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

Nikolay Iskrev
Banco de Portugal

This draft: March 11, 2010

PRELIMINARY AND INCOMPLETE. COMMENTS WELCOME

Abstract

The strength of identification in structural models is a reflection of the empirical relevance of the model features represented by the parameters. Weak identification arises when some parameters are nearly irrelevant or nearly redundant with respect to the aspects of reality the model is intended to explain. The strength of identification is therefore not only a crucial requirement for the reliable estimation of models, but also has important implications for model development. This paper proposes a new framework for evaluating the strength of identification in linearized dynamic stochastic general equilibrium (DSGE) models prior to their estimation. In a parametric setting, the empirical implications of a model are contained in the likelihood function, which, for DSGE models, is completely characterized by the underlying structural model. I show how to use standard asymptotic theory to evaluate the theoretical properties of likelihood-based estimators at any point in the parameter space associated with the model. Furthermore, in addition to assessing the informativeness of the likelihood as a whole, I show how to determine which particular features of the data, such as moments of a given variable or a set of variables, are most important for the identification of a given parameter. The methodology is illustrated using a medium-scale business cycle model.

Keywords: DSGE models, Weak identification, Information matrix, Cramér-Rao lower bound

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)).

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".

Most of the existing research on identification in DSGE models follows the econometric literature where weak identification is treated as a sampling problem, i.e. as an issue within the realm of statistical inference (see e.g. Stock and Yogo (2005) and the survey in Andrews and Stock (2005)). For this reason the effort has been devoted to either devising tests for detecting weak identification (Inoue and Rossi (2008)), or to developing methods for inference that are robust to identification problems (Guerron-Quintana, Inoue, and Kilian (2009)). This paper proposes an alternative approach, based on the premise that identification in DSGE models can be treated as a property of the underlying economic model, and as such may be studied without reference to particular sample of data. This approach is in the spirit of the classical literature on identification, and is based on the fact that DSGE models provide a complete characterization of the data generating process. This is in contrast to the typical situation in structural econometrics where the mapping from the economic model to the data is known only partially. For instance, the degree of correlation between instruments and endogenous variables in the simple linear instrumental variables model depends on nuisance parameters which, in the absence of a fully-articulated economic model, have no structural interpretation. In a general equilibrium setting, all reduced-form parameters become functions of the structural parameters, and one can investigate how the instruments' strength is determined by the properties of the underlying model.

In the context of DSGE models, important identification-related question include: (1) which model parameters are identified and which are not; (2) how well identified are the identifiable parameters; (3) if the identification of some parameters fails or is weak, is this due to data limitations, or is it intrinsic to the structure of the model; (4) which features of the data are most important for the identification (or lack thereof) of a given parameter; (5) how the answers to (1)-(4) vary across different regions in the parameter space and for different sets of observables. The purpose of this paper is to show how answers to questions like these can be obtained for any linearized DSGE model.

A central tool in the proposed approach is the expected Fisher information matrix, the use of which for identification analysis was first suggested by Rothenberg (1971). As Rothenberg points out, the information matrix "is a measure of the amount of information about the unknown parameters available in the sample". Information deficiency results in identification problems and is associated with singular or nearly-singular information matrix. In addition to the purely statistical dimension of these problems there is also an economic modeling aspect, which is often more important. Parameters are unidentifiable or weakly identified if the economic features they represent 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 overparameterized

(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) and Komunjer and Ng (2009), which address the parameter identifiability question, and Canova and Sala (2009), which is focused on the weak identification problem. Iskrev (2009) presents 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. Komunjer and Ng (2009) derive a similar rank condition for identification using the spectral density matrix. 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 may be 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 group. 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 introduces the class of linearized DSGE models, and outlines the derivation of the log-likelihood function and the Fisher information matrix for Gaussian models. Section 3 explains the role of the Fisher information matrix in the analysis of identification, and discusses different aspect of the a priori approach to the issue. In particular, I show how to measure the strength of identification of each parameter, how to determine the sources of identification, and how to detect the causes for identification problems. 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 Preliminaries

This section provides a brief discussion of the class of linearized DSGE models and the restrictions they imply on the first and second order moments of the observed variables.

2.1 Setup

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. Let $\hat{\boldsymbol{z}}_t$ be a $m-$dimensional vector of the stationary variables, and $\hat{\boldsymbol{z}}^*$ be the steady state value of $\hat{\boldsymbol{z}}_t$. Once linearized, most DSGE models can be written in the following form

$$\boldsymbol{\Gamma}_0(\boldsymbol{\theta})\boldsymbol{z}_t = \boldsymbol{\Gamma}_1(\boldsymbol{\theta}) \operatorname{E}_t \boldsymbol{z}_{t+1} + \boldsymbol{\Gamma}_2(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_3(\boldsymbol{\theta})\boldsymbol{\epsilon}_t \tag{2.1}$$

where $\boldsymbol{z}_t$ is a $m-$dimensional vector of endogenous and exogenous state variables, and the structural shocks $\boldsymbol{\epsilon}_t$ are independent and identically distributed n-dimensional random vectors with $\operatorname{E}\boldsymbol{\epsilon}_t = \boldsymbol{0}$, $\operatorname{E}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t' = I_n$. 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), King and Watson (1998), 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

$$\boldsymbol{z}_t = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t \tag{2.2}$$

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

In most applications the model in (2.2) cannot be taken to the data directly since some of the variables in $\boldsymbol{z}_t$ are not observed. Instead, the solution of the DSGE model is expressed in a state space form, with transition equation given by (2.2), and a measurement equation

$$\boldsymbol{x}_t = \boldsymbol{C}\boldsymbol{z}_t + \boldsymbol{D}\boldsymbol{u}_t + \boldsymbol{\nu}_t \tag{2.3}$$

where $\boldsymbol{x}_t$ is a l-dimensional vector of observed variables, $\boldsymbol{u}_t$ is a q-dimensional vector of exogenous variables, and $\boldsymbol{\nu}_t$ is a p-dimensional random vectors with $\operatorname{E}\boldsymbol{\nu}_t = \boldsymbol{0}$, $\operatorname{E}\boldsymbol{\nu}_t\boldsymbol{\nu}_t' = \boldsymbol{Q}$, where $\boldsymbol{Q}$ is $p \times p$ symmetric semi-positive definite matrix.

For a given value of $\boldsymbol{\theta}$, the matrices $\boldsymbol{A}$, $\boldsymbol{\Omega} := \boldsymbol{B}\boldsymbol{B}'$, and $\hat{\boldsymbol{z}}^*$ completely characterize the equilibrium dynamics and steady state properties of all endogenous variables in the linearized model. Typically, some elements of these matrices are constant, i.e. independent of $\boldsymbol{\theta}$. For instance, if the steady state of some variables is zero, the corresponding elements of $\hat{\boldsymbol{z}}^*$ will be zero as well. Furthermore, if there are exogenous autoregressive (AR) shocks in the model, the matrix $\boldsymbol{A}$ will have rows composed of zeros and the AR coefficients. As a practical matter, it is useful to separate the solution parameters that depend on $\boldsymbol{\theta}$ from those that do not. I will use $\boldsymbol{\tau}$ to denote the vector collecting the non-constant elements of $\hat{\boldsymbol{z}}^*$, $\boldsymbol{A}$, and $\boldsymbol{\Omega}$, i.e. $\boldsymbol{\tau} := [\boldsymbol{\tau}_z', \ \boldsymbol{\tau}_A', \ \boldsymbol{\tau}_\Omega']'$, where $\boldsymbol{\tau}_z$, $\boldsymbol{\tau}_A$, and $\boldsymbol{\tau}_\Omega$ denote the elements of $\hat{\boldsymbol{z}}^*$, vec$(\boldsymbol{A})$ and vech$(\boldsymbol{\Omega})$ that depend on $\boldsymbol{\theta}$.

2.2 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 $\boldsymbol{e}_{t|t-1} = \boldsymbol{x}_t - \boldsymbol{C}\hat{\boldsymbol{z}}_{t|t-1} - \boldsymbol{D}\boldsymbol{u}_t$ 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 $\boldsymbol{\epsilon}_t$ 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

$$\ell_T(\boldsymbol{\theta}) = \text{const.} - \frac{1}{2}\sum_{t=1}^{\mathrm{T}} \log\det(\boldsymbol{S}_{t|t-1}) - \frac{1}{2}\sum_{t=1}^{\mathrm{T}} \boldsymbol{e}_{t|t-1}^{'}\boldsymbol{S}_{t|t-1}^{-1}\boldsymbol{e}_{t|t-1} \tag{2.4}$$

The ML estimator $\hat{\boldsymbol{\theta}}_T$ is the value of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ which maximizes (2.4). As I will discuss in Section 3.1, the precision of $\hat{\boldsymbol{\theta}}_T$ is determined by 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 1. *The expected Fisher information matrix is given by*

$$\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \mathrm{E}\left[\left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)' \boldsymbol{S}_t^{-1} \left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)\right] + \frac{1}{2}\sum_{t=1}^{T}\left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_t)}{\partial \boldsymbol{\theta}'}\right)' \left(\boldsymbol{S}_t \otimes \boldsymbol{S}_t\right)^{-1} \left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_t)}{\partial \boldsymbol{\theta}'}\right) \tag{2.5}$$

The asymptotic information matrix, defined as the limit of (2.5), can be computed using the following result (see Ljung (1999))

Theorem 2. *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*

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \mathrm{E}\left[\left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)' \boldsymbol{S}_{\infty}^{-1}\left(\frac{\partial \boldsymbol{e}_{t|t-1}}{\partial \boldsymbol{\theta}'}\right)\right] + \frac{1}{2}\left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_{\infty})}{\partial \boldsymbol{\theta}'}\right)'(\boldsymbol{S}_{\infty} \otimes \boldsymbol{S}_{\infty})^{-1}\left(\frac{\partial \mathrm{vec}(\boldsymbol{S}_{\infty})}{\partial \boldsymbol{\theta}'}\right) \tag{2.6}$$

To evaluate either (2.5) or (2.6), 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 given my the difference between the vector of theoretical first and second order moments and the vector of their sample counterparts. The optimal weighting matrix, given by the inverse of the moment function, is not available in closed-form unless Gaussianity is assumed. It can be shown that the inverse of the information matrix (2.6) 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 Identification Analysis

3.1 General principles

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 $\boldsymbol{\theta}^o \in \boldsymbol{\Theta}$ is identified if

$$f(\boldsymbol{X};\boldsymbol{\theta}) = f(\boldsymbol{X};\boldsymbol{\theta}^o) \text{ with probability } 1 \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}^o \tag{3.1}$$

This definition is made operational by using the following property of the log-likelihood function[1] $\ell_T(\boldsymbol{\theta}) := \log f(\boldsymbol{X};\boldsymbol{\theta})$

$$\mathrm{E}_0\, \ell_T(\boldsymbol{\theta}^o) \geq \mathrm{E}_0\, \ell_T(\boldsymbol{\theta}), \text{ for any } \boldsymbol{\theta} \tag{3.2}$$

It follows that the function $H(\boldsymbol{\theta}^o, \boldsymbol{\theta}) := \mathrm{E}_0\,(\ell_{\mathrm{T}}(\boldsymbol{\theta}) - \ell_T(\boldsymbol{\theta}^o))$ achieves a maximum at $\boldsymbol{\theta} = \boldsymbol{\theta}^o$, and $\boldsymbol{\theta}^o$ 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 $\boldsymbol{\theta}^o$ may be established by verifying the usual first and second order conditions, namely: (a) $\frac{\partial H(\boldsymbol{\theta}^o,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^o} = \mathbf{0}$, (b) $\frac{\partial^2 H(\boldsymbol{\theta}^o,\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^o}$ is negative definite. If the maximum at $\boldsymbol{\theta}^o$ is locally unique we say that $\boldsymbol{\theta}^o$ is locally identified. This means that there exists an open neighborhood of $\boldsymbol{\theta}^o$ where (3.1) holds for all $\boldsymbol{\theta}$. Global identification, on the other hand, extends the uniqueness of $f(\boldsymbol{X};\boldsymbol{\theta}^o)$ 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, given by

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

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

Theorem 3. *Let $\boldsymbol{\theta}^o$ be a regular point of the information matrix $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta})$ Then $\boldsymbol{\theta}^o$ is locally identifiable if and only if $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}^o)$ 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

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

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. Singularity of the information matrix means that likelihood function is flat at $\boldsymbol{\theta}^o$ and one has no hope of finding the true values of some of the parameters even with an infinite number of observations. Intuitively, this may occur for one of two reasons: either some parameters do not affect the likelihood at all, or different parameters have the same effect on the likelihood. This reasoning may be formalized by using the fact that the information matrix is equal to the covariance matrix of the scores, and therefore can be expressed as

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^o) = \boldsymbol{\Delta}^{\frac{1}{2}} \tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\theta}^o) \boldsymbol{\Delta}^{\frac{1}{2}} \tag{3.3}$$

where $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}^o))$ is a diagonal matrix containing the variances of the elements of the score vector, and $\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\theta}^o)$ is the correlation matrix of the score vector.

Hence, a parameter θ_i is locally unidentifiable if:

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

$$\Delta_i := \text{var}\left(\frac{\partial \ell_T(\boldsymbol{\theta}^o)}{\partial \theta_i}\right) = 0 \tag{3.4}$$

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

$$\boldsymbol{\varrho}_i := \text{corr}\left(\frac{\partial \ell_T(\boldsymbol{\theta}^o)}{\partial \theta_i}, \frac{\partial \ell_T(\boldsymbol{\theta}^o)}{\partial \boldsymbol{\theta}_{-i}}\right) = 1 \tag{3.5}$$

where $\frac{\partial \ell_T(\boldsymbol{\theta}^o)}{\partial \boldsymbol{\theta}_{-i}}$ is the partial derivative of the log-likelihood with respect to $[\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k]$.

Weak identification arises when the likelihood is not completely flat, but exhibits very low curvature with respect to some parameters. That is, when $\Delta_i \approx 0$ or $\boldsymbol{\varrho}_i \approx 1$

for some θ_i. The problem of detecting and measuring weak identification problems is discussed next.

3.2 Evaluating the identification strength

The rank condition ensures that the expected log-likelihood function is not flat and achieves a locally unique maximum at the true value of $\boldsymbol{\theta}$. In general this suffices for the consistent estimation of $\boldsymbol{\theta}$. However, the precision with which $\boldsymbol{\theta}$ may be estimated in finite samples depends on the degree of curvature of log-likelihood surface in the neighborhood of $\boldsymbol{\theta}^o$, of which the rank condition provides no information. Nearly flat likelihood means that small changes in the value of $\ell_T(\boldsymbol{\theta}^o)$, due to random variations in the sample, 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 that can be applied in order 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. Formally, this relationship is revealed by the fact that the covariance of any unbiased estimator of $\boldsymbol{\theta}^o$ is bounded from below by the inverse of the Fisher information matrix, which measures the expected curvature of the log-likelihood. This is known as Cramér-Rao lower bound

Theorem 4. *Let $\boldsymbol{X}$ be a $N-$dimensional random vector with a probability density function $f(\boldsymbol{X};\boldsymbol{\theta})$.If $\hat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$, and $\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} := \mathrm{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$, then*

$$\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{\mathcal{I}}_N(\boldsymbol{\theta})^{-1} \geq \mathbf{O} \quad (\textit{i.e. } \boldsymbol{V}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{\mathcal{I}}_N(\boldsymbol{\theta})^{-1} \textit{ is positive semi-definite}) \tag{3.6}$$

provided that the Fisher information matrix $\boldsymbol{\mathcal{I}}_N(\boldsymbol{\theta})$ is not singular when evaluated at $\boldsymbol{\theta}$.

A proof of this result can be found in Schraf (1991). Using (3.6) and the factorization of the information matrix in (3.3), it is straightforward to show that $\mathrm{std}(\hat{\theta}_i) :=$

$\sqrt{\mathrm{E}(\hat{\theta}_i - \theta_i)^2}$ is bounded from below as follows[2]

$$\mathrm{std}(\hat{\theta}_i) \geq \sqrt{\{\boldsymbol{\mathcal{I}}_N\left(\boldsymbol{\theta}\right)^{-1}\}_{ii}} = \frac{1}{\sqrt{\Delta_i}}\left(\frac{1}{\sqrt{1-\boldsymbol{\varrho}_i^2}}\right) \tag{3.7}$$

This decomposition confirms the direct link between the curvature of the likelihood with respect to θ_i and the precision with which θ_i may be estimated. It also provides a way to assess the relative importance of the two possible causes of identification problems - $\Delta_i \approx 0$ and $\boldsymbol{\varrho}_i \approx 1$. This distinction is useful if one wishes to determine what features of the model lead to identification problems for some parameters. Consider the first term. A very small variance of $\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \theta_i}$ 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. The second term in (3.7) 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.

Note that, unlike $\boldsymbol{\varrho}_i$ which is scale-invariant, the value of Δ_i depends on the units in which θ_i is measured. The variance term is made scale-invariant by dividing it by the absolute value of θ_i. The relative Cramér-Rao bounds, obtained by dividing both sides of (3.7) by $|\theta_i|$, can be used to assess the relative strength of identification of the deep parameters, as well as to study how the identification of a given θ_i varies across regions in the parameter space.

A further interpretation of the bounds on the variances in terms of confidence intervals for the parameters is possible if one is willing to resort to asymptotic theory. The asymptotic version of the Cramér-Rao lower bound is obtained when $\boldsymbol{\mathcal{I}}_N$ in (3.6) is replaced by the asymptotic information matrix. This gives a lower bound on the asymptotic covariance matrix of any consistent estimator of $\boldsymbol{\theta}$. Furthermore, the bound is achieved by the ML estimator $\hat{\boldsymbol{\theta}}_T$ whose asymptotic distribution is given by

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^o) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}, \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}^o)\right) \tag{3.8}$$

[2]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)

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^o) = \operatorname*{plim}_{T\to\infty} \frac{1}{T}\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}^o)$ denotes the asymptotic expected Fisher information matrix. The asymptotic normality of $\hat{\boldsymbol{\theta}}_T$ implies that

$$T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^o)'\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^o)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^o) \xrightarrow{d} \chi^2(k) \tag{3.9}$$

This result allows us to construct 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

$$T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})'\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^o)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \leq c_\alpha \tag{3.10}$$

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 (3.10) onto the parameter axes. This leads to intervals of the form

$$\theta_i - \sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)} \leq \hat{\theta}_i \leq \theta_i + \sqrt{c_\alpha \operatorname{var}(\hat{\theta}_i)} \tag{3.11}$$

where $\operatorname{var}(\hat{\theta}_i)$ is the asymptotic variance of the ML estimator of θ_i.

These intervals have 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 population 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 (3.11) includes points that lie outside the hyperellipsoid defined by (3.10). Third, the confidence intervals 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 asymptotic confidence intervals do not take into account such restrictions and may include values that are theoretically inadmissible. Therefore, though useful as a way for quantifying the curvature of the expected likelihood, the confidence intervals in (3.11) may not be appropriate to use for statistical inference.

3.3 Determining the sources of identification

To evaluate the strength of identification of the parameters we take into account all model-implied restrictions on the joint probability distribution of the sample. A natural step further in the analysis is to ask which characteristics of the distribution are most important for the identification of a given parameter. In principle, one may be able to answer this question by reasoning alone, by trying to establish the link between the feature of the model represented by the parameter, and the regularities in the observed data the model is designed to explain. In practice, however, the association between parameters and empirical implications of the model is often difficult to discern, especially for large models. It is therefore desirable to have a formal method for doing that. In this section I discuss one such approach, based on the model-implied relationship between the deep parameters and the moments of the observables.

Let $\boldsymbol{m}$ denote a vector collecting the unique elements of the mean and the covariance matrix of $\boldsymbol{X}_T$. Consider first the following question: which one among the moments in $\boldsymbol{m}$ is the most important for the identification of a given parameter θ_j? The answer depends on two factors: how well identified are the moments, and how informative for θ_j is each moment. Well identified moments are the ones that can be estimated accurately, using the restrictions implied by the model. A measure of the accuracy is the asymptotic covariance matrix $\boldsymbol{V_m}$ for the moments, which is easy to obtain from the covariance matrix for $\boldsymbol{\theta}$. Furthermore, the model predicts that a moment is very informative about a deep parameter if the value of the moment is very sensitive to changes in the value of that parameter. The sensitivity of any moment m_i to a parameter θ_j is given by

$$\frac{\partial m_i}{\partial \theta_j}\frac{1}{m_i} \tag{3.12}$$

Analytical formulae for computing the partial derivatives of the moments with respect to $\boldsymbol{\theta}$ are given in Iskrev (2009). The most important for the identification of θ_j moment maximizes

$$\frac{\partial m_i}{\partial \theta_j}\frac{1}{\sqrt{\boldsymbol{V_m}(i,i)}} \tag{3.13}$$

where $\boldsymbol{V_m}(i,i)$ is the $i-$th diagonal element of the covariance matrix of $\boldsymbol{m}$. Note that the expression in (3.13) comprises two terms: the sensitivity ($\frac{\partial m_i}{\partial \theta_j}\frac{1}{m_i}$), and the relative precision of an estimator of m_i ($\frac{m_i}{\sqrt{\boldsymbol{V_m}(i,i)}}$). It can be interpreted as the asymptotic standard error of an estimator of θ_j derived from the model-implied restrictions on m_i.[3]

The measure in (3.13) can be extended to multivariate functions of the moments. Suppose that $\boldsymbol{p}(\boldsymbol{m})$ is a continuously differentiable functions of $\boldsymbol{m}$. Then the weighted sensitivity measure is given by

$$\sqrt{\left(\frac{\partial \boldsymbol{p}}{\partial \theta_j}\right)' \boldsymbol{V}_{\boldsymbol{p}}^{-1} \left(\frac{\partial \boldsymbol{p}}{\partial \theta_j}\right)} \tag{3.14}$$

Using (3.14), one can evaluate the importance for the identification of θ_i of more complicated patterns in the data, not captures by a single moment. For instance, $\boldsymbol{p}$ may be defined as a function selecting the moments - mean, variance and autocovariances up to lag $T-1$, of only one of the observed variables. Comparing (3.14) computed for each variable would indicate which observable is most useful for identifying a given deep parameter.

3.4 Discussion

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 (2.5) and (2.6). Taking the linearized structural model in (2.1) 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 (2.5), 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 $\boldsymbol{\Theta}$. 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

[3]Consider the problem of estimating θ_j by matching the sample value of m_i with the one implied by the model, conditional on $\boldsymbol{\theta}_{-i}$. The asymptotic standard error of the estimator is equal to the inverse of the expression in (3.13).

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 (3.10), 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 (2.3)). 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 necessary 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:

$$\frac{\partial \ell_T(\boldsymbol{\theta}^o)}{\partial \boldsymbol{\theta}'} = \frac{\partial \ell_T}{\partial \boldsymbol{\tau}'} \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'} \tag{3.15}$$

and therefore the information matrix may be written as

$$\boldsymbol{\mathcal{I}}_T = \left(\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'}\right)' \mathrm{E}\left\{\left(\frac{\partial \ell_T}{\partial \boldsymbol{\tau}'}\right)' \left(\frac{\partial \ell_T}{\partial \boldsymbol{\tau}'}\right)\right\} \left(\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}'}\right) \tag{3.16}$$

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 properties of the Jacobian matrix, when it has full column rank, has implications for the strength of identification of $\boldsymbol{\theta}$. From (3.15) it is clear that the two types of weak identification problems discussed in Section 3.1 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 3.4. Specifically, θ_i is weakly identified in the model if either one of the following two conditions holds:

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

$$\left\| \frac{\partial \boldsymbol{\tau}}{\partial \theta_i} \right\| \approx 0 \tag{3.17}$$

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

$$\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}} \right) \approx 1 \tag{3.18}$$

If (a) is true, changing θ_i while keeping the other deep parameters fixed has almost no effect on $\boldsymbol{\tau}$. 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 (3.18) 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 apply the rank conditions developed above to a medium-scale DSGE model estimated in Smets and Wouters (2007) (SW07 henceforth). I start with an outline of the main components 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 exogenous term in the measurement equation (2.3) is given by $\boldsymbol{u}_t = 1$ for all t, and

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

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. Since there is no measurement error, the last term in (2.3) is omitted.

The deep parameters of the model are collected in a 41-dimensional vector $\boldsymbol{\theta}$ given by[4]

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

4.2 Identification Analysis

The identifiability of the parameters in the SW07 model was studied in Iskrev (2009), and the following was found: 37 out of the 41 parameters in (4.3) are locally identified; the remaining four parameters - ξ_w, ξ_p, ϵ_w andϵ_p, are not separately identifiable in the sense that in the linearized model ξ_w cannot be distinguished from ϵ_w, and ξ_p cannot be distinguished from ϵ_p. As in SW07, I will assume that ϵ_w and ϵ_p are known and are both equal to 10. The purpose of this section is to study the strength of identification of the remaining 39 parameters.

Proceeding as discussed in Section 3.4, I draw randomly $100,000$ points from $\boldsymbol{\Theta}$, which is defined using the prior distribution in SW07 (see Table 3), and compute the theoretical Cramér-Rao lower bounds on $\frac{std(\hat{\theta}_i)}{|\theta_i|}$ at each point. The expected information matrix is evaluated for $T = 156$. Table 4 presents a summary of the results, showing the mean, the variance, and the nine deciles of the distribution of the bounds for each parameter.

The first thing to notice about the table is the large variability of the bounds for some parameters, namely $\bar{l}$, λ_w, μ_p, ρ_p, ξ_w, ρ_w, and μ_w. This implies that the strength

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

of identification of these parameters is not a constant feature of the model, but varies greatly across different points in the parameter space. Particularly striking is the variability for ρ_w, and μ_w, for which the variances of the bounds are several orders of magnitude greater than those of most other parameters. On the other extreme are γ, Φ, ρ, λ and the standard deviations of the shock processes, whose bounds exhibit very low variability.

Many of the later group of parameters, and in particular γ, Φ, ρ and λ, are also consistently among the best identified parameters in the model. To a lesser extend this is also true for the standard deviations of the shocks to government purchases and to TFP. Other parameters that are frequently among those with the lowest bounds are ξ_p and α. Similarly, there are parameters whose bounds are both very volatile and large; among these are ρ_w, ρ_p, λ_w, ξ_w, μ_p and μ_w. Other parameters that are consistently among the worst identified are: $\bar{l}$, β, σ_l, r_y and ρ_{ga}. For an easier comparison of the relative strength of identification, Table 5 shows the order the parameters at each decile according to the values of the bounds reported in Table 4.

As was shown in Section 3.2, the value of the theoretical Cramér-Rao bound for a given parameter can be expressed as the product of two factors, of which one reflects the sensitivity of the likelihood to that parameter, and the other measures the degree of collinearity between the effects of different parameters on the likelihood. Large values of the lower bounds may be a consequence of low sensitivity, high collinearity, or both, and the factorization can be used to determine the relative importance of these two causes. This is done in Table 6, where the two factors, referred to as sensitivity and collinearity components of the bounds, are reported for the values of the bounds in Table 4. The results show very large collinearity components for all of the worst identified parameters that were highlighted above. Particularly large are the values for the wage stickiness parameter (ξ_w) and the wage markup parameter (λ_w), which at the same time have some of the smallest values of the sensitivity component. This implies that each one of these parameters represents a structural feature that is well identified empirically, but can be captured in the model by a set of other deep parameters. To a lesser degree the same conclusion applies to the elasticity of labor supply σ_l, and to the wage and price markup shock parameters ρ_w, ρ_p, μ_p, and μ_w. On the other end of the spectrum are parameters like ρ_{ga} and $\bar{l}$, which tend to have very low collinearity components and very large sensitivity components. This implies that, although these parameters play a distinct role in the model, their effects on the likelihood are too weak to allow for a precise estimation.

Another way to think about the results in Table 6 is that the sensitivity component for a given parameter shows what the value of the Cramér-Rao lower bound would be if all other parameters were known. From the table it is clear that the collinearity effects are large and play a substantial role in determining the strength of identification of most deep parameter in the model. This can be seen more clearly from the values of the multiple correlation coefficients ϱ_i associated with the collinearity components (see equation (3.7)), which are bounded between 0 and 1. From Table 7, which shows the values of ϱ_i corresponding to the collinearity components in Table 6, we can see that the multiple correlations exceed .9 for most parameters; for ξ_w, λ_w and σ_c the values are above .99 in all nine points. A more complete picture of the distributions of the correlation coefficients is provided in Table 8, which presents the mean, the variance and the deciles of ϱ_i. As can be seen from the values in the upper deciles of the distributions, for almost all parameters there are points in the parameter space where the values of ϱ_i exceeds .99. At the same time, for many of the parameters there are also points where the collinearity is much weaker, and thus the variance of ϱ_i across the parameter space is relatively large. On the other hand, several parameters have very high multiple collinearity coefficients everywhere in the parameter space, as indicated by the large mean values of ϱ_i, combined with very low variances. In addition to ξ_w, λ_w and σ_c mentioned above, the other parameters in this group are σ_l, β, ρ, φ, λ and r_π.

The large values of $\boldsymbol{\varrho}_i$ in Table 8 suggest that many structural features are difficult to distinguish on the basis of the empirical implications of the model, as summarized by the likelihood function of the seven observables I have considered. While for most parameters this is only a local problem, affecting only certain regions in the parameter space, for others this appears to be a global problem. One possible explanation is that the moments of the seven observables do not represent fully the properties of the economic model, and with a different or a larger set of observables one may be able to better capture the distinct roles the parameters play in it. However, it may also be that some parameters play very similar roles in the structural model, and would therefore be difficult to distinguish with any empirical model. As was discussed in Section 3.4, the second possibility may be investigated by studying the Jacobian matrix of the reduced-form parameters $\boldsymbol{\tau}$ with respect to $\boldsymbol{\theta}$. Since $\boldsymbol{\tau}$ fully characterizes the steady state properties and equilibrium dynamics of all endogenous variables, deep parameters that have similar effects on $\boldsymbol{\tau}$ will be hard to distinguish on the basis of any subset of the model variables. As discussed in Section , this is similar to having strong collinearity among the derivatives of the log-likelihood, and can be assessed using $\cos\left(\frac{\partial\boldsymbol{\tau}}{\partial\theta_i}, \frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}_{-i}}\right)$, 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 $\frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}}$.

An evaluation of the degree of collinearity in the model is provided in Table 9, which presents the mean, the variance and the deciles of $\cos\left(\frac{\partial\boldsymbol{\tau}}{\partial\theta_i}, \frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}_{-i}}\right)$. While the results for most parameters are consistent with those in Table 8, there are some remarkable differences. The degree of collinearity in the model for $\bar{l}$ (steady state hours worked) is zero for all $\boldsymbol{\theta}$, while with respect to the likelihood the degree of collinearity is .85 on average, and reaches .99 for some parameter values. The zero collinearity in the model is due to the fact that $\bar{l}$ affects only one element of $\boldsymbol{\tau}$ - the steady state level, or equivalently, the mean of hours worked, and is the only parameter that does so. The non-zero collinearity with respect to the likelihood arises because parameters other than $\bar{l}$ affect the information in the likelihood about $\bar{l}$ through their effect on the moments of observables other than l_t. In other words, the information matrix for $\boldsymbol{\tau}$ has non-zero off-diagonal elements in the row and column corresponding to $\bar{l}$. According to the decomposition in (3.16), this results in a non-zero multiple correlation coefficient for $\bar{l}$.

The results in Table 9 suggest that the SW07 model contains too many features, some of which are nearly redundant given the other features present in the model. The most severely affected parameters are: σ_c, λ_w, ξ_w,λ, 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 correlation 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 ?? 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 parameters listed above. The result can be seen in the last column of Table 10. The table also reports the values of $\cos\left(\frac{\partial\boldsymbol{\tau}}{\partial\theta_i}, \frac{\partial\boldsymbol{\tau}}{\partial\boldsymbol{\theta}_{-i}}\right)$ when $\boldsymbol{\theta}_{-i}$ includes only the selected parameters. To facilitate the comparison with Table ??, the coefficients are evaluated at the same points in the parameter space. Therefore, for instance, at the point where the collinearity for σ_c with respect to all other deep parameters is .98, the coefficient is .93 if only g_y, λ and σ_l are included.

The list of parameters shown in the last column of Table 10 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 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 10 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 described 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, many 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.

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)	$c_t^* = \dfrac{\lambda/\gamma}{1+\lambda/\gamma} c_{t-1}^* + \dfrac{1}{1+\lambda/\gamma} \mathrm{E}_t\, c_{t+1}^* + \dfrac{w^{ss} l^{ss}(\sigma_c - 1)}{c^{ss}\sigma_c(1+\lambda/\gamma)}(l_t^* - \mathrm{E}_t\, l_{t+1}^*)$
	$-\frac{1-\lambda/\gamma}{(1+\lambda/\gamma)\sigma_c} r_t^* - \frac{1-\lambda/\gamma}{(1+\lambda\lambda/\gamma)\sigma_c} \varepsilon_t^b$
(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: Theoretical Cramér-Rao lower bounds

			deciles								
Parameter	mean	var.	1	2	3	4	5	6	7	8	9
ρ_{ga}	0.7	16.7391	0.073	0.105	0.139	0.182	0.237	0.312	0.433	0.652	1.234
$\bar{l}$	1.9	15865.6	0.054	0.081	0.110	0.145	0.192	0.259	0.368	0.575	1.187
$\bar{\pi}$	0.3	0.1022	0.073	0.101	0.129	0.161	0.199	0.247	0.313	0.410	0.601
$100(\beta^{-1}-1)$	1.0	0.3923	0.394	0.507	0.609	0.708	0.818	0.945	1.100	1.317	1.709
μ_w	4.9	111339	0.061	0.137	0.254	0.404	0.580	0.817	1.165	1.782	3.508
μ_p	3.8	20127.3	0.066	0.136	0.262	0.432	0.631	0.880	1.233	1.863	3.536
α	0.1	0.0029	0.034	0.044	0.054	0.063	0.074	0.085	0.101	0.122	0.157
ψ	0.2	0.0175	0.055	0.079	0.102	0.126	0.151	0.178	0.214	0.262	0.348
φ	0.3	0.0111	0.143	0.179	0.206	0.232	0.258	0.286	0.317	0.355	0.410
σ_c	0.3	0.0116	0.125	0.161	0.190	0.217	0.243	0.272	0.304	0.343	0.400
λ	0.1	0.0014	0.011	0.019	0.027	0.035	0.043	0.053	0.064	0.078	0.100
Φ	0.1	0.0009	0.018	0.026	0.033	0.039	0.046	0.053	0.062	0.074	0.091
ι_w	0.3	0.1028	0.072	0.102	0.129	0.159	0.194	0.239	0.299	0.393	0.576
ξ_w	2.9	21581.3	0.168	0.234	0.296	0.363	0.445	0.559	0.741	1.077	2.078
ι_p	0.2	0.0694	0.050	0.070	0.090	0.112	0.139	0.173	0.222	0.297	0.456
ξ_p	0.1	0.0035	0.032	0.042	0.049	0.056	0.064	0.074	0.086	0.104	0.141
σ_l	0.4	0.1273	0.100	0.147	0.191	0.238	0.295	0.361	0.450	0.588	0.825
r_π	0.2	0.0210	0.056	0.085	0.110	0.137	0.167	0.200	0.241	0.296	0.385
$r_{\triangle y}$	0.3	0.7487	0.072	0.103	0.130	0.160	0.194	0.236	0.292	0.379	0.563
r_y	0.9	3.1517	0.241	0.339	0.427	0.520	0.617	0.738	0.895	1.127	1.604
ρ	0.1	0.0011	0.019	0.027	0.035	0.042	0.049	0.058	0.067	0.079	0.098
ρ_a	0.2	0.0439	0.062	0.082	0.102	0.121	0.143	0.170	0.206	0.260	0.365
ρ_b	0.2	0.0525	0.031	0.056	0.080	0.107	0.136	0.170	0.213	0.274	0.395
ρ_g	0.2	0.0454	0.062	0.082	0.101	0.121	0.142	0.169	0.203	0.255	0.362
ρ_I	0.2	0.0550	0.071	0.095	0.117	0.140	0.165	0.195	0.236	0.296	0.418
ρ_r	0.2	0.0608	0.024	0.046	0.071	0.099	0.131	0.169	0.217	0.284	0.408
ρ_p	3.7	20127.6	0.133	0.225	0.317	0.423	0.554	0.741	1.024	1.573	3.176
ρ_w	4.8	111339	0.124	0.212	0.298	0.399	0.524	0.712	0.997	1.545	3.193
γ	0.0	0.0000	0.002	0.002	0.003	0.003	0.004	0.004	0.005	0.006	0.008
σ_a	0.1	0.0002	0.073	0.074	0.076	0.078	0.080	0.084	0.088	0.094	0.105
σ_b	0.1	0.0005	0.090	0.100	0.107	0.113	0.119	0.125	0.131	0.138	0.149
σ_g	0.1	0.0001	0.072	0.073	0.074	0.075	0.076	0.078	0.080	0.083	0.088
σ_I	0.1	0.0002	0.101	0.107	0.111	0.114	0.117	0.120	0.123	0.126	0.132
σ_r	0.1	0.0010	0.072	0.073	0.076	0.080	0.085	0.091	0.100	0.113	0.134
σ_p	0.1	0.0020	0.087	0.091	0.094	0.098	0.102	0.108	0.115	0.126	0.150
σ_w	0.1	0.0009	0.085	0.088	0.092	0.096	0.101	0.106	0.113	0.123	0.142
δ	0.3	0.0413	0.098	0.140	0.177	0.216	0.257	0.305	0.365	0.447	0.578
λ_w	3.0	18971.2	0.180	0.248	0.313	0.383	0.472	0.592	0.779	1.135	2.222
g_y	0.2	0.0198	0.074	0.101	0.124	0.147	0.173	0.202	0.237	0.285	0.367

Note: Each row of the table shows the mean, the variance and the nine deciles of the theoretical Cramér-Rao lower bounds for $\frac{std(\hat{\theta}_i)}{|\theta_i|}$. The results are based on $100,000$ draws from $\boldsymbol{\Theta}$.

Table 5: Order of parameters according to the strength of identification

deciles of the Cramér-Rao bounds								
1	2	3	4	5	6	7	8	9
γ	γ	γ	γ	γ	γ	γ	γ	γ
λ	λ	λ	λ	λ	λ	Φ	Φ	σ_g
Φ	Φ	Φ	Φ	Φ	Φ	λ	λ	Φ
ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ
ρ_r	ξ_p	ξ_p	ξ_p	ξ_p	ξ_p	σ_g	σ_g	λ
ρ_b	α	α	α	α	σ_g	ξ_p	σ_a	σ_a
ξ_p	ρ_r	ρ_r	σ_g	σ_g	σ_a	σ_a	ξ_p	σ_I
α	ρ_b	σ_g	σ_a	σ_a	α	σ_r	σ_r	σ_r
ι_p	ι_p	σ_a	σ_r	σ_r	σ_r	α	α	ξ_p
$\bar{l}$	σ_g	σ_r	σ_w	σ_w	σ_w	σ_w	σ_w	σ_w
ψ	σ_r	ρ_b	σ_p	σ_p	σ_p	σ_p	σ_p	σ_b
r_π	σ_a	ι_p	ρ_r	σ_I	σ_I	σ_I	σ_I	σ_p
μ_w	ψ	σ_w	ρ_b	σ_b	σ_b	σ_b	σ_b	α
ρ_a	$\bar{l}$	σ_p	ι_p	ρ_r	ρ_g	ρ_g	ρ_g	ψ
ρ_g	ρ_a	ρ_g	σ_b	ρ_b	ρ_r	ρ_a	ρ_a	ρ_g
μ_p	ρ_g	ρ_a	σ_I	ι_p	ρ_a	ρ_b	ψ	ρ_a
ρ_I	r_π	ψ	ρ_g	ρ_g	ρ_b	ψ	ρ_b	g_y
$r_{\triangle y}$	σ_w	σ_b	ρ_a	ρ_a	ι_p	ρ_r	ρ_r	r_π
σ_r	σ_p	$\bar{l}$	ψ	ψ	ψ	ι_p	g_y	ρ_b
σ_g	ρ_I	r_π	r_π	ρ_I	ρ_I	ρ_I	r_π	σ_c
ι_w	σ_b	σ_I	ρ_I	r_π	r_π	g_y	ρ_I	ρ_r
σ_a	g_y	ρ_I	$\bar{l}$	g_y	g_y	r_π	ι_p	φ
ρ_{ga}	$\bar{\pi}$	g_y	g_y	$\bar{l}$	$r_{\triangle y}$	$r_{\triangle y}$	σ_c	ρ_I
$\bar{\pi}$	ι_w	$\bar{\pi}$	ι_w	$r_{\triangle y}$	ι_w	ι_w	φ	ι_p
g_y	$r_{\triangle y}$	ι_w	$r_{\triangle y}$	ι_w	$\bar{\pi}$	σ_c	$r_{\triangle y}$	$r_{\triangle y}$
σ_w	ρ_{ga}	$r_{\triangle y}$	$\bar{\pi}$	$\bar{\pi}$	$\bar{l}$	$\bar{\pi}$	ι_w	ι_w
σ_p	σ_I	ρ_{ga}	ρ_{ga}	ρ_{ga}	σ_c	φ	$\bar{\pi}$	δ
σ_b	μ_p	δ	δ	σ_c	φ	δ	δ	$\bar{\pi}$
δ	μ_w	σ_c	σ_c	δ	δ	$\bar{l}$	$\bar{l}$	σ_l
σ_l	δ	σ_l	φ	φ	ρ_{ga}	ρ_{ga}	σ_l	$\bar{l}$
σ_I	σ_l	φ	σ_l	σ_l	σ_l	σ_l	ρ_{ga}	ρ_{ga}
ρ_w	σ_c	μ_w	ξ_w	ξ_w	ξ_w	ξ_w	ξ_w	r_y
σ_c	φ	μ_p	λ_w	λ_w	λ_w	λ_w	r_y	β
ρ_p	ρ_w	ξ_w	ρ_w	ρ_w	ρ_w	r_y	λ_w	ξ_w
φ	ρ_p	ρ_w	μ_w	ρ_p	r_y	ρ_w	β	λ_w
ξ_w	ξ_w	λ_w	ρ_p	μ_w	ρ_p	ρ_p	ρ_w	ρ_p
λ_w	λ_w	ρ_p	μ_p	r_y	μ_w	β	ρ_p	ρ_w
r_y	r_y	r_y	r_y	μ_p	μ_p	μ_w	μ_w	μ_w
β	β	β	β	β	β	μ_p	μ_p	μ_p

Note: The parameters are ordered according to the increasing values of the theoretical Cramér-Rao lower bounds for $\frac{std(\hat{\theta}_i)}{|\theta_i|}$. See Table 4 for the values of the bounds.

Table 6: Sensitivity and collinearity components of the Cramér-Rao lower bounds

	1 decile		2 decile		3 decile		4 decile		5 decile		6 decile		7 decile		8 decile		9 decile	
Parameter	sens.	col.	sens.	col.	sens.	col.	sens.	col.	sens.	col.	sens.	col.	sens.	col.	sens.	col.	sens.	col.
ρ_{ga}	.07	1.1	.10	1.0	.10	1.4	.18	1.0	.23	1.0	.30	1.1	.43	1.0	.62	1.1	.78	1.6
$\bar{l}$	.01	4.1	.03	2.6	.08	1.4	.01	11.5	.12	1.5	.25	1.0	.11	3.5	.06	9.9	.09	13.9
$\bar{\pi}$	.06	1.2	.05	2.0	.10	1.3	.07	2.4	.02	9.0	.03	8.0	.03	10.6	.04	11.1	.07	8.8
$100(\beta^{-1}-1)$	.04	10.4	.12	4.2	.11	5.6	.10	6.8	.16	5.2	.20	4.7	.20	5.5	.05	24.2	.20	8.6
μ_w	.05	1.2	.04	3.6	.03	9.1	.11	3.6	.04	14.9	.18	4.5	.10	11.6	.12	15.4	1.43	2.5
μ_p	.02	3.1	.02	6.2	.04	7.4	.10	4.4	.06	10.1	.23	3.9	.31	3.9	.33	5.6	.75	4.7
α	.02	1.9	.01	5.5	.01	8.4	.02	2.9	.02	3.4	.01	6.8	.02	5.1	.02	7.0	.04	3.6
ψ	.02	2.8	.02	3.7	.02	5.2	.01	12.3	.01	15.5	.05	3.3	.07	3.0	.08	3.3	.04	7.9
φ	.02	6.5	.05	3.8	.06	3.5	.07	3.2	.03	9.3	.03	9.3	.02	14.4	.05	6.5	.08	4.8
σ_c	.01	13.0	.01	17.3	.01	21.8	.02	10.3	.02	10.3	.02	17.0	.01	25.4	.05	7.5	.03	13.3
λ	.00	7.8	.00	3.9	.00	5.6	.00	8.7	.00	10.1	.01	5.8	.02	2.6	.01	5.8	.03	2.9
Φ	.00	11.0	.00	5.5	.01	4.0	.02	1.8	.01	6.6	.01	6.5	.02	3.9	.02	3.9	.01	7.0
ι_w	.05	1.5	.04	2.5	.09	1.5	.08	2.1	.16	1.2	.17	1.4	.07	4.2	.11	3.5	.35	1.7
ξ_w	.00	52.1	.00	120.5	.03	8.9	.02	15.8	.01	47.1	.01	93.1	.03	23.1	.01	77.7	.01	281.6
ι_p	.01	6.8	.05	1.6	.05	1.9	.04	2.9	.08	1.7	.04	4.8	.15	1.5	.15	2.0	.12	3.7
ξ_p	.01	2.2	.02	2.1	.01	3.4	.01	4.2	.03	1.9	.04	2.1	.04	2.2	.04	2.7	.03	4.4
σ_l	.03	3.2	.02	8.2	.01	18.5	.03	6.9	.04	8.3	.05	6.8	.09	5.2	.14	4.3	.13	6.4
r_π	.01	4.9	.03	3.3	.02	6.0	.02	6.1	.05	3.2	.04	4.7	.05	5.0	.10	2.9	.11	3.6
$r_{\triangle y}$	.00	15.1	.03	3.1	.03	4.7	.09	1.8	.11	1.8	.07	3.4	.19	1.5	.25	1.5	.18	3.2
r_y	.05	4.5	.02	17.0	.23	1.9	.18	2.9	.19	3.3	.06	12.6	.20	4.4	.31	3.7	.21	7.7
ρ	.01	3.4	.00	10.4	.01	3.9	.00	12.5	.00	14.3	.01	10.1	.02	3.1	.02	4.1	.02	5.0
ρ_a	.05	1.2	.07	1.2	.09	1.2	.09	1.4	.07	2.1	.17	1.0	.20	1.0	.26	1.0	.30	1.2
ρ_b	.00	15.1	.02	3.4	.02	3.9	.01	11.1	.02	7.1	.06	3.0	.09	2.4	.12	2.2	.13	3.0
ρ_g	.05	1.2	.07	1.2	.09	1.1	.11	1.1	.13	1.1	.16	1.1	.20	1.0	.25	1.0	.36	1.0
ρ_I	.04	1.8	.06	1.7	.06	1.9	.09	1.6	.09	1.8	.11	1.7	.15	1.5	.17	1.8	.27	1.5
ρ_r	.01	4.6	.03	1.7	.03	2.1	.04	2.5	.13	1.0	.12	1.4	.15	1.4	.28	1.0	.27	1.5
ρ_p	.03	4.7	.06	3.8	.06	5.5	.30	1.4	.03	17.3	.08	9.8	.96	1.1	.09	17.8	.39	8.1
ρ_w	.03	3.8	.04	5.1	.04	6.7	.06	6.9	.08	6.7	.57	1.2	.13	8.0	.29	5.4	.11	28.5
γ	.00	1.0	.00	1.0	.00	1.0	.00	1.0	.00	1.0	.00	1.1	.01	1.0	.01	1.0	.01	1.0
σ_a	.07	1.0	.07	1.0	.07	1.1	.07	1.1	.07	1.1	.07	1.2	.07	1.2	.07	1.3	.07	1.5
σ_b	.07	1.3	.07	1.4	.07	1.5	.07	1.6	.07	1.7	.07	1.8	.07	1.8	.07	1.9	.07	2.1
σ_g	.07	1.0	.07	1.0	.07	1.0	.07	1.1	.07	1.1	.07	1.1	.07	1.1	.07	1.2	.07	1.2
σ_I	.07	1.4	.07	1.5	.07	1.6	.07	1.6	.07	1.6	.07	1.7	.07	1.7	.07	1.8	.07	1.9
σ_r	.07	1.0	.07	1.0	.07	1.1	.07	1.1	.07	1.2	.07	1.3	.07	1.4	.07	1.6	.07	1.9
σ_p	.07	1.2	.07	1.3	.07	1.3	.07	1.4	.07	1.4	.07	1.5	.07	1.6	.07	1.8	.07	2.1
σ_w	.07	1.2	.07	1.2	.07	1.3	.07	1.4	.07	1.4	.07	1.5	.07	1.6	.07	1.7	.07	2.0
δ	.04	2.8	.07	1.9	.08	2.1	.09	2.4	.18	1.4	.15	2.0	.05	6.7	.31	1.4	.28	2.0
λ_w	.00	50.6	.01	45.8	.01	46.2	.01	31.5	.01	77.9	.01	75.0	.01	108.7	.02	49.7	.01	219.3
g_y	.01	8.5	.03	3.2	.02	5.7	.09	1.6	.08	2.2	.03	6.7	.12	2.0	.04	6.6	.21	1.7

Note: Sensitivity and collinearity components of the theoretical Cramér-Rao lower bounds for $\frac{std(\hat{\theta}_i)}{|\theta_i|}$. See Table 4 for the values of the bounds.

Table 7: Multiple collinearity coefficients

	deciles								
Parameter	1	2	3	4	5	6	7	8	9
ρ_{ga}	.34889	.16911	.69704	.17516	.22522	.31300	.18205	.32757	.77298
$\bar{l}$	.97044	.92587	.68036	.99621	.76204	.25033	.95773	.99490	.99740
$\bar{\pi}$	.54484	.86722	.60090	.90639	.99382	.99214	.99552	.99593	.99352
$100(\beta^{-1}-1)$	.99538	.97078	.98379	.98925	.98118	.97688	.98323	.99914	.99325
μ_w	.50072	.96059	.99391	.96009	.99774	.97490	.99625	.99789	.91361
μ_p	.94491	.98677	.99080	.97378	.99512	.96644	.96720	.98413	.97724
α	.85036	.98349	.99296	.94035	.95598	.98922	.98086	.98982	.96113
ψ	.93498	.96368	.98160	.99669	.99791	.95240	.94092	.95348	.99189
φ	.98813	.96432	.95866	.95105	.99420	.99417	.99759	.98821	.97834
σ_c	.99703	.99833	.99895	.99530	.99531	.99826	.99922	.99114	.99719
λ	.99167	.96588	.98390	.99334	.99511	.98506	.92582	.98503	.94037
Φ	.99587	.98304	.96894	.82599	.98843	.98806	.96697	.96600	.98978
ι_w	.73874	.91695	.73397	.88002	.58603	.68123	.97184	.95831	.79557
ξ_w	.99982	.99997	.99361	.99799	.99977	.99994	.99906	.99992	.99999
ι_p	.98916	.76545	.84639	.93685	.79933	.97810	.75254	.86668	.96204
ξ_p	.89210	.87618	.95615	.97080	.84320	.87586	.88533	.92686	.97368
σ_l	.95108	.99245	.99854	.98950	.99264	.98898	.98160	.97228	.98786
r_π	.97923	.95181	.98620	.98655	.94923	.97722	.98004	.93752	.96133
$r_{\triangle y}$	.99781	.94761	.97687	.82413	.82739	.95530	.75649	.74898	.94952
r_y	.97510	.99826	.84210	.93825	.95173	.99687	.97382	.96229	.99145
ρ	.95539	.99536	.96698	.99678	.99754	.99504	.94792	.97051	.98018
ρ_a	.56234	.51164	.50296	.69052	.88490	.10987	.15622	.13278	.58544
ρ_b	.99779	.95540	.96688	.99596	.98996	.94153	.91114	.89153	.94204
ρ_g	.59524	.57312	.38481	.47500	.38962	.37211	.24060	.08117	.16934
ρ_I	.82942	.80820	.84877	.79079	.82296	.81017	.76222	.82155	.75369
ρ_r	.97655	.81956	.87734	.91687	.08487	.71034	.71245	.22686	.74457
ρ_p	.97723	.96494	.98319	.69400	.99833	.99478	.35835	.99842	.99243
ρ_w	.96515	.98034	.98873	.98957	.98865	.59724	.99208	.98272	.99939
γ	.13764	.27543	.29052	.09467	.11063	.36883	.11248	.23372	.28593
σ_a	.18608	.27205	.35141	.41083	.47799	.53188	.59128	.65188	.73784
σ_b	.63065	.71203	.75355	.78391	.80676	.82588	.83997	.85812	.87947
σ_g	.18026	.24242	.29077	.32056	.36139	.41520	.46464	.52561	.59419
σ_I	.71646	.74876	.76895	.78406	.79465	.80824	.81543	.82595	.84343
σ_r	.16592	.25311	.36349	.45466	.55935	.63216	.71383	.78121	.84929
σ_p	.56814	.62221	.65632	.69057	.71645	.75267	.78881	.82399	.88177
σ_w	.55465	.58871	.63540	.67322	.70626	.74168	.77816	.81717	.86570
δ	.93258	.85432	.88466	.91265	.70615	.87227	.98893	.71292	.87275
λ_w	.99980	.99976	.99977	.99950	.99992	.99991	.99996	.99980	.99999
g_y	.99311	.94928	.98464	.78904	.88773	.98871	.86211	.98859	.82025

Note: The multiple correlation coefficients are defined as $\boldsymbol{\varrho}_i := \text{corr}\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i}, \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{-i}}\right)$, where ℓ is the log-likelihood function. The table shows the values of $\boldsymbol{\varrho}_i$ associated with the collinearity components of the Cramér-Rao lower bounds for $\frac{std(\hat{\theta}_i)}{|\theta_i|}$. See Table 4 for the values of the bounds.

Table 8: Distribution of the multiple correlation coefficients

			deciles								
Parameter	mean	var.	1	2	3	4	5	6	7	8	9
ρ_{ga}	.361	.044	.139	.1798	.2185	.2590	.3072	.36360	.43211	.52862	.68437
$\bar{l}$	.846	.034	.559	.7174	.8189	.8852	.9281	.95588	.97400	.98618	.99398
$\bar{\pi}$	.874	.031	.617	.7812	.8712	.9221	.9529	.97171	.98375	.99147	.99644
$100(\beta^{-1}-1)$	.974	.001	.938	.9607	.9721	.9794	.9847	.98868	.99189	.99463	.99705
μ_w	.940	.011	.810	.9274	.9538	.9687	.9791	.98710	.99291	.99692	.99923
μ_p	.950	.008	.837	.9430	.9654	.9768	.9845	.99015	.99448	.99752	.99935
α	.932	.004	.844	.8911	.9186	.9377	.9522	.96417	.97498	.98413	.99216
ψ	.940	.006	.853	.9084	.9365	.9540	.9664	.97573	.98303	.98879	.99403
φ	.971	.001	.921	.9508	.9676	.9784	.9856	.99069	.99409	.99654	.99834
σ_c	.996	.000	.990	.9935	.9952	.9963	.9972	.99791	.99849	.99898	.99943
λ	.967	.001	.926	.9483	.9615	.9705	.9771	.98247	.98693	.99089	.99468
Φ	.964	.003	.908	.9451	.9630	.9741	.9815	.98689	.99121	.99464	.99735
ι_w	.816	.027	.559	.6763	.7581	.8204	.8671	.90589	.93689	.96157	.98078
ξ_w	.998	.000	.996	.9982	.9991	.9995	.9997	.99984	.99992	.99997	.99999
ι_p	.868	.008	.746	.7889	.8226	.8519	.8799	.90656	.93133	.95449	.97605
ξ_p	.947	.003	.868	.9057	.9310	.9501	.9652	.97659	.98559	.99213	.99675
σ_l	.976	.001	.941	.9629	.9747	.9824	.9878	.99156	.99447	.99668	.99839
r_π	.966	.002	.916	.9455	.9609	.9717	.9795	.98529	.99007	.99387	.99706
$r_{\triangle y}$	.914	.006	.802	.8537	.8877	.9143	.9361	.95472	.96987	.98243	.99219
r_y	.936	.005	.844	.8943	.9244	.9450	.9598	.97147	.98061	.98819	.99428
ρ	.972	.001	.932	.9561	.9688	.9772	.9834	.98829	.99223	.99527	.99777
ρ_a	.406	.068	.097	.1502	.2091	.2761	.3564	.44582	.54993	.66728	.80864
ρ_b	.941	.002	.868	.8961	.9188	.9380	.9538	.96756	.97863	.98708	.99358
ρ_g	.417	.060	.119	.1800	.2409	.3055	.3769	.45616	.55000	.65681	.79032
ρ_I	.815	.005	.739	.7704	.7899	.8051	.8187	.83249	.84744	.86471	.89215
ρ_r	.757	.050	.405	.5686	.6778	.7651	.8313	.88473	.92425	.95407	.97839
ρ_p	.928	.022	.775	.9334	.9614	.9751	.9840	.99021	.99463	.99757	.99936
ρ_w	.915	.030	.726	.9198	.9532	.9696	.9804	.98781	.99327	.99705	.99924
γ	.216	.013	.085	.1156	.1436	.1705	.1993	.22921	.26450	.30713	.36788
σ_a	.467	.039	.198	.2808	.3480	.4097	.4676	.52630	.58750	.65349	.73333
σ_b	.774	.013	.628	.7098	.7542	.7834	.8056	.82382	.84109	.85870	.87965
σ_g	.378	.025	.174	.2314	.2784	.3235	.3689	.41479	.46475	.52058	.59736
σ_I	.783	.004	.711	.7496	.7693	.7831	.7947	.80550	.81615	.82823	.84451
σ_r	.527	.064	.164	.2610	.3592	.4593	.5514	.63537	.71127	.78084	.85021
σ_p	.725	.013	.575	.6198	.6564	.6894	.7203	.75236	.78637	.82653	.88157
σ_w	.706	.014	.539	.5874	.6329	.6724	.7088	.74278	.77900	.81849	.86739
δ	.866	.011	.717	.7941	.8390	.8708	.8958	.91704	.93500	.95230	.97007
λ_w	.999	.000	.997	.9988	.9993	.9996	.9998	.99987	.99993	.99997	.99999
g_y	.912	.009	.777	.8514	.8965	.9260	.9474	.96409	.97665	.98654	.99396

Note: Each row of the table shows the mean, the variance and the nine deciles of $\boldsymbol{\varrho}_i$ computed on the basis of 100,000 draws from $\boldsymbol{\Theta}$.

Table 9: Collinearity in the model

			deciles								
Parameter	mean	var.	1	2	3	4	5	6	7	8	9
ρ_{ga}	.861	.035	.586	.776	.866	.9138	.9429	.9623	.9754	.98516	.99244
$\bar{l}$	.000	.000	.000	.000	.000	.0000	.0000	.0000	.0000	.00000	.00000
$\bar{\pi}$	.682	.001	.658	.673	.680	.6849	.6887	.6917	.6945	.69719	.70022
$100(\beta^{-1}-1)$	.905	.003	.833	.857	.874	.8895	.9039	.9181	.9348	.95668	.98444
μ_w	.469	.025	.253	.330	.387	.4364	.4792	.5204	.5613	.60519	.65870
μ_p	.640	.030	.393	.496	.567	.6210	.6676	.7090	.7479	.78861	.83938
α	.916	.003	.843	.872	.890	.9054	.9208	.9367	.9546	.97203	.98805
ψ	.334	.015	.195	.228	.257	.2861	.3158	.3488	.3866	.43072	.49693
φ	.886	.008	.763	.812	.846	.8744	.9003	.9250	.9489	.97141	.99043
σ_c	.998	.000	.995	.997	.998	.9988	.9992	.9995	.9998	.99990	.99998
λ	.996	.000	.989	.995	.997	.9982	.9989	.9994	.9997	.99988	.99997
Φ	.946	.001	.900	.919	.932	.9428	.9519	.9605	.9688	.97691	.98552
ι_w	.896	.007	.782	.827	.858	.8842	.9070	.9295	.9527	.97573	.99299
ξ_w	.997	.000	.992	.996	.998	.9988	.9993	.9996	.9998	.99993	.99999
ι_p	.912	.005	.818	.854	.879	.9009	.9204	.9402	.9605	.97983	.99431
ξ_p	.971	.001	.934	.953	.965	.9725	.9787	.9838	.9882	.99244	.99671
σ_l	.979	.000	.955	.967	.974	.9790	.9834	.9873	.9910	.99466	.99808
r_π	.981	.001	.949	.970	.981	.9874	.9919	.9950	.9972	.99879	.99968
$r_{\triangle y}$	.823	.023	.602	.701	.769	.8206	.8632	.8989	.9294	.95656	.98046
r_y	.971	.001	.924	.952	.967	.9771	.9844	.9899	.9940	.99701	.99909
ρ	.962	.002	.897	.935	.956	.9704	.9805	.9878	.9931	.99684	.99914
ρ_a	.659	.019	.479	.554	.602	.6406	.6756	.7062	.7371	.77071	.81490
ρ_b	.944	.010	.850	.924	.954	.9701	.9810	.9882	.9935	.99704	.99918
ρ_g	.560	.032	.328	.406	.463	.5125	.5597	.6083	.6573	.71274	.79002
ρ_I	.460	.054	.172	.244	.308	.3676	.4289	.4968	.5785	.67938	.80756
ρ_r	.704	.067	.302	.442	.567	.6757	.7726	.8533	.9140	.95911	.98746
ρ_p	.783	.031	.517	.632	.714	.7784	.8301	.8738	.9115	.94282	.97109
ρ_w	.595	.043	.318	.398	.466	.5323	.5966	.6601	.7239	.79130	.87359
γ	.678	.023	.494	.551	.591	.6269	.6605	.6964	.7386	.80213	.92244
σ_a	.789	.047	.441	.638	.750	.8239	.8745	.9113	.9386	.96060	.97850
σ_b	.926	.007	.800	.859	.901	.9325	.9580	.9761	.9878	.99464	.99840
σ_g	.814	.034	.529	.685	.778	.8372	.8808	.9136	.9391	.96047	.97843
σ_I	.409	.032	.218	.263	.300	.3351	.3713	.4117	.4607	.52932	.65880
σ_r	.859	.008	.734	.782	.816	.8448	.8701	.8944	.9175	.94171	.96897
σ_p	.533	.023	.351	.408	.448	.4851	.5201	.5570	.5990	.65165	.73532
σ_w	.331	.019	.195	.230	.257	.2815	.3059	.3325	.3638	.40613	.48614
δ	.875	.008	.749	.804	.840	.8674	.8911	.9128	.9341	.95637	.97706
λ_w	.998	.000	.993	.996	.998	.9988	.9993	.9996	.9998	.99993	.99998
g_y	.887	.004	.812	.836	.854	.8697	.8848	.9006	.9189	.94305	.97384

Note: The collinearity in the model is measured by $\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}}\right)$, where $\boldsymbol{\tau}$ is a vector of parameter describing the solution of the structural model. Each row of the table shows the mean, the variance and the nine deciles of $\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}}\right)$ computed on the basis of $100,000$ draws from $\boldsymbol{\Theta}$.

Table 10: Parameters with strongest collinearity in the model

worst identified parameters	Quantiles of $\cos\left(\frac{\partial \boldsymbol{\tau}}{\partial \theta_i}, \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}_{-i}}\right)$ 0	.2	.4	.6	.8	1	functionally similar 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
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

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.

Andrews, D. W., and J. H. Stock (2005): "Inference with Weak Instruments," Cowles Foundation Discussion Papers 1530, Cowles Foundation, Yale University.

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.

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.

Guerron-Quintana, P. A., A. Inoue, and L. Kilian (2009): "Frequentist Inference in Weakly Identified DSGE Models," CEPR Discussion Papers 7447, C.E.P.R. Discussion Papers.

Inoue, A., and B. Rossi (2008): "Testing for Weak Identification in Possibly Nonlinear Models," unpublished manuscript.

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.

King, R. G., and M. W. Watson (1998): "The Solution of Singular Linear Difference Systems under Rational Expectations," *International Economic Review*, 39(4), 1015–26.

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.

Komunjer, I., and S. Ng (2009): "Dynamic Identification of DSGE Models," unpublished manuscript.

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.

SCHRAF, C. (1991): *Statistical Signal Processing*. Addison-Wesley Publishing Company, Inc.

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.

STOCK, J. H., AND M. YOGO (2005): *Testing for weak instruments in linear IV regression*chap. 5. Cambridge University Press.

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.