frictions and real rigidities. These include: monopolistic competition in goods and labor markets, sticky prices and wages, partial indexation of prices and wages, investments adjustment costs, habit persistence and variable capacity utilization. The endogenous variables in the model, expressed as log-deviations from steady state, are: output (y_t), consumption (c_t), investment (i_t), utilized and installed capital ($k_t^s,\ k_t$), capacity utilization (z_t), rental rate of capital (r_t^k), Tobin's q (q_t), price and wage markup ($\mu_t^p,\ \mu_t^w$), inflation rate(π_t), real wage (w_t), total hours worked (l_t), and nominal interest rate (r_t). The log-linearized equilibrium conditions for these variables are presented in Table 1. The last equation in the table gives the policy rule followed by the central bank, which sets the nominal interest rate in response to inflation and the deviation of output from its potential level. To determine potential output, defined as the level of output that would prevail in the absence of the price and wage mark-up shocks, the set of equations in Table 1 is extended with their flexible price and wage version (see Table 2). The model has seven exogenous shocks. Five of them - to total factor productivity, investment-specific technology, government purchases, risk premium, and monetary policy - follow AR(1) processes; the remaining two shocks - to wage and price markup, follow ARMA$(1,1)$ processes.

The model is estimated using data of seven variables: output growth, consumption growth, investment growth, real wage growth, inflation, hours worked and the nominal

interest rate. Thus, the vector of observables is given by

$$\boldsymbol{x}_t = [y_t - y_{t-1}, c_t - c_{t-1}, i_t - i_{t-1}, w_t - w_{t-1}, \pi_t, l_t, r_t,]' \tag{4.1}$$

and the constant term in the measurement equation (3.3) is given by

$$\boldsymbol{s} = \left[\bar{\gamma}, \bar{\gamma}, \bar{\gamma}, \bar{\gamma}, \bar{\pi}, \bar{l}, \bar{r}\right]' \tag{4.2}$$

where $\bar{\gamma}$ is the growth rate of output, consumption, investment and wages, $\bar{\pi}$ is the steady state rate of inflation, $\bar{l}$ is the steady state level of hours worked and $\bar{r}$ is the steady state nominal interest rate.

The deep parameters of the model are collected in a 41-dimensional vector $\boldsymbol{\theta}$ given by[5]

$$\begin{aligned}\boldsymbol{\theta} = \quad & [\delta, \lambda_w, g_y, \varepsilon_p, \varepsilon_w, \rho_{ga}, \beta, \mu_w, \mu_p, \alpha, \psi, \varphi, \sigma_c, \lambda, \Phi, \iota_w, \xi_w, \iota_p, \xi_p, \sigma_l, \\ & r_\pi, r_{\triangle y}, r_y, \rho, \rho_a, \rho_b, \rho_g, \rho_I, \rho_r, \rho_p, \rho_w, \gamma, \sigma_a, \sigma_b, \sigma_g, \sigma_I, \sigma_r, \sigma_p, \sigma_w, \bar{\pi}, \bar{l}]'\end{aligned} \tag{4.3}$$

4.2 Identification Analysis

Before proceeding with the identification analysis, I need to complete the description of the economic model, by specifying a parameter space, and to add a description of the statistical model, namely, what are the observed variables and what is the sample size. In my baseline specification I follow SW08. Therefore, the parameter space is determined as explained in Section 3.3, with $\boldsymbol{\Theta}'$ given by the prior distribution in SW07 (see Table 3).

The identifiability of the parameters in SW07 was studied in Iskrev (2009), and the following was found: 37 out of the 41 parameters in (4.3) locally identified; the remaining four parameters - ξ_w, ξ_p, ϵ_w andϵ_p, are not separately identifiable in the sense that ξ_w and ϵ_w cannot be distinguished in the linearized model, and neither can ξ_p and ϵ_p. Given these findings, below I study the strength identification of a 39-dimensional vector $\boldsymbol{\theta}$, obtained by removing ϵ_w, and ϵ_p from the list of parameters in (4.3).

[5]Note that by definition $\bar{\gamma} = 100(\gamma - 1)$, and $\bar{r}$ is determined from the values of $\beta,\ \sigma_c,\ \gamma$ and $\bar{\pi}$ from $\bar{r} = 100(\frac{\bar{\pi}\gamma^{\sigma_c}}{\beta} - 1)$.

The results I report below are based on $100,000$ randomly drawn points from $\boldsymbol{\Theta}$. Table 4 provides information about the strength of identification of $\boldsymbol{\theta}$ at three of these points, corresponding to the minimum, median and maximum value of $\bar{r}(\boldsymbol{\theta})$, computed, as in all of following tables, for $\alpha = .1$. In addition to the value of $\bar{r}(\boldsymbol{\theta})$, shown in the last row, the table presents the values of $r(\theta_i)$ for each parameter in the columns labeled with r. The results suggest that there are points in the parameter space where almost all parameters are reasonably well identified. This is seen in the first column, where the values of $r(\theta_i)$ are quite small for all parameters except ρ_{ga}. However, many parameters appear to be poorly identified in large regions of the parameter space. This is seen from the fourth column, corresponding to the median value of $\bar{r}(\boldsymbol{\theta})$, where the value of $r(\theta_i)$ exceeds 1.2 for one third of the parameters. This implies that the length of the 90% confidence interval is more than 240% of the true value of θ_i. Furthermore, the values of $r(\theta_i)$ at the worst identified point, shown in column 7, indicate that several parameters, namely λ_w, β, μ_p, ξ_w, and ρ_p may be extremely poorly identified.

To get a sense of why some parameters are well identified and other are not, I use the factorization of $\text{var}(\hat{\theta}_i)$ discussed in Section 2. In particular, I express $r(\theta_i)$ as follows

$$r(\theta_i) = \left(\frac{\sqrt{c_\alpha}}{|\theta_i| \sqrt{D_i}} \right) \left(\frac{1}{\sqrt{1 - \boldsymbol{\varrho}_i^2}} \right) \tag{4.4}$$

The first term, which I denote with r_1, measures the sensitivity of the likelihood with respect to θ_i, and tells us what $r(\theta_i)$ would be if θ_i was the only free parameter.[6] The second term, denoted with r_2, captures the interdependence among the parameters, and reflects the fact that there is a loss of information for each individual parameter when there are other free parameters.

The columns labeled r_1 and r_2 in Table 4 suggest that most poorly identified parameters suffer from a severe parameter interdependence problem. This is indicated by the very large values found in the r_2 columns of the table. For instance, $r_2 = 7$ implies that $\boldsymbol{\varrho}_i = .99$ in (4.4). The exceptions, for which poor identification is mostly due to a very low sensitivity of the likelihood, are ρ_{ga} and β.

[6]Note that r_2 is bounded from below by 1, and $r(\theta_i) = r_1(\theta_i)$ when θ_i is either independent from the other parameters, or is the only estimated parameter.

The results in Table 4 refer to the distribution of $\bar{r}(\boldsymbol{\theta})$ and are not necessarily fully informative about the individual parameters. In order to assess the identification of each element of $\boldsymbol{\theta}$ separately, we need to consider the distributions of $r(\theta_i)$. Table 5 presents the extreme values and the four quintiles of the distribution of $r(\theta_i)$ for each θ_i. The numbers confirm the conjecture made earlier that many of the parameters are quite poorly identified in most points of the parameter space, and some of them, namely λ_w, β, φ, σ_c, ξ_w, σ_l, r_y, ρ_p and ρ_w, are very poorly identified almost everywhere. Parameters that are relatively well identified are: α, h, Φ, ξ_p, ρ, and the volatility parameters of the exogenous shocks (σ_a, σ_b, σ_g, σ_I, σ_r, σ_p, σ_w).

The values of $r(\theta_i)$ in Table 5 result from the interaction of the two terms, r_1 and r_2, in the decomposition shown in (4.4). Table 6 shows the values of r_1 associated with the numbers for $r(\theta_i)$ in Table 5. It is apparent that almost all parameters would be well identified on their own, i.e. if they were the only ones we estimate. Exceptions are r_{ga}, β, r_y and to a lesser degree δ and ρ_g. This implies that in most cases the identification problems are caused by strong parameter interdependence, leading to large values of r_2. This is confirmed by the numbers in Table 7, where instead of r_2 I present the values of $\boldsymbol{\varrho}_i$, which are bounded between 0 and 1 and are therefore easier to interpret than r_2. For more than half of the parameters the value of $\boldsymbol{\varrho}_i$ exceeds .9 in at least 60% of the points, and for λ_w, σ_c, and ξ_w it is very close to 1 virtually everywhere in the parameter space.

Strong parameter interdependence indicates that many features of the model are difficult to distinguish on the basis of the particular statistical model I have considered. Perhaps using more or different economic variables in the analysis would help us with the identification of some parameters, if their statistical implications for those variables are stronger and more distinct. However, it may also be that some parameters play very similar roles in the structural model, and would be difficult to distinguish with any statistical model. As was explained in Section 3.3, we can answer the second question by studying the Jacobian matrix of the reduced-form parameters $\boldsymbol{\tau}$ with respect to $\boldsymbol{\theta}$. Since $\boldsymbol{\tau}$ fully characterizes the equilibrium dynamics of all endogenous variables, deep parameters that have similar effects on $\boldsymbol{\tau}$ will be hard to identify separately on the basis of any subset of the model variables. Table 8 provides a summary of the distribution of

the multiple collinearity coefficient, defined in (3.10), which measures the angle between $\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}$ and the projection of this vector onto the space spanned by the other columns of $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}}$. Values close to one imply that the two are nearly collinear, and the effect of varying θ_i can be offset to a large extent by changing other deep parameters. The results indicate that the severe interdependence problem observed in Table 7 largely originates in the structural model, and is not caused be limitations in the statistical model.

The results in Table 8 suggest that the SW07 model contains too many features, some of which are nearly redundant given the other features present in the model. Consequently, the parameters that represent such features are difficult to identify and estimate. The most severely affected parameters are: λ_w, σ_c, h, ξ_w, ξ_p, σ_l, r_π, r_y and ρ. We can go a step further and ask which ones among all other 35 parameters are the most important ones in replicating the effect of a given deep parameter on $\boldsymbol{\tau}$. It is reasonable to expect that only a small subset of them - those representing closely related features of the theoretical model, will be important, while the others have only a marginal contribution. Finding the important parameters, therefore, would shed light on the relationships among the model parameters and the features they represent. A simple way to address this question is to compute pairwise collinearity coefficients, that is, angles between two columns of the Jacobian matrix $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}}$, instead of the multiple collinearity coefficients used above. Doing this, we find that, for instance, the degree of similarity between the wage markup parameter λ_w and wage stickiness parameter ξ_w varies between .8211 and .9991. Furthermore, for both of these parameters the pairwise relationship is always stronger between the two than with any other parameter. Thus, we may conclude that the large multiple collinearity coefficients for these two parameters are primarily due to the strong pairwise dependence between them. However, it is not as simple to explain the large values in Table 8 for some of the other parameters listed above. For instance, at the point where the multiple collinearity coefficient for the price stickiness parameter ξ_p is .81, the strongest pairwise collinearity for that parameter is with the wage indexation parameter ι_w, with a coefficient of only .38. Moreover, at other points in the parameter space the strongest pairwise collinearity for ξ_p is not with ι_w but with either ι_p, ρ_p or Φ. Similar lack of a stable and strong two-parameter relationship is observed for most of the other deep parameters. This suggest that the parameter interdependence problem in the SW07 model in most cases involves more than two deep parameters.

The problem of selecting a small set of parameters that are most useful for approximating the effect of a given deep parameter θ_i on $\boldsymbol{\tau}$, is similar to that of choosing a few among many potential predictors in a linear regression model. Various methods for doing this have been developed in the variable selection literature. I use a method called elastic net, which, besides its simplicity, has the advantage of allowing for grouping among the predictors. This means that the method will select a variable even if its marginal contribution is small, if that variable is strongly correlated with another included predictors (see Zou and Hastie (2005) for details). For instance, since λ_w and ξ_w are strongly related in the SW07 model, both will be selected by the elastic net if any one of them is useful for approximating the effect of a third variable on $\boldsymbol{\tau}$. Similarly, additional parameters may be selected to approximate the effect of λ_w on $\boldsymbol{\tau}$, even though their marginal contribution is small once ξ_w is included.

The elastic net procedure was used to select small subsets of deep parameters that are most functionally similar to each one of the nine worst identified parameters listed above. The result can be seen in the last column of Table 9. The table also reports the values of the multiple correlation coefficients one obtains when these subsets of all elements of $\boldsymbol{\theta}$ are used. To facilitate the comparison with Table 8, the coefficients are evaluated at the same points in the parameter space. Therefore, for instance, at the point where the multiple correlation of σ_c with all other deep parameters is .98, the coefficient is .93 if only g_y, h and σ_l are included.

The list of parameters shown in the last column of Table 9 was compiled after an extensive experimentation with the tuning parameters of the elastic net procedure applied to many different points in the parameter space. It should be stressed, however, that it is always possible to improve the approximations by including many additional parameters. Moreover, there are points in the parameter space where some of the parameters included in the list may be replaced by other, and, by doing so, increase quite substantially the value of the collinearity coefficients. Table 9 is only an demonstration of the fact that it is possible to select a robust and yet parsimonious list of parameters, which generally manage to replicate quite well the role of a given deep parameter in the linearized model.

To summarize, the objective in this section was the illustrate the a priori approach to parameter identification analysis in DSGE models, and to demonstrate the types of questions one can investigate using the tools developed in the paper. Although this should not be considered to be a complete and comprehensive study of identification in the SW07 model, several conclusions emerge that appear quite robust. First, most identifiable parameters in the model seem to be very poorly identified. Interestingly, among the worst identified ones is the wage markup parameter λ_w, which Smets and Wouters (2007) assert to be "clearly unidentifiable". Strictly speaking this statement is false, but given how poorly identified λ_w is, it may be classified as practically unidentifiable. However, if we are to apply the same standard to all parameters, also unidentifiable are the following: the discount rate β, the wage stickiness ξ_w, and the response to output gap in the Taylor rule r_y. Second, in most cases the identification weakness stems from a very strong parameter interdependence problem. This means that different parameters have very similar empirical implications, and are thus difficult to distinguish when the model is estimated. Third, in principle the parameter interdependence problem may be alleviated if more variables are included in the analysis. This is not likely in the SW07 model however, since many deep parameters were found to have similar effects on the parameters describing the equilibrium law of motion of the economy. Thus, parameter interdependence would still be a serious problem even if all endogenous variables were observed.

An important point that should be remembered is that identification is a parameter-dependent property of most models. Therefore, in spite of the conclusion that the SW07 model is generally poorly identified, as we saw in Table 4 there exist points in the parameter space where all parameters are very well identified. One may ask how well identified are parameter values that are not only theoretically admissible but also empirically relevant. To find that out, we should examine the information matrix evaluated at the parameter values obtained when the model is estimated for a given set of data. This is done in Table 10 for the posterior mean value of $\boldsymbol{\theta}$ reported in Smets and Wouters (2007). In columns 2 to 4 are shown the mean, the 5-th and 95-th percentile of the posterior distribution; columns 5 and 6 show the lower and upper bound of the 90% confidence intervals; the last three columns show the relative measures of identification strength, defined in (4.4). Although they are conceptually quite different, it is interesting to compare the 90% confidence intervals with their Bayesian counterparts. Doing this

provides some indication of the contribution of the prior knowledge, including the fact that δ, λ_w and g_y are assumed known in Smets and Wouters (2007). On average, the confidence intervals are about 7.5 times wider than the corresponding Bayesian credibility intervals. Particularly striking is the difference for h, α, and β; for them the confidence intervals are 11, 13, and 65 times wider, respectively. Note, however, that for many parameters the confidence intervals cover regions which are not part of the parameter space. As was already discussed in section 2, such a priori restrictions on the parameter values are not accounted for in the measure of weak identification. Overall, the results suggest that the prior mean is very weakly identified from the likelihood only, and that this particular point is among the worst identified points in the parameter space. This can be seen by comparing the values of $r(\theta_i)$ with those reported in Table 5. For all parameters except $\mu_w, \mu_p, \rho_a, \rho_g$ and ρ_p the values of $r(\theta_i)$ are in the upper tails of the distributions.

5 Concluding Remarks

There are two main reasons why we should care about identification in DSGE models. First, using such models for policy analysis hinges upon having reliably estimated parameters. Obtaining such estimates is impossible when identification fails or is very weak. Second, identification failures often have their roots in the underlying model and the economic theory that motivated it. Thus, detecting identification problems and investigating the causes leading to them may provide useful insights to researchers who are not interested in estimation.

This paper develops a new framework for analyzing parameter identification in linearized DSGE models. By following the steps and applying the tools described here, researchers can assess how well identified the model parameters are, and determine the causes for identification problems when they occur. The main advantage of the methodology is that it does not involve simulation or estimation. This makes it suitable for analysis of large and complicated models prior to their empirical evaluation.

An important lesson learnt from the application of the methodology is that the identification properties of a model are strongly dependent on the parameter values, and may change quite dramatically across different regions in the parameter space.

Therefore, it is a mistake to label a model as "weakly identified" or "strongly identified", unless it is determined that either one of this conclusions applies to the large majority of the theoretically plausible parameter values. Unfortunately, the results indicate that the parameters in the Smets and Wouters (2007) model are quite poorly identified in most of the parameter space. The analysis also shows that the identification problems are largely due to the structure of the model, and could not be resolved by extending the set of observed variables. Thus, it may be concluded that this and other similar models are indeed nearly overparameterized, as has been suggested in the literature.

One limitation of the approach in this paper is that it cannot detect certain types of global identification problems. It is possible that some parameters are well identified locally, and yet globally unidentifiable or poorly identified. Such identification failures are less common, but not impossible. Unfortunately, they are very difficult to discover in large and highly non-linear models as those estimated in the DSGE literature.

Table 1: Log-linearized equations of the SW07 model (sticky-price-wage economy)

(1) $y_t = c_y c_t + i_y i_t + r^{kss} k_y z_t + \varepsilon_t^g$

(2) $c_t = \frac{\lambda/\gamma}{1+\lambda/\gamma} c_{t-1} + \frac{1}{1+\lambda/\gamma} \mathrm{E}_t\, c_{t+1} + \frac{w^{ss} l^{ss} (\sigma_c - 1)}{c^{ss} \sigma_c (1+\lambda/\gamma)} (l_t - \mathrm{E}_t\, l_{t+1})$
$- \frac{1-\lambda/\gamma}{(1+\lambda/\gamma)\sigma_c} (r_t - \mathrm{E}_t\, \pi_{t+1}) - \frac{1-\lambda/\gamma}{(1+\lambda\lambda/\gamma)\sigma_c} \varepsilon_t^b$

(3) $i_t = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} i_{t-1} + \frac{\beta\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} \mathrm{E}_t\, i_{t+1} + \frac{1}{\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})} q_t + \varepsilon_t^i$

(4) $q_t = \beta(1-\delta)\gamma^{-\sigma_c} \mathrm{E}_t\, q_{t+1} - r_t + \mathrm{E}_t\, \pi_{t+1} + (1 - \beta(1-\delta)\gamma^{-\sigma_c}) \mathrm{E}_t\, r_{t+1}^k - \varepsilon_t^b$

(5) $y_t = \phi_p(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a)$

(6) $k_t^s = k_{t-1} + z_t$

(7) $z_t = \frac{1-\psi}{\psi} r_t^k$

(8) $k_t = (1-\delta)/\gamma k_{t-1} + (1 - (1-\delta)/\gamma) i_t + (1 - (1-\delta)/\gamma)\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})\varepsilon_t^i$

(9) $\mu_t^p = \alpha(k_t^s - l_t) - w_t + \varepsilon_t^a$

(10) $\pi_t = \frac{\beta\gamma^{(1-\sigma_c)}}{1+\iota_p\beta\gamma^{(1-\sigma_c)}} \mathrm{E}_t\, \pi_{t+1} + \frac{\iota_p}{1+\beta\gamma^{1-\sigma_c}\iota_p} \pi_{t-1} - \frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_p)(1-\xi_p)}{(1+\iota_p\beta\gamma^{(1-\sigma_c)})(1+(\phi_p-1)\varepsilon_p)\xi_p} \mu_t^p + \varepsilon_t^p$

(11) $r_t^k = l_t + w_t - k_t$

(12) $\mu_t^w = w_t - \sigma_l l_t - \frac{1}{1-\lambda/\gamma}(c_t - \lambda/\gamma c_{t-1})$

(13) $w_t = \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} (\mathrm{E}_t\, w_{t+1} + \mathrm{E}_t\, \pi_{t+1}) + \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} (w_{t-1} + \iota_w \pi_{t-1}) - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}} \pi_t$
$- \frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_w)(1-\xi_w)}{(1+\beta\gamma^{(1-\sigma_c)})(1+(\phi_w-1)\varepsilon_w)\xi_w} \mu_t^w + \varepsilon_t^w$

(14) $r_t = \rho r_{t-1} + (1-\rho)(r_\pi \pi_t + r_y(y_t - y_t^*)) + r_{\triangle y}((y_t - y_t^*) - (y_{t-1} - y_{t-1}^*)) + \varepsilon_t^r$

(15) $\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \eta_t^a$

(16) $\varepsilon_t^b = \rho_a \varepsilon_{t-1}^b + \eta_t^b$

(17) $\varepsilon_t^g = \rho_g \varepsilon_{t-1}^a + \rho_{ga} \eta_t^a + \eta_t^g$

(18) $\varepsilon_t^i = \rho_I \varepsilon_{t-1}^I + \eta_t^I$

(19) $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \eta_t^r$

(20) $\varepsilon_t^p = \rho_p \varepsilon_{t-1}^p + \eta_t^p - \mu_p \eta_{t-1}^p$

(21) $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w$

Note: The model variables are: output (y_t), consumption (c_t), investment (i_t), utilized and installed capital ($k_t^s,\ k_t$), capacity utilization (z_t), rental rate of capital (r_t^k), Tobin's q (q_t), price and wage markup ($\mu_t^p,\ \mu_t^w$), inflation rate(π_t), real wage (w_t), total hours worked (l_t), and nominal interest rate (r_t). The shocks are: total factor productivity (ε_t^a), investment-specific technology (ε_t^i), government purchases (ε_t^g), risk premium (ε_t^b), monetary policy (ε_t^r), wage markup (ε_t^w) and price markup (ε_t^p).

Table 2: Log-linearized equations of the SW07 model (flexible-price-wage economy)

(1)	$y_t^* = c_y c_t^* + i_y i_t^* + r^{kss} k_y z_t^* + \varepsilon_t^g$
(2)	$c_t^* = \frac{\lambda/\gamma}{1+\lambda/\gamma} c_{t-1}^* + \frac{1}{1+\lambda/\gamma} \mathrm{E}_t\, c_{t+1}^* + \frac{w^{ss} l^{ss} (\sigma_c - 1)}{c^{ss} \sigma_c (1+\lambda/\gamma)} (l_t^* - \mathrm{E}_t\, l_{t+1}^*)$ $- \frac{1-\lambda/\gamma}{(1+\lambda/\gamma)\sigma_c} r_t^* - \frac{1-\lambda/\gamma}{(1+\lambda\lambda/\gamma)\sigma_c} \varepsilon_t^b$
(3)	$i_t^* = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} i_{t-1}^* + \frac{\beta\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} \mathrm{E}_t\, i_{t+1}^* + \frac{1}{\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})} q_t^* + \varepsilon_t^i$
(4)	$q_t^* = \beta(1-\delta)\gamma^{-\sigma_c} \mathrm{E}_t\, q_{t+1}^* - r_t^* + (1 - \beta(1-\delta)\gamma^{-\sigma_c}) \mathrm{E}_t\, r_{t+1}^{k*} - \varepsilon_t^b$
(5)	$y_t^* = \phi_p(\alpha k_t^{s*} + (1-\alpha) l_t^* + \varepsilon_t^a)$
(6)	$k_t^{s*} = k_{t-1}^* + z_t^*$
(7)	$z_t^* = \frac{1-\psi}{\psi} r_t^{k*}$
(8)	$k_t^* = (1-\delta)/\gamma k_{t-1}^* + (1-(1-\delta)/\gamma) i_t^* + (1-(1-\delta)/\gamma)\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})\varepsilon_t^i$
(9)	$\mu_t^{p*} = \alpha(k_t^{s*} - l_t^*) - w_t^* + \varepsilon_t^a$
(10)	$\mu_t^{p*} = 1$
(11)	$r_t^{k*} = l_t^* + w_t^* - k_t^*$
(12)	$\mu_t^{w*} = -\sigma_l l_t^* - \frac{1}{1-\lambda/\gamma}(c_t^* + \lambda/\gamma c_{t-1}^*)$
(13)	$w_t^* = \mu_t^{w*}$

Note: The model variables are: output (y_t^*), consumption (c_t^*), investment (i_t^*), utilized and installed capital ($k_t^{s*},\ k_t^*$), capacity utilization (z_t^*), rental rate of capital (r_t^{k*}), Tobin's q (q_t^*), price and wage markup ($\mu_t^{p*},\ \mu_t^{w*}$), real wage (w_t^*), and total hours worked (l_t^*).

Table 3: Prior Distribution and posterior mean

	prior					posterior
Parameter	distr.	mean	stdd.	lb	ub	mean
ρ_{ga}	$\mathcal{B}$	0.5000	0.2500	0.0100	2.0000	0.5211
$\bar{l}$	$\mathcal{N}$	0.0000	2.0000	-10.0000	10.0000	0.5416
$\bar{\pi}$	$\mathcal{G}$	0.6250	0.1000	0.1000	2.0000	0.7852
$100(\beta^{-1}-1)$	$\mathcal{G}$	0.2500	0.1000	0.0100	2.0000	0.1661
μ_w	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.8414
μ_p	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.6988
α	$\mathcal{N}$	0.3000	0.0500	0.0100	1.0000	0.1906
ψ	$\mathcal{B}$	0.5000	0.1500	0.0100	1.0000	0.5462
φ	$\mathcal{N}$	4.0000	1.5000	2.0000	15.0000	5.7439
σ_c	$\mathcal{N}$	1.5000	0.3750	0.2500	3.0000	1.3803
λ	$\mathcal{B}$	0.7000	0.1000	0.0010	0.9900	0.7140
Φ	$\mathcal{N}$	1.2500	0.1250	1.0000	3.0000	1.6043
ι_w	$\mathcal{B}$	0.5000	0.1500	0.0100	0.9900	0.5891
ξ_w	$\mathcal{B}$	0.5000	0.1000	0.3000	0.9500	0.7007
ι_p	$\mathcal{B}$	0.5000	0.1500	0.0100	0.9900	0.2437
ξ_p	$\mathcal{B}$	0.5000	0.1000	0.5000	0.9500	0.6503
σ_l	$\mathcal{N}$	2.0000	0.7500	0.2500	10.0000	1.8365
r_π	$\mathcal{N}$	1.5000	0.2500	1.0000	3.0000	2.0454
$r_{\triangle y}$	$\mathcal{N}$	0.1250	0.0500	0.0010	0.5000	0.2237
r_y	$\mathcal{N}$	0.1250	0.0500	0.0010	0.5000	0.0876
ρ	$\mathcal{B}$	0.7500	0.1000	0.5000	0.9750	0.8084
ρ_a	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.9577
ρ_b	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.2167
ρ_g	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.9764
ρ_I	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.7106
ρ_r	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.1513
ρ_p	$\mathcal{B}$	0.5000	0.2000	0.0100	0.9999	0.8914
ρ_w	$\mathcal{B}$	0.5000	0.2000	0.0010	0.9999	0.9682
γ	$\mathcal{N}$	0.4000	0.1000	0.1000	0.8000	0.4310
σ_a	$\mathcal{IG}$	0.1000	2.0000	0.0100	3.0000	0.4595
σ_b	$\mathcal{IG}$	0.1000	2.0000	0.0250	5.0000	0.2405
σ_g	$\mathcal{IG}$	0.1000	2.0000	0.0100	3.0000	0.5289
σ_I	$\mathcal{IG}$	0.1000	2.0000	0.0100	3.0000	0.4532
σ_r	$\mathcal{IG}$	0.1000	2.0000	0.0100	3.0000	0.2453
σ_p	$\mathcal{IG}$	0.1000	2.0000	0.0100	3.0000	0.1399
σ_w	$\mathcal{IG}$	0.1000	2.0000	0.0100	3.0000	0.2443
δ	$\mathcal{B}$	0.0250	0.0050	0.0100	0.4000	0.0250
λ_w	$\mathcal{N}$	1.5000	0.2500	1.0000	2.0000	1.5000
g_y	$\mathcal{N}$	0.1800	0.0500	0.1500	0.2500	0.1800

Note: $\mathcal{N}$ is Normal distribution, $\mathcal{B}$ is Beta-distribution, $\mathcal{G}$ is Gamma distribution, $\mathcal{IG}$ is Inverse Gamma distribution.

Table 4: Overall identification of $\boldsymbol{\theta}$

	Best			Median			Worst		
Parameter	r	r_1	r_2	r	r_1	r_2	r	r_1	r_2
δ	0.0044	0.0006	6.8	1.71	0.36	4.76	5.55	1.60	3.5
λ_w	0.0067	0.0000	164.3	1.96	0.08	24.72	711.86	0.05	14519.9
g_y	0.0080	0.0007	11.1	0.93	0.17	5.42	2.58	0.49	5.3
ρ_{ga}	1.8911	0.1665	11.4	3.36	1.20	2.81	1.20	0.39	3.1
$100(\beta^{-1}-1)$	0.0473	0.0024	19.7	15.60	2.15	7.27	74.47	12.77	5.8
μ_w	0.0539	0.0001	554.7	3.55	0.28	12.55	1.63	0.11	14.9
μ_p	0.0986	0.0110	9.0	2.20	0.15	14.64	69.77	0.26	266.2
α	0.0020	0.0002	11.6	0.70	0.12	5.83	2.93	0.37	8.0
ψ	0.0033	0.0002	19.1	0.96	0.15	6.36	2.30	0.49	4.7
φ	0.0080	0.0002	46.6	0.80	0.05	15.75	2.57	0.17	15.4
σ_c	0.0094	0.0001	73.2	0.59	0.05	12.43	2.62	0.10	26.1
h	0.0006	0.0000	33.5	0.31	0.02	12.35	0.45	0.05	8.7
Φ	0.0016	0.0001	17.3	0.25	0.03	7.52	0.28	0.05	5.7
ι_w	0.0111	0.0004	27.0	0.46	0.12	3.99	0.98	0.26	3.8
ξ_w	0.0101	0.0000	240.4	1.25	0.05	24.25	623.42	0.04	14520.1
ι_p	0.0202	0.0012	17.5	3.34	0.33	10.28	1.82	0.36	5.0
ξ_p	0.0095	0.0009	10.8	1.57	0.10	15.50	0.83	0.17	4.9
σ_l	0.0146	0.0002	81.4	1.24	0.17	7.42	2.48	0.39	6.3
r_π	0.0020	0.0002	10.8	0.76	0.02	36.18	2.32	0.20	11.7
$r_{\triangle y}$	0.0046	0.0008	5.5	0.76	0.06	13.43	1.29	0.23	5.7
r_y	0.0080	0.0007	11.0	1.98	0.28	7.16	4.05	0.79	5.1
ρ	0.0013	0.0001	10.3	0.25	0.01	28.35	0.59	0.05	12.4
ρ_a	0.0192	0.0015	12.8	0.40	0.13	3.17	3.32	1.21	2.8
ρ_b	0.0120	0.0002	54.7	0.74	0.16	4.53	1.48	0.21	7.1
ρ_g	0.0198	0.0026	7.5	0.67	0.25	2.67	0.64	0.23	2.8
ρ_I	0.0037	0.0005	7.2	0.90	0.27	3.29	2.14	0.58	3.7
ρ_r	0.0080	0.0021	3.8	0.03	0.01	4.53	1.97	0.51	3.9
ρ_p	0.0235	0.0017	13.5	0.17	0.03	6.47	67.79	0.26	265.4
ρ_w	0.0098	0.0009	11.6	1.71	0.13	12.82	3.14	0.21	15.3
σ_a	0.3397	0.0340	10.0	0.43	0.15	2.87	0.52	0.15	3.5
σ_b	0.0001	0.0000	548.6	0.57	0.15	3.84	0.63	0.13	4.7
σ_g	0.3769	0.0637	5.9	0.41	0.15	2.79	0.58	0.15	3.9
σ_I	0.2937	0.0187	15.7	0.45	0.15	3.04	0.51	0.15	3.4
σ_r	0.3884	0.1324	2.9	0.67	0.15	4.52	0.42	0.15	2.8
σ_p	0.3910	0.1483	2.6	0.97	0.15	6.54	0.59	0.15	4.0
σ_w	0.3366	0.0182	18.5	0.52	0.16	3.38	0.49	0.14	3.5
$\bar{r}(\boldsymbol{\theta})$		0.016			0.82			2.53	

Note: The table shows the results on the identification of $\boldsymbol{\theta}$ at the points in the parameters space with the minimum, the median and the maximum value of $\bar{r}(\boldsymbol{\theta}) = \left(\prod_{i=1}^{k} r(\theta_i)\right)^{1/k}$. The identification strength of each parameter θ_i is measured by $r(\theta_i) = \frac{\sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)}}{\theta_i}$, and is decomposed into the product of $r_1(\theta)$ and $r_2(\theta)$, which capture the likelihood sensitivity and the parameter interdependence aspect of identification, respectively. The results are based on $100,000$ draws from $\boldsymbol{\Theta}$ and apply to the baseline statistical model with $c_t, l_t, \pi_t, w_t, i_t, r_t$ and y_t as observed variables, and $T = 156$. c_α is the $(1-\alpha)$ quantile of the $\chi^2(36)$ distribution, and α is set to .1.

Table 5: Quantiles of the distribution of $r(\theta_i)$

Parameter	0	.2	.4	.6	.8	1
δ	0.0033	0.42	0.73	1.09	1.68	8.2
λ_w	0.0067	0.97	1.58	2.49	4.84	22433.2
g_y	0.0080	0.44	0.69	0.97	1.40	12.3
ρ_{ga}	0.0008	0.56	0.99	1.71	3.50	792.0
$100(\beta^{-1}-1)$	0.0366	4.13	6.53	9.35	14.21	169.3
μ_w	0.0077	0.42	1.37	2.73	5.83	68799.2
μ_p	0.0054	0.40	1.72	3.49	7.57	25032.2
α	0.0020	0.20	0.31	0.45	0.66	3.6
ψ	0.0031	0.33	0.52	0.73	1.04	5.9
φ	0.0080	0.69	0.95	1.21	1.54	4.0
σ_c	0.0094	0.62	0.88	1.14	1.47	3.7
h	0.0003	0.09	0.16	0.25	0.37	1.8
Φ	0.0016	0.10	0.15	0.20	0.28	1.3
ι_w	0.0087	0.48	0.75	1.12	1.81	44.8
ξ_w	0.0101	0.97	1.55	2.39	4.57	17304.5
ι_p	0.0050	0.32	0.52	0.80	1.37	22.0
ξ_p	0.0032	0.20	0.28	0.37	0.51	4.5
σ_l	0.0088	0.71	1.25	1.95	3.11	32.5
r_π	0.0010	0.38	0.63	0.92	1.39	10.2
$r_{\triangle y}$	0.0046	0.48	0.76	1.13	1.82	195.7
r_y	0.0080	1.56	2.48	3.66	5.81	859.1
ρ	0.0007	0.12	0.20	0.28	0.38	1.4
ρ_a	0.0101	0.39	0.58	0.82	1.26	25.1
ρ_b	0.0011	0.23	0.41	0.62	0.97	28.1
ρ_g	0.0024	0.40	0.60	0.85	1.30	31.1
ρ_I	0.0015	0.28	0.41	0.57	0.84	18.9
ρ_r	0.0002	0.19	0.38	0.63	1.02	19.0
ρ_p	0.0092	0.75	1.52	2.92	6.76	25021.2
ρ_w	0.0098	0.63	1.24	2.31	5.38	68800.7
σ_a	0.0048	0.39	0.40	0.42	0.45	1.1
σ_b	0.0000	0.38	0.48	0.53	0.59	1.2
σ_g	0.0142	0.40	0.41	0.42	0.46	0.9
σ_I	0.0092	0.44	0.49	0.52	0.54	1.3
σ_r	0.0012	0.39	0.41	0.46	0.55	2.4
σ_p	0.1309	0.44	0.48	0.52	0.57	3.7
σ_w	0.0001	0.44	0.48	0.52	0.63	31.3

Note: Each row of the table shows the extreme values and the four quintiles of the distribution of $r(\theta_i) = \frac{\sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)}}{\theta_i}$, which measures the identification strength of θ_i. The results are based on $100,000$ draws from $\boldsymbol{\Theta}$ and apply to the baseline statistical model with $c_t, l_t, \pi_t, w_t, i_t, r_t$ and y_t as observed variables, and $T = 156$. c_α is the $(1-\alpha)$ quantile of the $\chi^2(36)$ distribution, and α is set to .1.

Table 6: Sensitivity component in the distribution of $r(\theta_i)$

	Quantiles of $r(\theta_i)$					
Parameter	0	.2	.4	.6	.8	1
δ	0.00003048	0.20	0.39	0.64	1.04	5.44
λ_w	0.00000966	0.02	0.04	0.07	0.13	0.90
g_y	0.00003680	0.10	0.19	0.29	0.48	5.13
ρ_{ga}	0.00000167	0.46	0.87	1.51	3.01	560.14
$100(\beta^{-1}-1)$	0.00026126	0.78	1.53	2.54	4.31	65.06
μ_w	0.00000284	0.07	0.21	0.39	0.68	25.19
μ_p	0.00000052	0.09	0.26	0.47	0.82	32.59
α	0.00001303	0.05	0.08	0.13	0.21	1.08
ψ	0.00000898	0.08	0.14	0.23	0.36	3.87
φ	0.00002746	0.09	0.16	0.26	0.42	1.76
σ_c	0.00000701	0.04	0.07	0.12	0.17	0.47
h	0.00000143	0.01	0.03	0.05	0.10	0.68
Φ	0.00000311	0.02	0.03	0.05	0.07	0.40
ι_w	0.00004052	0.17	0.32	0.54	0.96	21.27
ξ_w	0.00001759	0.02	0.04	0.08	0.14	1.05
ι_p	0.00001052	0.14	0.27	0.43	0.66	5.09
ξ_p	0.00000459	0.03	0.07	0.12	0.20	0.75
σ_l	0.00006790	0.08	0.17	0.31	0.59	7.22
r_π	0.00001272	0.06	0.12	0.19	0.28	1.05
$r_{\triangle y}$	0.00007077	0.13	0.25	0.42	0.72	93.45
r_y	0.00004711	0.35	0.68	1.09	1.77	306.35
ρ	0.00000285	0.02	0.03	0.05	0.08	0.39
ρ_a	0.00005061	0.29	0.48	0.71	1.13	24.36
ρ_b	0.00000248	0.04	0.12	0.24	0.47	19.25
ρ_g	0.00007091	0.31	0.50	0.74	1.17	31.06
ρ_I	0.00000583	0.16	0.26	0.38	0.59	15.19
ρ_r	0.00001757	0.07	0.18	0.35	0.65	16.10
ρ_p	0.00000217	0.14	0.26	0.39	0.64	41.66
ρ_w	0.00000947	0.12	0.21	0.32	0.55	23.45
σ_a	0.00000116	0.39	0.39	0.39	0.39	0.39
σ_b	0.00000004	0.15	0.34	0.38	0.39	0.41
σ_g	0.00000135	0.38	0.39	0.39	0.39	0.39
σ_I	0.00004477	0.38	0.39	0.39	0.39	0.40
σ_r	0.00000054	0.34	0.38	0.39	0.39	0.39
σ_p	0.00002382	0.39	0.39	0.39	0.39	0.40
σ_w	0.00000056	0.35	0.38	0.39	0.40	0.50

Note: Each row of the table shows the value of $r_1(\theta) = \frac{\sqrt{c_\alpha / D_i}}{\theta_i}$ at the extreme values and the four quintiles of the distribution of $r(\theta_i)$. Large values of $r_1(\theta)$ imply low sensitivity of the log-likelihood function with respect to θ_i. The results are based on $100,000$ draws from $\boldsymbol{\Theta}$ and apply to the baseline statistical model with $c_t, l_t, \pi_t, w_t, i_t, r_t$ and y_t as observed variables, and $T = 156$. c_α is the $(1-\alpha)$ quantile of the $\chi^2(36)$ distribution, and α is set to .1. D_i is the variance of the $i-$th component of the score vector.

Table 7: Parameter interdependence component in the distribution of $r(\theta_i)$

	Quantiles of $r(\theta_i)$					
Parameter	0	.2	.4	.6	.8	1
δ	0.35	0.734	0.7910	0.8403	0.89460	0.99999987241
λ_w	0.92	0.998	0.9995	0.9998	0.99996	0.99999999996
g_y	0.34	0.871	0.9354	0.9678	0.98705	0.99999985796
ρ_{ga}	0.01	0.188	0.2912	0.4147	0.60224	0.99999999979
$100(\beta^{-1}-1)$	0.49	0.924	0.9570	0.9758	0.98826	0.99999998070
μ_w	0.56	0.957	0.9833	0.9936	0.99875	0.99999999998
μ_p	0.53	0.963	0.9836	0.9932	0.99837	0.99999999998
α	0.53	0.913	0.9470	0.9672	0.98253	0.99999994395
ψ	0.21	0.885	0.9364	0.9638	0.98249	0.99999995301
φ	0.66	0.938	0.9712	0.9868	0.99475	0.99999989013
σ_c	0.89	0.990	0.9946	0.9971	0.99871	0.99999997740
h	0.64	0.950	0.9718	0.9834	0.99153	0.99999977950
Φ	0.39	0.926	0.9612	0.9794	0.99108	0.99999999829
ι_w	0.20	0.709	0.8381	0.9147	0.96442	0.99999998284
ξ_w	0.87	0.998	0.9994	0.9998	0.99996	0.99999999996
ι_p	0.45	0.763	0.8247	0.8834	0.94352	0.99999999762
ξ_p	0.57	0.887	0.9383	0.9701	0.98952	0.99999999836
σ_l	0.48	0.966	0.9833	0.9920	0.99679	0.99999998947
r_π	0.60	0.950	0.9742	0.9867	0.99458	0.99999999018
$r_{\triangle y}$	0.48	0.842	0.9067	0.9494	0.97942	0.99999996394
r_y	0.45	0.895	0.9455	0.9717	0.98859	0.99999999802
ρ	0.69	0.960	0.9786	0.9888	0.99532	0.99999998021
ρ_a	0.02	0.236	0.3618	0.5158	0.72441	0.99999979811
ρ_b	0.46	0.842	0.9116	0.9583	0.98593	0.99999991612
ρ_g	0.03	0.233	0.3569	0.4999	0.69232	0.99999939471
ρ_I	0.32	0.655	0.7059	0.7562	0.82993	0.99999998246
ρ_r	0.19	0.645	0.7859	0.8893	0.95556	0.99999771107
ρ_p	0.14	0.962	0.9839	0.9932	0.99835	0.99999999998
ρ_w	0.15	0.960	0.9827	0.9926	0.99816	0.99999999998
σ_a	0.01	0.204	0.3124	0.4219	0.55873	0.99999999998
σ_b	0.19	0.677	0.7402	0.7964	0.90617	0.99999999998
σ_g	0.01	0.237	0.3461	0.4591	0.60713	0.99999999979
σ_I	0.05	0.546	0.6308	0.6736	0.71783	0.99999999325
σ_r	0.01	0.288	0.4874	0.6704	0.82687	0.99999999998
σ_p	0.04	0.525	0.6187	0.6808	0.75501	0.99999998344
σ_w	0.08	0.542	0.6268	0.7080	0.82918	0.99999999951

Note: Each row of the table shows the value of $\boldsymbol{\varrho}_i$ at the extreme values and the four quintiles of the distribution of $r(\theta_i)$. Values close to 1 indicate strong parameter interdependence problem and near-redundancy of θ_i. The results are based on 100,000 draws from $\boldsymbol{\Theta}$ and apply to the baseline statistical model with $c_t, l_t, \pi_t, w_t, i_t, r_t$ and y_t as observed variables.

Table 8: Parameter interdependence in the structural model

	Quantiles of $\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}}\right)$					
Parameter	0	.2	.4	.6	.8	1
δ	0.57	0.811	0.8764	0.9194	0.95715	0.9983464283
λ_w	0.96	0.995	0.9983	0.9996	0.99995	0.9999999882
g_y	0.76	0.866	0.9022	0.9237	0.96920	0.9999700745
ρ_{ga}	0.14	0.737	0.9121	0.9607	0.98460	0.9990421323
$100(\beta^{-1}-1)$	0.83	0.918	0.9405	0.9612	0.98226	0.9999994432
μ_w	0.07	0.326	0.4215	0.5086	0.60781	0.9671890433
μ_p	0.22	0.483	0.6537	0.7343	0.80572	0.9972787213
α	0.86	0.924	0.9411	0.9598	0.98229	0.9999977067
ψ	0.13	0.223	0.2793	0.3623	0.46999	0.9432934213
φ	0.59	0.800	0.8596	0.9176	0.96805	0.9999995522
σ_c	0.98	0.996	0.9981	0.9993	0.99987	0.9999999997
h	0.94	0.992	0.9974	0.9989	0.99983	0.9999999997
Φ	0.76	0.926	0.9487	0.9604	0.97855	0.9984902894
ι_w	0.67	0.816	0.8711	0.9188	0.97297	0.9999998643
ξ_w	0.95	0.995	0.9984	0.9996	0.99995	0.9999999859
ι_p	0.69	0.847	0.8973	0.9415	0.98025	0.9999993589
ξ_p	0.81	0.960	0.9722	0.9840	0.99164	0.9999995108
σ_l	0.90	0.963	0.9726	0.9864	0.99373	0.9999994155
r_π	0.74	0.971	0.9847	0.9946	0.99908	0.9999999812
$r_{\triangle y}$	0.27	0.651	0.7962	0.8780	0.96169	0.9999943718
r_y	0.76	0.950	0.9748	0.9878	0.99686	0.9999998917
ρ	0.73	0.931	0.9723	0.9859	0.99691	0.9999999085
ρ_a	0.04	0.557	0.6562	0.7099	0.78225	0.9999961581
ρ_b	0.12	0.889	0.9571	0.9822	0.99603	0.9999999575
ρ_g	0.05	0.414	0.5344	0.6244	0.71443	0.9999958296
ρ_I	0.06	0.239	0.3833	0.5256	0.68363	0.9981250333
ρ_r	0.03	0.499	0.6801	0.8597	0.94317	0.9999261878
ρ_p	0.19	0.604	0.7817	0.8933	0.95010	0.9987404769
ρ_w	0.11	0.373	0.5088	0.6665	0.84165	0.9995829154
σ_a	0.10	0.607	0.8258	0.9048	0.95859	0.9965723603
σ_b	0.54	0.823	0.9045	0.9681	0.99181	0.9999984997
σ_g	0.17	0.622	0.8169	0.9107	0.96345	0.9964866185
σ_I	0.13	0.265	0.3253	0.4126	0.52482	0.9931261729
σ_r	0.54	0.755	0.8272	0.8723	0.93071	0.9994093412
σ_p	0.14	0.367	0.4552	0.5211	0.60895	0.9596850368
σ_w	0.07	0.199	0.2514	0.3118	0.39339	0.9778976469

Note: Each row of the table shows the extreme values and the four quintiles of the distribution of $\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}}\right)$, which measures the degree of parameter interdependence in the structural model. Values close to 1 indicate that the effect of θ_i on $\boldsymbol{\tau}$ may be approximated very well by that the other deep parameters. The results are based on $100,000$ draws from $\boldsymbol{\Theta}$.

Table 9: Worst identified parameters

worst identified	Quantiles of $\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}}\right)$						functionally similar
parameters	0	.2	.4	.6	.8	1	parameters
λ_w	0.860	0.952	0.993	0.992	0.987	0.996923	σ_c, ξ_w, σ_l
σ_c	0.928	0.982	0.986	0.970	0.994	0.999480	g_y, h, σ_l
h	0.809	0.932	0.960	0.944	0.984	0.999410	σ_c, ρ_b
ξ_w	0.823	0.928	0.981	0.993	0.950	0.937484	λ_w, σ_l
ξ_p	0.448	0.910	0.924	0.969	0.973	0.954205	μ_p, φ, Φ, ι_p, ρ_p
σ_l	0.702	0.860	0.855	0.898	0.952	0.998846	λ_w, g_y, α, σ_c,σ_l
r_π	0.494	0.920	0.964	0.960	0.993	0.999995	$r_{\triangle y}$, r_y, ρ, ρ_b
r_y	0.670	0.879	0.935	0.973	0.989	0.999982	r_π, $r_{\triangle y}$, ρ, ρ_b, σ_b, σ_r
ρ	0.688	0.900	0.930	0.954	0.994	0.999918	r_π, $r_{\triangle y}$, r_y, ρ_r

Note: This table shows to what extent the results in Table 8 can be explained, for the worst identified parameters, using only a small set of other functionally similar parameters. Each row reports the multiple collinearity coefficients for the parameter in the first column with those in column 8. The multiple collinearity coefficient for θ_i is computed as the cosine of the angle between the vector $\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}$ and the space spanned by the columns of $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}}$ corresponding to the parameters in column 8. Values close to 1 indicate that the effect of θ_i on $\boldsymbol{\tau}$ may be approximated very well by that of the other deep parameters.

Table 10: Parameter identification at the posterior mean

	posterior distribution			90% CI		ratios		
Parameter	mean	5 percent.	95 percent.	lb	ub	r	r_1	r_2
δ^*	0.03	0.03	0.03	-0.04	0.09	2.51	1.10	2.28
λ_w^*	1.50	1.50	1.50	-3.31	6.31	3.21	0.31	10.41
g_y^*	0.18	0.18	0.18	-0.29	0.65	2.63	1.57	1.67
ρ_{ga}	0.52	0.37	0.66	-0.22	1.26	1.43	1.22	1.17
$100(\beta^{-1}-1)$	0.17	0.07	0.26	-6.01	6.34	37.18	10.12	3.68
μ_w	0.84	0.75	0.93	0.52	1.16	0.38	0.11	3.34
μ_p	0.70	0.54	0.85	-0.13	1.53	1.18	0.21	5.66
α	0.19	0.16	0.21	-0.14	0.53	1.76	0.41	4.30
ψ	0.55	0.36	0.72	-0.51	1.60	1.93	0.96	2.02
φ	5.74	3.97	7.42	-7.74	19.22	2.35	1.12	2.10
σ_c	1.38	1.16	1.59	-0.43	3.19	1.31	0.29	4.50
h	0.71	0.64	0.78	0.33	1.10	0.54	0.19	2.86
Φ	1.60	1.48	1.73	0.84	2.37	0.48	0.25	1.89
ι_w	0.59	0.38	0.78	-0.56	1.74	1.96	1.46	1.34
ξ_w	0.70	0.60	0.81	-0.43	1.83	1.62	0.15	10.44
ι_p	0.24	0.10	0.38	-0.49	0.98	3.02	1.40	2.17
ξ_p	0.65	0.56	0.74	0.29	1.01	0.55	0.22	2.53
σ_l	1.84	0.91	2.78	-4.59	8.27	3.50	1.40	2.51
r_π	2.05	1.74	2.33	-0.66	4.75	1.32	0.37	3.56
$r_{\triangle y}$	0.22	0.18	0.27	-0.08	0.53	1.36	0.70	1.93
r_y	0.09	0.05	0.12	-0.17	0.34	2.93	0.93	3.16
ρ	0.81	0.77	0.85	0.51	1.10	0.36	0.13	2.84
ρ_a	0.96	0.94	0.97	0.86	1.05	0.10	0.06	1.70
ρ_b	0.22	0.07	0.36	-0.16	0.60	1.75	1.10	1.60
ρ_g	0.98	0.96	0.99	0.91	1.04	0.06	0.04	1.53
ρ_I	0.71	0.61	0.80	0.40	1.02	0.43	0.23	1.86
ρ_r	0.15	0.04	0.24	-0.24	0.54	2.59	2.07	1.25
ρ_p	0.89	0.80	0.96	0.54	1.25	0.40	0.09	4.22
ρ_w	0.97	0.94	0.99	0.87	1.07	0.10	0.05	2.15
σ_a	0.46	0.41	0.50	0.24	0.68	0.48	0.39	1.24
σ_b	0.24	0.19	0.27	0.10	0.38	0.58	0.39	1.50
σ_g	0.53	0.48	0.58	0.27	0.79	0.48	0.39	1.25
σ_I	0.45	0.37	0.53	0.18	0.73	0.61	0.39	1.57
σ_r	0.25	0.22	0.27	0.13	0.36	0.47	0.39	1.20
σ_p	0.14	0.11	0.16	0.02	0.26	0.87	0.39	2.22
σ_w	0.24	0.20	0.28	0.09	0.40	0.63	0.39	1.61

Note: The posterior mean and percentiles are those reported in Smets and Wouters (2007). The lower (lb) and upper bound (ub) of the individual confidence intervals are obtained by projecting the 90% confidence ellipsoid onto the parameter axes. $r(\theta_i)$ is a unit-free measure of identification strength, defined as the half-length of the confidence interval, divided by the true parameter value. $r(\theta_i)$ is expressed as the product of $r_1(\theta)$ and $r_2(\theta)$, which capture the likelihood sensitivity and the parameter interdependence aspect of identification, respectively.

*These parameters are fixed in Smets and Wouters (2007)

References

Anderson, G., and G. Moore (1985): "A linear algebraic procedure for solving linear perfect foresight models," *Economics Letters*, 17(3), 247–252, available at http://ideas.repec.org/a/eee/ecolet/v17y1985i3p247-252.html.

Bekker, P. A., and D. S. G. Pollock (1986): "Identification of linear stochastic models with covariance restrictions," *Journal of Econometrics*, 31(2), 179–208, available at http://ideas.repec.org/a/eee/econom/v31y1986i2p179-208.html.

Beyer, A., and R. E. A. Farmer (2004): "On the indeterminacy of New-Keynesian economics," Working Paper Series 323, European Central Bank, available at http://ideas.repec.org/p/ecb/ecbwps/20040323.html.

Blanchard, O. J., and C. M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–11, available at http://ideas.repec.org/a/ecm/emetrp/v48y1980i5p1305-11.html.

Bowden, R. J. (1973): "The Theory of Parametric Identification," *Econometrica*, 41(6), 1069–74.

Canova, F., and L. Sala (2009): "Back to square one: identification issues in DSGE models," *Journal of Monetary Economics, forthcoming*.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2009): "New Keynesian Models: Not Yet Useful for Policy Analysis," *American Economic Journal: Macroeconomics*, 1(1), 242–66.

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45, available at http://ideas.repec.org/a/ucp/jpolec/v113y2005i1p1-45.html.

Christiano, L. J. (2002): "Solving dynamic equilibrium models by a method of undetermined coefficients," *Computational Economics*, 20(1-2).

Cochrane, J. H. (2007): "Identification with Taylor Rules: A Critical Review," NBER Working Papers 13410, National Bureau of Economic Research, Inc.

Del Negro, M., and F. Schorfheide (2008): "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," *Journal of Monetary Economics*, 55(7), 1191 – 1208.

Eichenbaum, M., and J. Fisher (2007): "Estimating the Frequency of Reoptimisation in Calvo-style Models," *Journal of Monetary Economics, forthcoming*.

Greenwood, J., Z. Hercowitz, and P. Krusell (2000): "The role of investment-specific technological change in the business cycle," *European Economic Review*, 44(1), 91–115, available at http://ideas.repec.org/a/eee/eecrev/v44y2000i1p91-115.html.

Guerron, P. A. (2007): "What You Match Does Matter: The Effects of Data on DSGE Estimation," Working Paper Series 012, North Carolina State University, Department of Economics.

Iskrev, N. (2009): "Local Identification in DSGE Models," unpublished manuscript.

Kim, J. (2003): "Functional equivalence between intertemporal and multisectoral investment adjustment costs," *Journal of Economic Dynamics and Control*, 27(4), 533–549.

Klein, A., and H. Neudecker (2000): "A direct derivation of the exact Fisher information matrix of Gaussian vector state space models," *Linear Algebra and its Applications*, 321(1-3), 233–238.

Klein, P. (2000): "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423, available at http://ideas.repec.org/a/eee/dyncon/v24y2000i10p1405-1423.html.

Koopmans, T. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17, 125–144.

Ljung, L. (1999): *System Identification: Theory for the User (2nd Edition)*. Prentice Hall, Upper Saddle River, NJ, USA.

Manski, C. F. (1995): *Identification Problems in Social Sciences*. Harvard University Press.

Pesaran, M. H. (1989): *The Limits to Rational Expectations*. Basil Blackwell, Oxford.

Rao, C. R. (1973): *Linear Statistical Inference and its Applications*. Wiley, New York.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–91, available at http://ideas.repec.org/a/ecm/emetrp/v39y1971i3p577-91.html.

Sargent, T. J. (1976): "The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics," *Journal of Political Economy*, 84(3), 631–40, available at http://ideas.repec.org/a/ucp/jpolec/v84y1976i3p631-40.html.

Shapiro, A., and M. Browne (1983): "On the investigation of local identifiability: A counterexample," *Psychometrika*, 48(2), 303–304.

Sims, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2), 1–20, available at http://ideas.repec.org/a/kap/compec/v20y2002i1-2p1-20.html.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175, available at http://ideas.repec.org/a/tpr/jeurec/v1y2003i5p1123-1175.html.

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

Tucker, L., L. Cooper, and W. Meredith (1972): "Obtaining squared multiple correlations from a correlation matrix which may be singular," *Psychometrika*, 37(2), 143–148.

Zou, H., and T. Hastie (2005): "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.