Essays On Identification And Estimation of Dynamic Stochastic General Equilibrium Models

by

Nikolay Ivanov Iskrev

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2008

Doctoral Committee:
Professor Serena Ng, Chair
Professor Robert B. Barsky
Professor Fred M. Feinberg
Assistant Professor Christopher L. House

to Tenzin Palmo

Acknowledgments

In writing this dissertation, I have benefitted from numerous discussions with my advisor, Serena Ng, and the other members of my dissertation committee - Bob Barsky, Chris House, and Fred Feinberg. I am indebted to all of them for their advice, guidance and patience.

I would also like to thank Alejandro Justiniano and Marco Ratto who were very very generous and encouraging discussants of earlier drafts of my job market paper, which eventually turned into Chapter 4 of this dissertation.

I am grateful to Matthew Shapiro, Giorgio Primiceri, Òscar Jorda, Tao Zha, Jinill Kim and participants at the Cleveland Fed DSGE models workshop, Third Dynare conference, and University of Michigan Macroeconomics Seminar for helpful comments and discussions.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Appendices

Appendix

Abstract

The estimation of dynamic stochastic general equilibrium (DSGE) models is the subject of a rapidly growing literature. This dissertation contributes to the existing body of work by focusing on issues related to parameter identification. Identification deals with the question of whether distinct parameter values lead to distinct probability distributions of the data. Lack of identification implies that consistent estimation is impossible. Therefore, verifying that the parameters of interest are identifiable is a condition that has to be checked before estimation is attempted. This is particularly important in the estimation of DSGE models because of their increasing role as a tool of economic policy. As such, DSGE models are useful only as long as their parameter values are empirically relevant. The main contribution of this dissertation is making explicit the link between the likelihood of the statistical model and the structure of the underlying theoretical macroeconomic model, and showing how this can be used in the empirical evaluation of DSGE models.

In the first essay I show that DSGE models are fully characterized by a set of cross-equation and covariance restrictions. Taking this as a starting point, I show how the restrictions can be used to determine the identifiability, and estimate the parameters of such models. I derive identification conditions with respect to the structural parameters, for which global identification may be established, and the deep parameters, for which identification can be only local. I also propose a two-step minimum distance method for estimating the parameters of DSGE models. I show that the estimator is asymptotically efficient, and provide simulation evidence

showing that it has good small sample properties.

In the second essay I show how the Information matrix of any DSGE model can be evaluated analytically. This is achieved by a factorization of the matrix as a product of two terms: the Jacobian matrix of the mapping from deep to reduced-form parameters, and the Information matrix of the reduced form model. I show that both terms can be derived analytically. This result is useful for the estimation of DSGE models, both in the classical and the Bayesian tradition, as well as for detecting identification problems.

In the third essay I develop a general methodology for analyzing parameter identification in linearized DSGE models. Specifically, I show how to address the following questions: first, are the parameters of the model identifiable; second, how strong is identification; third, if there are identification problems, do they originate in the model or in the data; and fourth, which parameters are not well-identified and what features of the model are responsible for that. I apply this methodology to study parameter identification of a model estimated recently in Smets and Wouters (2007). I find that identification is generally very weak, and the problems are largely embedded in the structure of the model, and thereby cannot be resolved by using more informative data. I estimate the model with maximum likelihood, and find substantial differences in the parameter estimates compared with those obtained with Bayesian methods. I conclude that the use of estimated DSGE models for policy analysis should be done with caution since, when identification is weak, the results are likely to be strongly influenced by the prior distribution.

Chapter 1

Introduction

Dynamic stochastic general equilibrium (DSGE henceforth) models are a standard tool of modern macroeconomics. Such models are used to study the determinants of the main economic aggregates - consumption, savings, and investments, as well as to analyze alternative economic policies. Unlike the macroeconomic models used in the 1960s and the 1970s, DSGE models are based on maximizing behavior of economic agents, and deal explicitly with uncertainty and agents' expectations. These features of the models are represented by a set of parameters, known as deep parameters. Because of their increasing complexity, however, DSGE models are rarely susceptible to analytical solution, and it is typically impossible to make general statements about the predictions of a model, that would be independent from the parameter values. Where these values come form is therefore a key question when evaluating the empirical and policy relevance of macroeconomic models.

Until recently complete DSGE models were never confronted with the data directly. Instead, either reduced-form vector autoregressions or single Euler equations were estimated and used for testing particular features of the underlying economic models. This has now changed, and in the last several years there has been a remarkable growth in the research on empirical evaluation of complete DSGE models. Nowadays researchers routinely estimate rich micro-founded models, that until re-

cently had to be calibrated. Unlike the reduced-form or single equation estimation methods, the full set of model parameters are being estimated in an internally-consistent fashion. This, together with the finding that empirical DSGE models can fit the data as well as model-free reduced-form vector autoregressions (VAR), has made them extremely popular in central banks and other policy-making institutions. Quite large and sophisticated DSGE models are being developed, estimated, and used for policy analysis in institutions such as the Federal Reserve Board, the European Central Bank, Bank of England, RiksBank, the Bank of Canada, and the IMF.

A question that is rarely addressed in the empirical DSGE literature is that of parameter identification. This is surprising as identifiability is a prerequisite for estimation of parameters of any structural model, and the ability to do that for fully articulated macroeconomic models is considered to be one of the main accomplishments of this line of research. That parameter identification is a potentially serious issue for DSGE models is not a new concern. Among the authors who have made this point are Sargent (1976) and Pesaran (1989). More recently Beyer and Farmer (2004) provide several examples of commonly used models that are unidentifiable. They argue that the problem is likely to be common in DSGE models.

In most empirical DSGE papers the question of parameter identification is not confronted directly. Usually, if some of the parameters are considered to be of lesser interest, and/or with potentially problematic identifiability, their values are calibrated and assumed known, instead of being estimated. Furthermore, since DSGE models are frequently estimated using Bayesian methods, potential identification problems remain hidden due to the use of priors. As a result, it is often unclear to what extent the reported estimates reflect information in the data instead of subjective beliefs or other considerations reflected in the choice of prior distribution for the parameters.

One reason why this is an important issue is that DSGE models are increasingly being used for analyzing policy-relevant questions, such as, for instance, the design of

optimal monetary policy. Such analysis often hinges crucially on the values assigned to the parameters of the model. It is, therefore, important to know how informative the data is for the parameters of interest, and whether there are any benefits from estimating instead of calibrating the models we use to address policy questions.

The main objective of this dissertation is to develop a methodology for studying identification issues in DSGE models. I begin in Chapter 2 by showing that any linearized DSGE model is completely characterized by a set of cross-equation and covariance restrictions on the parameters of the reduced-form solution of the model. Taking this as a starting point, I show how these restrictions can be used to analyze the identifiability, and also estimate the parameters of such models. An important simplifying assumption I make in this chapter is that the parameters of the reduced-form are identifiable, and therefore can be estimated directly. Under that assumption, I derive conditions that are both necessary and sufficient for identification. Moreover, I show that the relationship between the reduced-form and the deep parameters can be used to estimate the latter using a two-step minimum distance procedure. Unlike other minimum-distance estimators used in the DSGE literature, the estimator is asymptotically equivalent to full information maximum likelihood estimation, and is thus efficient. I also present Monte Carlo evidence showing that the estimator has good small sample properties.

While valid for some macroeconomic models, the assumption that the parameters of the reduced-form linear state space model are identifiable is too strong for many of the models estimated in the current empirical DSGE research. It does not hold for a general unrestricted linear state space model, and while linearized DSGE models often imply a large number of zero restrictions on the reduced-form, it is frequently very difficult to determine whether those restrictions are sufficient to guarantee identifiability. For that reason an alternative approach to the question of identification is required for general DSGE models. This leads us to Chapter 3 where, building upon

results from Chapter 2 I show how the Information matrix for any linearized DSGE model can be evaluated analytically. Non-singularity of the Information matrix is a necessary and sufficient condition for identification, a result shown in Rothenberg (1971). Apart from its use for identification analysis, which I pursue in Chapter 4, the Information matrix is important for estimation and inference, both in the classical and the Bayesian tradition. For instance, the asymptotic covariance matrix of the maximum likelihood estimator is given by the inverse of the Information matrix. Similarly, Bayesian methods such as Metropolis algorithm or Importance sampler, use the Information matrix to draw from the posterior distribution.

Knowing how to compute the Information matrix allows us to determine the identifiability of the parameters of any parametric model. In Chapter 4 I use the result from Chapter 3 to study identification of a model estimated recently in Smets and Wouters (2007). That model was selected because of its prominence and influence in the empirical DSGE literature. Furthermore, many of the other models estimated in the literature have a lot of features in common with the Smets and Wouters model. Using the Information matrix approach, I determine that the parameters of the model are generally identifiable, including a parameter previously believed to be unidentified.

The non-singularity of the Information matrix only guarantees that the parameters are identifiable in the strict sense, that is, asymptotically. In practice, however, the data available for estimation are relatively short, and it is thus very important to know how strong identification is. This matters greatly for the precision of the parameters estimates, and the reliability of the standard methods for constructing confidence intervals, testing hypothesis and inference in general. When poorly identified models are estimated using Bayesian methods, the estimation results are strongly influenced by the specification of the prior distribution. It Chapter 4 I argue that much can be learned about the strength of identification, as well as the causes for identification problems, prior to taking a model to the data. Such problems arise, for

instance, when different parameters play very similar roles in the model. I propose measures of the degree of similarity, and show how the problematic parameters can be determined. Applying this methodology to the Smets and Wouters model, I find that many of its parameters are very poorly identified. Finally, I estimate the model using maximum likelihood and compare the results to those reported in Smets and Wouters (2007) and obtained using Bayesian methods. The results show that the specification of the prior distribution has a strong influence on the estimation results.

Chapter 2

Identification and Estimation of DSGE models - an integrated approach

2.1 Introduction

The last several years have witnessed a remarkable growth in the research on empirical evaluation of DSGE models. Nowadays researchers routinely estimate rich micro-founded models, that until recently had to be calibrated. Unlike reduced-form or single equation estimation methods, the full set of model parameters are being estimated in an internally-consistent fashion. This, together with the finding that empirical DSGE models can fit the data as well as model-free reduced-form vector autoregressions (VAR), has made them extremely popular in central banks and other policy-making institutions.

A question that is rarely addressed in the empirical DSGE literature is that of parameter identifiability. This is surprising as identification is a prerequisite for estimation of the parameters, and the ability to do that for full-fledged structural models is believed to be one of the main accomplishments of this line of research. That parameter identification is a potentially serious issue for DSGE models is not a new

concern. Among the authors who have made this point are Sargent (1976) and Pesaran (1989). More recently Beyer and Farmer (2004) provide several examples of commonly used models that are unidentifiable. They argue that the problem is likely to be common in DSGE models.

In most empirical DSGE papers the question of parameter identification is not confronted directly. Usually, if some of the parameters are considered to be of lesser interest, and/or with potentially problematic identifiability, their values are calibrated and assumed known, instead of being estimated. Furthermore, since DSGE models are frequently estimated using Bayesian methods, potential identification problems remain hidden due to the use of priors. As a result, it is often unclear to what extent the reported estimates reflect information in the data instead of subjective beliefs or other considerations reflected in the choice of prior distribution for the parameters. One reason why this is an important issue is that DSGE models are increasingly being used for analyzing policy-relevant questions, such as, for instance, the design of optimal monetary policy. Such analysis often hinges crucially on the values assigned to the parameters of the model. It is, therefore, important to know how informative the data is for the parameters of interest, and whether there are any benefits from estimating instead of calibrating the models we use to address policy questions.

In general, DSGE models have the following form:

$$E_t J(\tilde{Z}_{t+1}, \tilde{Z}_t, \tilde{Z}_{t-1}, U_t; \theta) = 0 \tag{2.1}$$

where J is a non-linear function of the endogenous variables $\tilde{Z}$, and the exogenous shocks U, and θ is a vector of deep parameters. Since the model in (2.1) is, for most purposes, too difficult to work with, researchers use a linear or log-linear approxi-

mation of (2.1) around the steady state. The resulting system of linear stochastic equations is of the form

$$E_t \hat{J}(Z_{t+1}, Z_t, Z_{t-1}, U_t; \gamma) = 0 \tag{2.2}$$

where Z is the log-deviation of $\tilde{Z}$ from its steady-state level, and $\hat{J}$ is function linear in the variables Z and U, as well as in the vector of structural parameters γ. The mapping from the deep parameters θ to the structural parameters γ, is given by some known, usually non-linear function. Solving the linearized version of the DSGE model yields a reduced-form model, given by

$$R(Z_t, Z_{t-1}, U_t; \tau) = 0 \tag{2.3}$$

where R is a linear function of Z and U, parameterized by the $m \times 1$ vector of reduced-form parameters τ.

In this paper I present an integrated approach for determining identifiability and estimating the parameters of linearized DSGE models, by using the restrictions that the linearized model (2.2) imposes on the reduced form (2.3). In particular, I present necessary and sufficient conditions for local identification of θ, and discuss when its global identifiability can be ascertained. The conditions for partial identification, that is, when only a subset of all parameters is of interest, or some of the parameters are unidentifiable, are also examined. Second, I present a minimum distance estimator of the parameters in DSGE models, derive its asymptotic properties, and compare its performance in small samples with that of the maximum likelihood estimator by using Monte Carlo simulations.

Several authors have studied identification issues in models with rational expectations. Wallis (1980) provides a rank condition for identification of a system of

simultaneous equations with current expectations [1] of the endogenous variables. Pesaran (1981) and Wegge and Feldman (1983) extend the Wallis' analysis by allowing for more general identification restrictions, and for models with lagged exogenous variables. In addition they state the rank condition for identification in terms of the structural coefficients, unlike Wallis (1980), where it is given in terms of the reduced-form coefficients. Only Wegge and Feldman (1983) considers identification by a priori restrictions on the structural covariance matrix. Pesaran (1989) provides a rank condition for identification of a single equation in a system with future expectations. In all of the above papers identification conditions are derived by determining when the mapping from reduced-form to structural parameters is unique, which is also the approach in the present paper. Unlike those studies however, where identification only of the structural parameters γ is considered, here I present conditions for identifiability for both the structural coefficients and the underlying deep parameters θ, which is typically what DSGE modelers are interested in estimating. Moreover, I present identification conditions for models with future expectations and covariance restrictions, which is also in line with the models in the current DSGE literature.

On the estimation side, this paper has predecessors both in the literature on estimation of standard linear simultaneous equations, as well as that on estimating models with rational expectations. Estimating the reduced-form parameters first, and then solving for the structural ones, is known as indirect least squares, in the simultaneous equations literature (see Judge, Griffiths, Hill, Lutkepohl, and Lee (1982, chap. 15) for a textbook treatment of ILS), and classical minimum distance estimator (see Wooldridge (2002, chap. 14)) in general econometrics. Sargent (1979) was the first to point out that rational expectations lead to cross-equation restrictions which can be used to estimate structural coefficients. Sbordone in several papers (Sbordone,

[1] Models with current expectations contain predictions for the current value of the endogenous variable, made in the previous period. Models with future expectations contain expectations about future values of the variables, formed in the current period.

2002, 2005a,b) estimates models of inflation dynamics using a two-step minimum distance procedure, based on the cross-equation restrictions on the reduced-form VAR representation of the data. Li (2004) extends this estimation approach to models with time variation in the reduced-form coefficients, and applies it to the estimation of investment Euler equation. The approach in these papers is a limited information one since not all of the cross-equation restrictions, or those on the structural covariance, implied by the DSGE model are utilized. In contrast, in this paper I use a two step minimum distance procedure based on all restrictions implied by the underlying general equilibrium model.

The remaining of this chapter is organized as follows. In section 2 I introduce the general model and notation. I also discuss parameter identification in general terms, and introduce a result from Rothenberg (1971) which will serve as a basis for my approach to identification of DSGE models. In Section 2 I present conditions for identifiability of the parameters in DSGE models. I follow a two step approach in which the identification of the deep parameters θ is conditioned first on the identifiability of the structural parameters γ, and then on the uniqueness of the mapping from γ to θ. I also make a distinction between identification based on the cross-equation restrictions only, and identification when covariance restrictions are used as well. Condition for identification are given for the complete system of equations as well as for the parameters in a single equation. Finally, I discuss how identification failures can be detected in practice, and describe an algorithm for a thorough identification analysis of DSGE models. In section 2 I introduce a full information minimum distance estimator of the parameters in DSGE models, and discuss its asymptotic properties. Section 2 concludes.

2.2 Preliminaries

2.2.1 Structural and Reduced Form of DSGE Models

A typical DSGE model is represented by a system of non-linear equation of the form in equation (2.1). Currently, most analyses involving either simulation or estimation of DSGE models use linear approximations the original models. Specifically, the model is first expressed in terms of stationary variables, and then linearized or log-linearized around the steady-state values of these variables.

Typically, the linearized system can be written in the form

$$\Gamma_0 Z_t = \Gamma_1 E_t Z_{t+1} + \Gamma_2 Z_{t-1} + \Gamma_3 U_t \tag{2.4}$$

where Z_t is a $m \times 1$ vector of endogenous variables, and the structural errors, U_t, are i.i.d. n-dimensional random vectors with $E[U_t] = 0,\ E[U_t U_t'] = I$. The coefficient matrices $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3 are functions of the $k \times 1$ vector of deep parameters θ.

I use the term *structural parameters* for the components of the coefficient matrices $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3. The a priori unknown elements of $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ are collected in the $l \times 1$ vector γ_1, while γ_2 is a $n \times 1$ vector collecting the a priori unknown components of Γ_3. Also, let $\gamma = [\gamma_1',\ \gamma_2']'$ be a vector of all structural parameters. The structural parameters in turn depend on *deep parameters*, reflecting properties of the original DSGE model such as preferences or technologies. I denote the deep parameters with θ, and will assume that the mapping from θ to γ is given by the function $\tilde{g}$, i.e.

$$\gamma = \tilde{g}(\theta) \tag{2.5}$$

I distinguish between the model in (2.4), when it is written in terms of the structural parameters γ, and the same model when the parameters of interest are θ . I use the term "structural model" in the first case, and "behavioral model" in the second.

[2]

In most models the structural shocks are assumed to be independent, and thus Γ_3 is a diagonal matrix with the standard deviations of the shocks in the diagonal. Therefore there is no distinction between the deep parameters that enter in Γ_3 and the structural parameters in γ_2. Hence we can write $\theta = [\theta_1', \theta_2']'$, where $\theta_2 = \gamma_2$. This implies that the function $\tilde{g}$ can be written as $\tilde{g} = \begin{bmatrix} g \\ I_n \end{bmatrix}$, where g maps θ_1 into γ_1,

$$\gamma_1 = g(\theta_1) \tag{2.6}$$

and I_n is a $n-$dimensional identity matrix.

There are several algorithms for solving linear rational expectations models like (2.4) (see for instance Blanchard and Kahn (1980), Anderson and Moore (1985), Klein (2000), Christiano (2002), Sims (2002)). Depending on the value of θ, there may exist zero, one, or many stable solutions. Assuming that a unique solution exists, it can be cast in the following form

$$Z_t = AZ_{t-1} + BU_t \tag{2.7}$$

where A and B are functions of θ, and are unique for each value of θ. I let $\Omega = BB'$, and define $\tau_1 = \mathbf{vec}(A)$, and $\tau_2 = \mathbf{vech}\Omega$. Lastly, let $\tau = [\tau_1', \ \tau_2']'$ be the vector of all *reduced-form parameters.*

Throughout the chapter I will assume that the linearized model (2.4) has a unique reduced-form solution (2.7). I state this as

Assumption 2.2.1 (Uniqueness of the solution)**.** *The vector of reduced-form parameters τ is unique for each admissible value of θ.*

[2]I follow Chris Sims in distinguishing between "structural" and "behavioral" model. By *structural model* I mean a model invariant to policy interventions. The *behavioral model*, in contrast, is the one where the parameters have behavioral interpretations.

It is important to note that this assumption does not mean that the inverse mapping, from τ to θ, is unique. In fact, as I will explain shortly, the identifiability of the deep parameters θ is implied by the uniqueness of this mapping.

In some cases it will be convenient to assume that the structural shocks U_t are jointly normally distributed. I state this assumption below

Assumption 2.2.2 (Normality)**.** $U_t \sim N(0, \Sigma)$*, where* $\Sigma = \Gamma_3 \Gamma_3'$*, for* $t = 1, \ldots, T$*.*

The three models and the parameters associated with them are summarized in table below.

Model	Parameters (dimension)			Parameter space
	mean	covariance	All	
Behavioral	θ_1 $(k \times 1)$	θ_2 $(n \times 1)$	$\theta = [\theta_1', \ \theta_2']'$	$\Theta \in \mathbb{R}^{k+n}$
Structural	γ_1 $(l \times 1)$	γ_2 $(n \times 1)$	$\gamma = [\gamma_1', \ \gamma_2']'$	$\mathcal{S}_\gamma \in \mathbb{R}^{l+n}$
Reduced-form	τ_1 $(m^2 \times 1)$	τ_2 $(m^2 \times 1)$	$\tau = [\tau_1', \ \tau_2']'$	$\mathcal{S}_\tau \in \mathbb{R}^{2m^2}$

Table 2.1: Models and parameters

Other notations that I use in the chapter are:

- function f defining implicitly the mapping from the deep parameters θ, to the reduced-form parameters τ

  $$f(\theta, \tau) = 0$$

  In addition, I will use the function f^* to denote the mapping between the structural coefficients γ and τ

  $$f^*(\gamma, \tau) = f(g(\theta), \tau) = 0$$

- In certain occasions it will be convenient to assume that an explicit function mapping the elements of θ into those of τ is available; I use h_1 and h_2 to denote the functions mapping θ into τ_1 and τ_2, respectively, i.e.

  $$\tau_1 = h_1(\theta)$$

  and

  $$\tau_2 = h_2(\theta)$$

  and will define $h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$, so that

  $$\tau = h(\theta)$$

In addition, I define the following three vectors: $Z = \begin{bmatrix} Z_2 & \dots & Z_T \end{bmatrix}$, $X = \begin{bmatrix} Z_1 & \dots & Z_{T-1} \end{bmatrix}$, and $V = \begin{bmatrix} V_2 & \dots & V_T \end{bmatrix}$

2.2.2 Identification

To make precise the language I use when discussing identification, I provide a few definitions of terms that will be used in the next section. For a more details see Rothenberg (1971).

Let Z be a vector valued random variable with a parametric distribution function $\mathfrak{F}$. A *structure* S is a parametric point which implies a unique distribution function $\mathfrak{F}(S)$. A *model* is the set of all admissible structures. A structure S is *globally identifiable* if no other structure in the model implies the same probability distribution, that is, if whenever $S \neq S_1$, we have $\mathfrak{F}(S) \neq \mathfrak{F}(S_1)$. A structure S is *locally identifiable* if there exists an open neighborhood of S such that for every structure S_1 in it, if $S \neq S_1$, then $\mathfrak{F}(S) \neq \mathfrak{F}(S_1)$. Structures that induce the same probability distribution are called *observationally equivalent*. If all structures in a model are identifiable, the model is *identified*.

The above definition implies that in principle a structure must include all parameters that completely characterize the distribution of the data. Unless we do so, we can never conclude that a model is not identified. In what follows I am going to use only parameters that specify the first two moments of the data. Thus the identification conditions that I find are in general only sufficient, and not necessary. However, if it is assumed that the data is generated from a normal distribution, then the conditions become necessary and sufficient for identifiability. The reason is that a normal distribution is completely described by its first two moments [3].

[3]Although it is not required in the estimation method used in this chapter, the normality assumption is very common in the literature on estimation of DSGE models, most of which follows the likelihood approach

With the last remark as caveat, I define structures for each of the three models as follow:

1. Behavioral model - a structure is a known vector θ;
2. Structural model - a structure is a known vector γ;
3. Reduced-form model - a structure is a known vector τ.

The focus of this chapter is on θ and γ - the parameters of the behavioral model and the structural model respectively. However, I will not deal with the identification of the reduced-form model, and will assume that it is globally identified. I state this as follows

Assumption 2.2.3. *The reduced-form coefficients A and Ω are globally identifiable for all admissible values of the parameters θ.*

In a seminal paper Rothenberg (1971) provides a general condition for identification of parametric models. The condition is that the information matrix must be non-singular at the true value of the parameters. This result can be applied for detecting identification problems for other extremum estimators as well, and not only likelihood-based methods. The general criterion for identifiability is that the Hessian of the objective function is of full rank. Due to its generality, however, the information matrix approach has some important drawbacks. First, in many situations, and in particular in the case of DSGE models, the Hessian is very difficult, and often impossible to obtain in analytical form. As a result, researchers have to resort to numerical methods to compute second derivatives, which inevitably leads to inaccuracy in the results. Moreover, this approach is in most cases limited to local identification only, and therefore does not allow researchers to determine whether their model is globally or only locally identified. And finally, due to the intricacies involved in obtaining the Hessian, it is usually hard to pinpoint the source of identification problems in the underlying model.

Another, less general result in Rothenberg (1971) provides the basis for an alternative method for determining identification in parametric models, without involving

the information matrix [4]. The approach is based on the relationship between the parameters of interest, and characteristics of the probability distribution of the data, and boils down to a question of uniqueness of a solution to a system of equations. The well known rank and order conditions for identification of systems of linear simultaneous equations are an example of this approach to identification. A very useful feature of this approach, when applied to DSGE models, is that it allows for a more transparent and intuitive interpretation of the necessary identification condition, and makes it easier for researchers to locate sources of potential identification problems. Moreover, this identification framework extends in a natural way into an estimation procedure that could potentially be useful for empirical validation of DSGE models. I will return to the last observation later on in the chapter.

To state Rothenberg's result formally, let h be a continuously differentiable function, such that

$$h(\theta) = \gamma, \tag{2.8}$$

and denote with H its Jacobian, that is $H = \frac{\partial h}{\partial \theta}$. Then we have [5]

Theorem. *Suppose that the density of Z depends on the parameter vector θ only through the reduced-form parameter γ, and assume that γ is globally identifiable. Then a structure θ_0 is locally identifiable if the Jacobian H, evaluated at θ_0, has a full column rank. If θ_0 is a regular point of H, this condition is also necessary.* [6] *If h is linear, then θ_0 is globally identifiable.*

The intuition behind this result is straightforward. Since the probability distribution of the data depends on θ only through γ, the identifiability of θ is equivalent

[4] Rothenberg (1971) was not the first to suggest this identification approach. The classical treatment of identification of systems of linear simultaneous equations in Koopmans and Reiersol (1950) and Fisher (1966) was based on the same idea

[5] see Rothenberg (1971, Assumption 7 and Theorem 6)

[6] A regular point of the matrix is a point around which there exists an open neighborhood where the rank of the matrix remains constant. For the Jacobian matrices of analytic functions the set of irregular points has a measure of zero, and thus almost all points are regular.(see Bekker and Pollock (1986)).

to whether knowing γ allows for determining uniquely the value of θ. Another way to say this, is that (2.8) has a unique solution. For a general function h, we can only find out whether the solution is locally unique, by determining the rank of H. However, if the function is linear, the rank condition is necessary and sufficient for global uniqueness.

The identification approach in this chapter is based on the above theorem. However, to make it more suitable for our purpose, I present two straightforward extensions of Theorem 2. First, I relax the assumption that the data density depends on θ through γ *only*. The consequence of doing that is that in general the rank condition in the theorem becomes only sufficient for identification of θ, and not necessary. Second, I generalize the above result to account for situations where the mapping between the parameters in θ and those in τ is given by an implicit function, such as

$$f(\theta, \tau) = 0 \tag{2.9}$$

instead of the explicit function in (2.8). Let f_θ be the Jacobian of f with respect to θ. Then we have the following

Results. *Suppose that the density of Z depends on the parameter vector θ through the reduced-form parameter τ, and assume that τ is globally identifiable. Assume also that the mapping from θ to τ is defined implicitly as in* (2.9)*. Then,*

(R1) a structure θ_0 is locally identifiable if the Jacobian f_θ, evaluated at θ_0, has a full column rank.
(R2) The last condition is also necessary for identification of θ_0, if the density of Z depends on θ only through τ, and θ_0 is a regular point of f_θ.
(R3) If f_θ has a full column rank for all values of θ, then θ_0 is globally identifiable.

The first result (R1) tells us that the existence of a locally unique mapping from τ to θ is a sufficient condition for the identification of θ. When the data density depends on θ only through τ, the latter is also a necessary condition, as implied by the second result (R2). Otherwise it is not. I will make use of this distinction in the following way. A density function is completely determined by its moments. If τ

contains the parameters necessary to characterize all moments of the data, then we are in a situation when R2 is applicable, and the rank condition is both necessary an sufficient. However, we often cannot use all moments, either because some of the higher moments are difficult to estimate precisely, or because we do not want to assume a particular density function, and, therefore, do not know what those moments are. If that is the case, we can still use the fact that, if some of the moments we do use imply a unique value of θ, then this is sufficient for identification. As an example, suppose that τ contains all parameters that uniquely determine the first two moments, and that τ is globally identified. If we are willing to assume that the data is normally distributed, then the rank condition would be necessary and sufficient for θ to be identifiable because the normal density is completely determined by the first two moments. Otherwise, it is only a sufficient one, because θ may me unidentifiable from the first two moments, but identified with the higher moments.

The last result (R3) is useful when the mapping from the reduced form parameters τ to the parameters of interest θ is linear. Then, if the rank condition holds, the solution for θ will be globally unique. This result is analogous to Theorem 5 in Rothenberg (1971).

Presenting these results for an implicit instead of explicit function makes them convenient for studying the identification of the parameters in DSGE models. As I discuss later in the chapter, in these models it is usually impossible to find an explicit function h as in (2.8), and easy to find an implicit function f as in (2.9).

2.3 Identification in Linear Rational Expectations models

As I indicated in the introduction, the problem of parameter identification has been largely neglected in the current empirical DSGE literature. There are two main rea-

sons for that. The first one is that, because the mapping from structural to reduced form parameters is extremely complicated, and, except for very special cases, not available in an analytical form, using the Iinformation matrix criterion for identification is very difficult. In the words of Lubik and Schorfheide "It is difficult to directly detect identification problems in large DSGE models, since the mapping from the vector of structural parameters - into the state-space representation that determines the joint probability distribution of Z is highly nonlinear and typically can only be evaluated numerically" (Lubik and Schorfheide, 2004). Computing the Information matrix requires the second derivative of the log-likelihood function which is cannot be obtained directly if the function mapping structural into reduced-form parameters is not available.[7] The second reason is that the large majority of papers in this literature apply Bayesian approach to the estimation of DSGE models. Identification problems can be alleviated with the use of priors for the structural parameters. Intuitively, an aptly selected prior can make a parameter identifiable by placing low probability on values which would render it unidentifiable.

In this section I show that identification of DSGE models can be studied directly, as a problem of uniqueness of solution to a system of equations, instead of trying to do that through the Information matrix. I start by deriving the cross-equation and covariance restrictions that relate the parameters of interest to those of the reduced-form model. Those restrictions are the basis of the identification results, as well as the estimation approach I pursue in the second part of the paper.

[7]Canova and Sala (2006), who study identification in DSGE models following this approach, compute the Hessian as the outer product of the gradient produced by the minimization routine used in estimation

2.3.1 Cross-equation and covariance restrictions

Consider again the structural and the reduced-form models, which I rewrite for convenience below.

$$\Gamma_0 Z_t = \Gamma_1 E_t Z_{t+1} + \Gamma_2 Z_{t-1} + \Gamma_3 U_t \tag{2.10}$$

$$Z_t = A Z_{t-1} + B U_t \tag{2.11}$$

Equation (2.11) implies that the expectation of Z_{t+1} is $E_t Z_{t+1} = A Z_t$. Using this in (2.10) yields:

$$\Gamma_0 Z_t = \Gamma_1 A Z_t + \Gamma_2 Z_{t-1} + \Gamma_3 U_t$$

or

$$(\Gamma_0 - \Gamma_1 A) Z_t = \Gamma_2 Z_{t-1} + \Gamma_3 U_t \tag{2.12}$$

Comparing the last equation with the one in (2.11), we find the following two sets of restrictions

$$(\Gamma_0 - \Gamma_1 A) A - \Gamma_2 = 0 \tag{2.13}$$

$$(\Gamma_0 - \Gamma_1 A) \Omega (\Gamma_0 - \Gamma_1 A)' - \Sigma = 0 \tag{2.14}$$

Unless additional assumptions about the probability distribution of U are made, the two equations (2.13) and (2.14), contain all we know about the relationship between the parameters of interest θ, and the reduced-form parameters τ. Moreover, since the reduced-form is all we know about the distribution of the data, the identification and the estimation of θ has to be based on these restriction. In the next two sections I explore further the implications of this observation, and present operational conditions for checking identification in DSGE models.

2.3.2 Identification of the structural parameters

In this section I derive the conditions for identification of the structural parameters γ. There two reasons why I am interested in this. First, the identifiability conditions for γ I derive here, will be useful in the next section, where the identification of the deep parameters is studied. And second, the identification and estimation of the structural parameters may be of independent interest for researchers. If, for instance, the goal is to study the economy's response to different shocks, the structural model is all we need. Furthermore, note that with $\Gamma_1 = \mathbf{0}$ the model in (2.10) reduces to the canonical structural vector autoregressive model (SVAR), which is widely used in the empirical macroeconomic literature. Thus, the results regarding the identification of γ are directly applicable to identification in SVAR models.

To find conditions for identification of γ, I use results R1 - R3 from section 2. Note that even though the statements there are made for the deep parameters θ , they are also valid for the structural parameters γ if $f(\theta, \pi)$ in (2.9) is replaced with $f^*(s, \pi)$.

It is straightforward to see from (2.14) that, if γ_1 is identified, γ_2 will be as well. Our first result provides a sufficient condition for global identification of γ, which follows from the identifiability of γ_1 from the cross-equation restrictions (2.13) only. To make this result self-evident, observe that (2.13) can be re-written as

$$\Xi_2 \gamma_1 = \Xi_1 \tag{2.15}$$

where the matrices Ξ_1 and Ξ_2 contain known constants and reduced-form parameters τ, which are identified by assumption 2.2.3. The derivation of (2.15) is shown in the Appendix. The following result follows immediately from (2.15), R1 and R3.

Lemma 2.3.1 (**Identification of γ by cross-equation restrictions**). *The vector of structural parameters γ is globally identifiable, by the cross-equation restrictions*

only, if the rank of Ξ_2 *is equal* l*, the dimension of* γ_1*.*

Alternatively, we can see this result as following directly from the well-known rank condition for uniqueness of a solution of a system of linear equations, which, in our notation, is that Ξ_2 has a full column rank. Lemma 2.3.1is analogous to the rank condition for identification of systems of linear simultaneous equations.[8]

As I discussed earlier, the covariance restrictions could also be useful for the identification of γ_1. For this to be the case we need to have some a priori restrictions on the structural covariance matrix Σ. If, for instance Σ is known to be diagonal, (2.14) could be useful for identifying γ_1, and, in fact, could help identify structural parameters which are otherwise unidentifiable. The difficulty with using the covariance restrictions is that, in general the system of equations (2.14) is nonlinear in γ_1, and therefore there are no general conditions for global uniqueness of its solution. Nevertheless, it is possible to find necessary and sufficient conditions for local uniqueness of the solution. To state a more general condition for identifiability of γ_1, we need some additional notation. Let f_1 and f_2 be the vectorized versions of (2.13) and (2.14) respectively. [9]

Also, reorder the elements of f_2 so that the ones corresponding to the a priori constrained components of Σ, denoted with f_{2c}, come before those corresponding to the unconstrained components, denoted with f_{2u}. Then, as I state in the following lemma, the identification of γ_1, when all available restrictions are used, depends on the rank of the Jacobian of the following matrix

$$\begin{bmatrix} f_1 \\ f_{2c} \end{bmatrix} \tag{2.16}$$

[8]The rank condition for identification of linear simultaneous equations models can be found in most econometrics textbooks. See for instance Theorem 1 on page 29 of Sargen (1988).

[9]$f_1 = \mathbf{vec}(\Gamma_0 A - \Gamma_1 A^2 - \Gamma_3)$; $f_2 = \mathbf{vec}((A - B\Phi)\Omega(A - B\Phi)' - \Sigma)$

Lemma 2.3.2 (**Identification of γ by cross-equation and covariance restrictions**). *The vector of structural parameters γ is locally identifiable, when both the cross-equation and the covariance restrictions are used, if the Jacobian with respect to γ_1 of the matrix in* (2.16) *has full column rank. This condition is also necessary for local identification, if assumption 2.2.2 is made, that is, if the density of the structural shocks is assumed to be Gaussian.*

The easiest way to understand this result is by appealing to the implicit function theorem. We can write the mean and covariance restrictions in a vectorized form as

$$f(\gamma, \tau) = 0 \tag{2.17}$$

The implicit function theorem says that a solution $(\gamma_0,\ \tau_0)$ of (2.17) is locally unique if the Jacobian $f_\gamma(\gamma, \pi)$ has full column rank when evaluated at $(\gamma_0,\ \pi_0)$.

To see that lemma 2.3.2 implies lemma 2.3.1, note that from the definition of f_1 and equation (2.15), we have $f_1 = \Xi_2\gamma_1 - \Xi_1$. Hence the Jacobian of f_1 with respect to γ_1 is equal to Ξ_2. If Ξ_2 has full column rank, the matrix in (2.16) will also have full column rank. Also, since Ξ_2 does not depend on the structural parameters, identification in that case is global.

To illustrate these points, consider the following simple rational expectations model.

$$y_t = \alpha E_t y_{t+1} + (1-\alpha)x_t + u_t \tag{2.18}$$

$$x_t = \beta x_{t-1} + \epsilon_t \tag{2.19}$$

The endogenous variable y is determined by its expected value in the next period, and by the current values of the observed exogenous variable x, and by an $i.i.d$ shock u. The exogenous variable x is governed by a stationary $AR(1)$ process. Both α and β are assumed to be positive and smaller than 1.

To put the model in the standard form of 2, define $Z = [y, x]'$ and $A,\ B$ and C as follows:

$$\Gamma_0 = \begin{bmatrix} 1 & -(1-\alpha) \\ 0 & 1 \end{bmatrix}, \Gamma_! = \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix}, \Gamma_2 = \begin{bmatrix} 0 & 0 \\ 0 & \beta \end{bmatrix}$$

Solving the model, we find that the reduced-form matrix A is given by

$$A \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 0 & \frac{(1-\alpha)\beta}{1-\alpha\beta} \\ 0 & \beta \end{bmatrix} \tag{2.20}$$

It is straightforward to express the cross-equation restrictions (2.13) in the form of (2.15) [10]:

$$\underbrace{\begin{bmatrix} a(1-a_{12}) & 0 \\ 0 & -1 \end{bmatrix}}_{\Xi_2} \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\gamma_1} = -\underbrace{\begin{bmatrix} \phi_{12} - \phi_{22} \\ \phi_{22} \end{bmatrix}}_{\Xi_1} \tag{2.21}$$

Note that for this example there is no difference between structural and reduced-form parameters, that is, $\gamma = \theta$. The rank condition for identification requires that the rank Ξ_2 is equal to two at the true value of A. Therefore the mean restrictions will be sufficient for identification of α and β only if $a_{22}(1 - a_{12}) \neq 0$. Using $a_{12} = \frac{(1-\alpha)\beta}{1-\alpha\beta}$ and $a = \beta$, we find that the last condition is equivalent to $\beta(1-\beta) \neq 0$, or $\beta \neq 0$, and $\beta \neq 1$. These two conditions are also necessary if there are no a priori restrictions on Σ.

[10] The equality in (2.20) implies the following relationships between the reduced-form parameters $a_{12},\ a_{22}$, and the structural parameters $\alpha,\ \beta$:

$$\begin{aligned} (\alpha a_{12} + 1 - \alpha)\beta &= a_{12} \\ \beta &= a_{22} \end{aligned}$$

which can be re-written as (2.21)

To see the effect of having restrictions on Σ, suppose that ϵ and u are uncorrelated, which implies that Σ is diagonal. Then, from the covariance restrictions (2.14) we obtain an additional equation that could be useful for the identification of α and β. The equation is: [11]

$$\Omega_{12}(1-\alpha\phi_{11}) - \Omega_{22}(1-\alpha(1-\phi_{12}) = 0 \tag{2.22}$$

Thus, the Jacobian of the matrix in (2.16) is:

$$\begin{bmatrix} \Xi_2 \\ \frac{\partial f_{2c}}{\partial \gamma_1} \end{bmatrix} = \begin{bmatrix} a_{22}(1-a_{12}) & 0 \\ 0 & -1 \\ \Omega_{22}(1-a_{12}) & 0 \end{bmatrix} \tag{2.23}$$

Therefore the model is identified if and only if the rank of this matrix is equal to 2. We have the following three cases:

1. $\beta \neq 1$ and $\beta \neq 0$. Then α is identified by both the cross-equation and the covariance restrictions;
2. $\beta = 0$. Then $a_{12} = a_{22} = 0$, and since $\Omega_{22} = \sigma^2_\epsilon \neq 0$[12]
   , α is identified by the covariance restrictions only;
3. $\beta = 1$. Then $a_{12} = a_{22} = 1$, and therefore α is not identified

An interesting feature of this example is that identification, even if it depends on the covariance restrictions, is always global. This is not generally true as I noted above. It is true here because of the recursive structure of the model, which makes $\Gamma_0 - \Gamma_1 A$ a triangular matrix, and, therefore the equation implied by the restriction on Σ is linear in the structural parameters. [13]

[11] This is the second equation in $f_2 \equiv \mathbf{vec}((\Gamma_0 - \Gamma_1 A)\Omega(\Gamma_1 - \Gamma_1 A)') - \mathbf{vec}(\Sigma) = 0$

[12] Solving for Ω we find

$$\Omega \equiv \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} \sigma^2_u + \frac{(1-\alpha)^2}{(1-\alpha\beta)^2}\sigma^2_\epsilon & \frac{(1-\alpha)}{1-\alpha\beta}\sigma^2_\epsilon \\ \frac{(1-\alpha)}{1-\alpha\beta}\sigma^2_\epsilon & \sigma^2_\epsilon \end{bmatrix}$$

[13] One way to see this is as follows: let $\bar{\Omega}$ and $\bar{\Sigma}$ be lower triangular matrices, such that $\bar{\Omega}\bar{\Omega}' = \Omega$ and $\bar{\Sigma}\bar{\Sigma}' = \Sigma$. We know that $\bar{\Omega}$ and $\bar{\Sigma}$ are unique, because of the uniqueness of the Cholesky de-

2.3.3 Identification of the deep parameters

As I discussed in section 2, the coefficient matrices $\Gamma_0,\ \Gamma_1$ and Γ_2 of the model in (2.4) are typically functions of behavioral or deep parameters. In the current empirical DSGE literature researchers are usually interested in estimating them, and not the structural parameters the identification of which was studied in the previous section [14]. In this section I consider the identification of the deep parameters $\theta = [\theta_1',\ \theta_2']'$. As in the previous section, I am primarily concerned with the identification of θ_1 in view of the fact that, if it is identified, θ_2 will be identified as well.

We know from section 2 that the identifiability of θ depends on whether the equations in (2.13) and (2.14) have unique solutions for θ. A sufficient condition for that is given in the following lemma.

Lemma 2.3.3 (**Identification of θ by cross-equation restrictions**). *The vector of deep parameters θ is locally identifiable by the cross-equation restrictions only, if the rank of the matrix*

$$\Gamma_2 \frac{\partial g}{\partial \theta_1}(\theta_1) \tag{2.24}$$

is equal to k, the dimension of θ_1. This condition is sufficient for global identifiability, only if the rank is equal to k for all admissible values of θ_1.

As defined in section 2 (see (2.6)), g is the function mapping θ_1 into γ_1; To understand the lemma, suppose first that γ_1 is identifiable by the mean restrictions only, which, by virtue of Lemma 2.3.1 implies that Ξ_2 is of full rank. Then θ_1 will be identified if the equation

$$\gamma_1 = g(\theta_1)$$

composition of a positive definite matrices. In general it is not true that $A\bar{\Omega}\bar{\Omega}'A' = \bar{\Sigma}\bar{\Sigma}'$ implies $A\bar{\Omega} = \bar{\Sigma}$. If, however, A is full rank and lower-triangular, the last equality is true, since then $A\bar{\Omega}$ is lower triangular, and thus unique by the uniqueness of the Cholesky decomposition of a positive definite matrices. Therefore, in this case the covariance restrictions can be expressed as a system of linear equations. For more on this, see section 2.

[14] There are exceptions though, for example Lubik and Schorfheide (2004) estimate structural parameters only

has a unique solution. For a general function g we can only determine whether a solution θ_1^* is locally unique, that is, for any other θ_1^{**} in some neighborhood of θ_1^*, $g_1(\theta_1^{**}) \neq q$. The condition is that the Jacobian $\frac{\partial g}{\partial \theta_1}(\theta_1)$, is of full column rank when evaluated at θ_1^* [15].

It is important to note, however, that, unless γ_1 and θ_1 are of the same dimension, we do not have to know the whole vector γ_1 to identify θ_1. In other words, Ξ_2 does not have to be of full column rank, as long as the matrix in (2.24) is. Using the fact that the rank of a product of two matrices is no greater than the smaller of the two ranks, we obtain the following corollary

Corollary 2.3.1. *A necessary condition for θ to be locally identifiable by the cross-equation restrictions only, is that both Ξ_2 and $\frac{\partial g}{\partial \theta_1}(\theta_1)$ have ranks greater or equal to k, the dimension of θ_1.*

Lemma 2.3.3 tells us when the mean restrictions (2.13) are sufficient for the identification of θ. As before, the covariance restrictions (2.14) are also potentially useful for identifying θ_1, if there are a priori restrictions on the structural covariance Σ. As in the previous section, the identifiability of θ, when both the mean and the covariance restrictions are used, depends on the rank of the Jacobian of the matrix in (2.16)

Lemma 2.3.4 (**Identification of θ by cross-equation and covariance restrictions**). *The vector of deep parameters θ is locally identifiable when both the cross-equation and the covariance restrictions are used, if the Jacobian with respect to θ_1 of the matrix in* (2.16) *is equal to k, the dimension of θ_1. This condition is also necessary for local identifiability of θ_1, if assumption 2.2.2 is made, that is, if the*

[15] This follows from the mean value theorem, according to which there exists a $\bar{\theta}_1$, such that

$$g_1(\theta_1^{**}) = g_1(\theta_1^*) + \frac{\partial g}{\partial \theta_1}(\bar{\theta}_1)(\theta_1^{**} - \theta_1^*) \tag{2.25}$$

If $\frac{\partial g}{\partial \theta_1}(\bar{\theta}_1)$ has a full column rank, then there exists a neighborhood Δ of θ_1^*, such that $\theta_1^* \in \Delta$, and $\frac{\partial g}{\partial \theta_1}(\bar{\theta}_1)$ has a full column rank for any θ_1 in that neighborhood, and therefore the only solution of $g(\theta_1^*) = g(\theta_1^{**}) = 0$ for any θ_1^{**} in Δ is $\theta_1^{**} = \theta_1^*$

density of the structural shocks is assumed to be Gaussian.

The difference between Lemma 2.3.4 and Lemma 2.3.2 is that here the derivatives are taken with respect to the deep parameters θ_1, instead of γ_1. As with Lemma 2.3.3, the following corollary follows immediately

Corollary 2.3.2. *A necessary condition for θ to be locally identifiable by the cross-equation and covariance restrictions, is that both the Jacobian with respect to γ_1 of the matrix in* (2.16)*, and $\frac{\partial g}{\partial \theta_1}(\theta_1)$ have ranks greater or equal to k, the dimension of θ_1.*

Clearly, there are order conditions associated with the rank conditions given above and in Corollary 2.3.1. For a matrix to have a rank equal to k, its row and column dimensions must be greater or equal to k. This implies that the row dimension of the matrix in (2.16) must not be less than k, and also, that there must be at least k structural parameters, i.e. we must have $l \geq k$.

Since γ depends on θ_1 only through γ_1, it should be clear that the covariance restrictions are useful for the identification of θ only as far as they help identify γ_1. If the cross-equation restrictions are sufficient for the identification of γ_1, that is, if Ξ_2 is of full column rank, all one needs to determine is whether there is a unique solution of the equation

$$\gamma_1 = g(\theta_1)$$

in that case the rank condition on the Jacobian of g is both necessary and sufficient for identification of θ_1.

The above results illustrate the fact that the identification of deep parameters in DSGE models involves two steps. First, the reduced-form parameters τ are used to identify as many of the structural parameters γ_1 as possible; second, the identified structural parameters are in turn used to identify the deep parameters θ. [16] The

[16] There is also a third transition - from data to reduced-form, which is equally essential.

identification of θ will fail, if, at any of these steps, the mapping from what is known to what we want to determine is not locally or globally unique. It is important to realize that the conditions for identification I have presented, are not specific to the method used to estimate the parameters of interest. In particular, we do not have to use a two step estimation procedure for identification to fail, if the necessary conditions in Corollary 2.3.2 are violated. Similarly, if the sufficient conditions in Lemma 2.3.3 or Lemma 2.3.4 are met, θ is identifiable, and can be estimated, as long as all cross-equation restrictions or cross-equation and covariance restrictions are used.

To illustrate these points, I consider the following linearized DSGE models

$$y_t = \gamma_{11} E_t y_{t+1} + \gamma_{12} y_{t-1} - \gamma_{13} (r_t - E_t \pi_{t+1}) + v_t \tag{2.26}$$

$$\pi_t = \gamma_{21} E_t \pi_{t+1} + \gamma_{22} y_t + \gamma_{23} \pi_{t-1} + e_t \tag{2.27}$$

$$r_t = \gamma_{31} \pi_{t-1} + \gamma_{32} y_{t-1} + \gamma_{33} r_{t-1} + \xi_t \tag{2.28}$$

$$v_t = \gamma_{41} v_{t-1} + \nu_t \tag{2.29}$$

$$e_t = \gamma_{51} e_{t-1} + \epsilon_t \tag{2.30}$$

where the vector of structural parameters γ_1 is related to the deep parameters θ_1 as follows

$$\begin{aligned}
&\gamma_{11} = \frac{\chi}{1+\chi}, \quad \gamma_{12} = \frac{1}{1+\chi}, \quad \gamma_{13} = -\frac{1}{\psi}; \\
&\gamma_{21} = \frac{\beta}{1+\varpi\beta}, \quad \gamma_{22} = \frac{(\psi+\nu)(1-\zeta\beta)(1-\zeta)}{(1+\varpi\beta)\zeta}, \quad \gamma_{23} = \frac{\varpi}{1+\varpi\beta}; \\
&\gamma_{31} = (1-\lambda_r)\lambda_\pi, \quad \gamma_{32} = (1-\lambda_r)\lambda_y, \quad \gamma_{33} = \lambda_r; \\
&\gamma_{41} = \rho_1, \quad \gamma_{51} = \rho_2;
\end{aligned} \tag{2.31}$$

There are 14 deep parameters in this model, 11 of them are $\theta_1 = [\beta, \chi, \varpi, \psi, \nu, \zeta, \lambda_r, \lambda_{pi}, \lambda_y, \rho_1, \rho_2]'$, and the other 3 are $\sigma = [\sigma_\nu, \sigma_\epsilon, \sigma_\xi]'$.

I start by checking the necessary condition for the rank of $\frac{\partial g}{\partial \theta_1}(\theta_1)$. According to

Corollary 2.3.2, if the rank is less than 11, then θ is not identified. Since there are 11 structural parameters γ_1, the dimension of $\frac{\partial g}{\partial \theta_1}(\theta_1)$ is 11×11, so it passes the order condition. An expression for the Jacobian is given in the Appendix. That the necessary rank condition fails follows from the fact that two of the structural parameters - γ_{11} and γ_{12}, depend on single deep parameter - χ. Because of that, the first two rows of $\frac{\partial g}{\partial \theta_1}(\theta_1)$ are not linearly independent, and, therefore, the rank is less than 11. Another way to see that is by noting that two of the deep parameters - ζ and ν, appear in only one of the structural parameters - γ_{22}. Therefore, unless additional information is available, ζ and ν are not identifiable. Proceeding in the same manner, I find that all other parameters can be solved for if γ_1 is known. All except β and ϖ are globally uniquely determined, and β and ϖ are only locally uniquely determined.[17] This implies that $\chi, \psi, \lambda_r, \lambda_{pi}, \lambda_y, \rho_1, \rho_2, \sigma_\nu, \sigma_\epsilon, \sigma_\xi$ will be globally identifiable, β and ϖ - locally identifiable, and ζ and ν - unidentifiable, if γ_1 can be identified. From Lemma 2.3.3, a sufficient condition for identification of γ_1 is that matrix Ξ_2 from (2.15) has a full column rank. As I show in the Appendix, Ξ_2 is of full column rank for the parameter values used in Canova and Sala (2006). Therefore, for those parameter values, γ_1 is globally identified by the mean restrictions only.

2.3.4 Identification in single equations

In this section I discuss the conditions for identification of the parameters in a single equation in DSGE models. As before, I will assume that the true reduced form of the complete model is known, and is identifiable. However, I will not assume that the structural or deep parameters in the other equations of the model or identifiable, or that the form of the cross-equation or covariance restrictions implied by these equations are known.

[17] ϖ can be determined by solving $\gamma_{21}\varpi^2 - \varpi + \gamma_{23} = 0$, and then β can be found from $\beta = \frac{\gamma_{21}}{\gamma_{23}}\varpi$. If either γ_{21} or γ_{23} is equal to zero, β and ϖ can be determined globally uniquely.

Suppose that we are interested in equation i from the model in (2.4):

$$\Gamma_{0i} Z_t = \Gamma_{1i} E_t Z_{t+1} + \Gamma_{2i} Z_{t-1} + \Gamma_{3i} u_{it} \tag{2.32}$$

Here $\Gamma_{0i},\ \Gamma_{1i},\ \Gamma_{2i}$ and Γ_{3i} are the i-th rows of $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3 respectively, and u_{it} is the i-th element of the vector U_t.

Following the steps from section 2, we obtain the following set of cross-equation restrictions:

$$\Gamma_{0i} A - \Gamma_{1i} A^2 - \Gamma_{2i} = 0 \tag{2.33}$$

Let θ_1^i be a $k_i \times 1$ vector of the components of θ_1 included in the i-th equation, and, similarly, γ_1^i be a $l_i \times 1$ vector collecting the structural parameters included in the i-th equation. Also, denote the mapping from θ_1^i to γ_1^i with the function g_1^i:

$$\gamma_1^i = g_1^i(\theta_1^i) \tag{2.34}$$

Ultimately, we are interested in the identification of θ_1^i. As before, this depends on, first, which of the elements of γ_1^i are identifiable, and second, can we determine uniquely θ_1^i from the identifiable components of γ_1^i. Before turning to the general result, I should point out that equation (2.32) is unidentifiable if all non-zero components of the vectors Γ_{0i}, Γ_{1i}, and Γ_{2i} are unknown parameters. This can be seen from equation (2.33), for which, given A, if $(\Gamma_{0i}, \Gamma_{1i}, \Gamma_{2i})$ is a solution, then $(x\Gamma_{0i}, x\Gamma_{1i}, x\Gamma_{2i})$ is also a solution for any $x \in \mathbb{R}$. If some of the non-zero components of Γ_{0i}, Γ_{1i}, and Γ_{2i} are known constants, we can write (2.33) in the following form:

$$\Xi_2^i \gamma_1^i = \Xi_1^i \tag{2.35}$$

The following lemma is the counterpart of Lemmas 2.3.1 and Lemma 2.3.3 for the single equation case.

Lemma 2.3.5 (Identification of γ_1^i and θ_1^i).

(i) The vector γ_1^i of structural parameters included in the i-th equation of the DSGE model is identifiable if and only if the rank of Ξ_2^i is equal to l_i.
(ii) The vector θ_1^i of deep parameters included in the i-th equation of the DSGE model are identifiable if and only if the rank of the product $\Xi_2^i \frac{\partial g_i}{\partial \theta_1^i}$ is equal to k_i.

As before, we have a necessary order condition - that the dimension of θ_1^i must not be larger that the dimension of γ_1^i, i.e. $k_i \leq l_i$

To illustrate, consider again the model in (2.26)-(2.30), and suppose that we want to know whether the Phillips curve (equation (2.27)) is identified. The equation is

$$\pi_t = \gamma_{21} E_t \pi_{t+1} + \gamma_{22} y_t + \gamma_{23} \pi_{t-1} + e_t \tag{2.36}$$

where mapping between structural parameters ($\gamma_{21},\ \gamma_{22},\ \gamma_{23}$), and deep parameters ($\beta,\ \varpi,\ \psi,\ \nu,\ \zeta$) is given by

$$\gamma_{21} = \frac{\beta}{1+\varpi\beta}, \quad \gamma_{22} = \frac{(\psi+\nu)(1-\zeta\beta)(1-\zeta)}{(1+\varpi\beta)\zeta}, \quad \gamma_{23} = \frac{\varpi}{1+\varpi\beta} \tag{2.37}$$

Let e_j be the j-th column of a 5×5 identity matrix. Then cross-equation restrictions are given by

$$e_2' A - \gamma_{21} e_2' A^2 - \gamma_{22} e_2' A - \gamma_{23} e_2' = 0 \tag{2.38}$$

which can be re-written as

$$\underbrace{\begin{bmatrix} A'^2 e_2 & A' e_1 & e_2 \end{bmatrix}}_{\Xi_2^i} \begin{bmatrix} \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} = \Xi' e_2 \tag{2.39}$$

The structural parameters $\gamma_{21},\ \gamma_{22}$ and γ_{23} are identifiable is the rank of $\begin{bmatrix} \Xi'^2 e_2 & \Xi' e_1 & e_2 \end{bmatrix}$ is equal to 3 [18]. The Jacobian of g_1^i is a 3×4 matrix and therefore does not meet the

[18] It is easy to show that this rank condition is equivalent to the existence of valid instruments for

order condition. Thus the deep parameters in this example are not globally identifiable. Examining the relationships between the γ's and the deep parameters given in (2.37), we find that there are two solutions for ϖ and β, and ψ, ν and ζ are completely unidentified.

This example demonstrates an approach to analyzing identification that is frequently very convenient in practice. Instead of computing the Jacobian and verifying the rank condition from Lemma 2.3.4, which would only tell us whether the parameter vector is locally identified as a whole, we could do the following:

1. examine which of the structural parameters are globally identified using the conditions in Lemma 2.3.1 or Lemma 2.3.5 (i);
2. go through each deep parameter and find whether it can be identified from knowing the identifiable structural parameters only;
3. find what other deep parameters can be identified from the known structural parameters and the deep parameters identified in the previous step.

Following this procedure would give us more information about the nature of identification failures if there are any, and may point at ways of dealing with them in estimation.

2.4 Estimation

To establish parameter identifiability we ask whether knowing the *true* characteristics of the data density allows for unique determination of the parameters values. For the models considered in this paper this boils down to the uniqueness of the mapping from reduced-form parameters γ to the deep parameters θ. In section 2 I presented conditions for this to be true. If the true value of γ were known, we would not have do any estimation, and would simply solve for θ, assuming of course that it is identified. In practice the true reduced-form coefficients γ are unknown and have to be estimated. Because of the estimation uncertainty, we can no longer obtain θ

estimation of equation (2.36) with GMM

by simply solving the system of equation. In this section I introduce a minimum distance approach for estimating the deep parameters in linear rational expectations models. Minimum distance estimation of forward-looking relationships has been used previously by Sbordone (2002, 2005a,b) and Li (2004). Here I extend this estimation approach to fully-fledged DSGE models. The estimation proceeds in two steps. First, the reduced form of the model is estimated. Then, the parameters of interest are obtained by making the restrictions implied by the model hold as closely as possible.

I start by reviewing the first step estimation, and after that I turn to the estimation of the deep parameters.

2.4.1 Estimation of A and Ω

The reduced form of the model is given in equation (2.7), reproduced for convenience below

$$Z_t = AZ_{t-1} + V_t, \quad \text{for } \mathrm{t} = 2 \ldots \mathrm{T} \tag{2.40}$$

The method of estimating A and $\Omega = \mathbb{E}V_t V_t'$ depends on whether or not Z is fully observable. A brief discussion of the two cases follows.

Z is fully observable

When all components of Z are observable (2.40) is a standard VAR model and can be estimated by OLS. We can write (2.40) compactly as follows:

$$Z = AX + V \tag{2.41}$$

where $Z = \begin{bmatrix} Z_2 & \ldots & Z_T \end{bmatrix}$, $X = \begin{bmatrix} Z_1 & \ldots & Z_{T-1} \end{bmatrix}$, and $V = \begin{bmatrix} V_2 & \ldots & V_T \end{bmatrix}$.

The OLS estimators [19] of A and Ω are

$$\hat{A} = ZX'(XX')^{-1} \tag{2.42}$$

$$\hat{\Omega} = \frac{1}{T}(Z - \hat{A}X)(Z - \hat{A}X)' \tag{2.43}$$

The asymptotic distribution of the two estimators is given by

$$\sqrt{T}\left(\begin{bmatrix} \hat{\mathbf{a}} - \mathrm{a}) \\ \hat{\omega} - \omega) \end{bmatrix} \right) \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} M_{XX}^{-1} \otimes \Omega & 0 \\ 0 & 2\big(\Omega \otimes \Omega\big) \end{bmatrix} \right) \tag{2.44}$$

where $\mathbf{a} = \mathbf{vec}(\Phi)$, $\omega = \mathbf{vec}(\Omega)$, and M_{XX} is defined to be the probability limit of $\frac{1}{T}XX'$.

Z is not fully observable

Most of the current generation of DSGE models imply a reduced form where not all components of Z are observable. This would be the case, for instance, if the structural shocks are persistent. Then the structural shocks would have to be included in Z, making it a partially observable state vector. A general way for dealing with this complication is to complement (2.40) with an observation equation

$$Y_t = GZ_t, \quad \text{for } \mathrm{t} = 2 \ldots \mathrm{T} \tag{2.45}$$

and use Kalman filter and maximum likelihood estimation to estimate of Φ and Ω. G in (2.45) is a matrix selecting the observable components of Z. Under standard assumptions, the ML estimates are asymptotically normally distributed, with asymptotic covariance equal to the inverse of the information matrix.

In following I will assume that consistent estimates of the reduced form param-

[19] These are also the ML estimators

eters, denoted with $\hat{\tau} = (\hat{\mathbf{a}},\ \hat{\omega})$, and of the asymptotic covariance matrix, denoted with $\hat{V}(\hat{\tau})$, are available.

Assumption 2.4.1. *An estimator $\hat{\tau}$ of τ is available, such that $\sqrt{T}\big(\hat{\tau} - \tau\big) \xrightarrow{d} \mathbb{N}\big(0, V(\hat{\tau})\big)$.*

2.4.2 Minimum Distance Estimation of the deep parameters

The cross-equation and covariance restrictions in (2.13) and (2.13) must hold for the true values of τ and θ. We can express this compactly, as in (2.9), by writing

$$f(\theta, \tau) = 0 \tag{2.46}$$

The minimum distance estimation of θ proceeds in two steps. First, a consistent estimate $\hat{\tau}$ of τ is obtained. Second, θ is estimated by minimizing the distance of $f(\theta, \hat{\tau})$ from zero. More formally, in the second step we solve the following optimization problem:

$$\min_{\theta} \underbrace{f(\hat{\tau}, \theta)' W_T f(\hat{\tau}, \theta)}_{S_1(\theta, \hat{\tau}, W_T)} \tag{2.47}$$

where W_T is a symmetric positive semidefinite weighting matrix, assumed to converge in probability to a symmetric positive definite matrix W. Let $\hat{\theta}_{MD}$ be the solution of (2.47). The asymptotic properties of $\hat{\theta}_{MD}$ are summarized in the following lemma

Lemma 2.4.1 (**Asymptotic properties of $\hat{\theta}_{MD}$**). *Suppose that Assumptions 2.2.1 and 2.4.1 are satisfied, that Θ is compact, and that θ_0 is identified. Then*

1. *$\hat{\theta}_{MD}$ is consistent,*
$$\hat{\theta}_{MD} \xrightarrow{p} \theta_0$$
2. *$\hat{\theta}_{MD}$ is asymptotically normally distributed,*
$$\sqrt{T}(\hat{\theta}_{MD} - \theta_0) \xrightarrow{d} \mathbb{N}\big(0, V(\theta_0, \tau_0)\big)$$

where

$$V(\theta_0, \tau_0) = V_1^{-1} V_2 V_1^{-1} \tag{2.48}$$

and (2.49)

$$V_1 = f_\theta(\theta_0, \tau_0)' W_T f_\theta(\theta_0, \tau_0) \tag{2.50}$$

$$V_2 = f_\theta(\theta_0, \tau_0)' W_T f_\tau(\theta_0, \tau_0) V(\hat{\tau}) f_\tau(\theta_0, \tau_0)' W_T f_\theta(\theta_0, \tau_0) \tag{2.51}$$

3. *If* $W = W^*(\theta)$ *where* $W^*(\theta) = \left[f_\tau(\theta_0, \tau_0) V(\hat{\tau}) f_\tau(\theta_0, \tau_0)' \right]^{-1}$, $\hat{\theta}_{MD}$ *is asymptotically efficient, and its asymptotic covariance matrix is given by*

$$V(\theta_0, \tau_0) = \left[f_\theta(\theta_0, \tau_0)' \left(f_\tau(\theta_0, \tau_0) V(\hat{\tau}) f_\tau(\theta_0, \tau_0)' \right)^{-1} f_\theta(\theta_0, \tau_0) \right]^{-1} \tag{2.52}$$

In order to use the efficient estimator in practice, we need a consistent estimate $\hat{W}_T^*$ of the weighting matrix in (2.52). One approach is to use a two step estimator, by first obtaining a preliminary inefficient but consistent estimate $\tilde{\theta}$ of θ, and then solving (2.47), with $W_T = W^*(\tilde{\theta})$. Alternatively, we can write W_T in terms of the unknown θ, and solve (2.47) with respect to θ. These two estimating approaches are analogues of the two-step and the continuous-updating (CU) GMM estimators respectively. There are no conclusive evidence that show superiority of one estimator over the other. The continuous-updating estimator is less biased, but has a flatter distribution than the two-step estimator. Another advantage of the CU estimator is that, unlike the two-step GMM, it is invariant to reparametrization (Hansen, Heaton, and Yaron (1996)). An advantage of the two-step estimator is that it is typically easier to use. This is true for the models we are concerned with in situations when for a given weighting matrix, the minimum distance estimator is available in analytical form; I discuss when this is possible in subsection 2 below.

An alternative form of the minimum distance estimator is obtained if, instead implicitly as in (2.46), the mapping from θ to τ is given explicitly, e.g.

$$\tau = h(\theta) \tag{2.53}$$

Then the efficient MD estimator solves

$$\min_{\theta} \underbrace{(\hat{\tau} - h(\theta))' W^{**} (\hat{\tau} - h(\theta))}_{S_2(\theta, \hat{\tau}), W^{**}} \tag{2.54}$$

where the optimal weighting matrix is $W^{**} = V(\hat{\tau})^{-1}$. Note that in this case W^{**} does not depend on θ. Moreover, since $V(\tau)$ is block diagonal, we can define h_1 and h_2 as [20]

$$\phi = h_1(\theta_1) \tag{2.55}$$

$$\omega = h_2(\theta) \tag{2.56}$$

and re-write the objective function in (2.54) as

$$S_2(\theta, \hat{\tau}, W^{**}) = (\hat{\phi} - h_1(\theta_1))' V(\hat{\phi})^{-1} (\hat{\phi} - h_1(\theta_1)) + (\hat{\omega} - h_2(\theta))' V(\hat{\omega})^{-1} (\hat{\omega} - h_2(\theta)) \tag{2.57}$$

As in section 2, let H be the Jacobian of h, that is

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial h_1}{\partial \theta_1} & 0 \\ \frac{\partial h_2}{\partial \theta_1} & \frac{\partial h_2}{\partial \theta_2} \end{bmatrix} \tag{2.58}$$

Then the asymptotic covariance of the estimator defined by (2.54) is

$$\left[H(\theta_0)' \big(V(\hat{\tau}) \big)^{-1} H(\theta_0) \right]^{-1} \tag{2.59}$$

The equivalence between (2.59) and (2.48) follows as a straightforward application of the Implicit function theorem. The estimator obtained by solving (2.47) is generally more convenient to use in practice because the function f is available in analytical form. In some cases this could greatly simplify the minimization problem, and, in

[20] Note that ϕ is not affected by θ_2, while, in general, both θ_1 and θ_2 affect ω.

fact, make numerical optimization unnecessary. Function h, on the other hand, can only be obtained numerically, and the minimization problem in (2.54) has to be solved numerically.

Unlike the first minimum distance estimator, for which the function $f(\theta, \tau)$ is available in analytical form, when using the second minimum distance estimator we have to compute $h(\theta)$ numerically in order to evaluate the objective function. This could be done by using some of the available algorithms for solving linear rational expectations models. In practice going through this additional complication may be justified by the fact that the weighting matrix is of much simpler form and does not depend on the unknown parameters, thus making the optimization problem easier than the one in (2.47).

The minimum distance estimators are asymptotically efficient only if all a priori known information about θ is used in the estimation. In general, this requires that function f in (2.47) contain all cross-equation and covariance restrictions from (2.13) and (2.14). In practice it is often much simpler to estimate the parameters using the cross-equation restrictions only, and thus it may be tempting to do so provided that they are sufficient for identification. Next I will examine the consequences in terms of efficiency of doing that.

Estimation ignoring the covariance restrictions

To examine the effect of ignoring the information contained in the covariance restrictions, it is more convenient to assume that the function h in (2.53) is available. Since the two minimum distance estimators, solving (2.47) and (2.54), are equivalent, this assumption is without loss of generality.

Suppose that cross-equation restrictions (2.13) are sufficient for identification of the deep parameters θ. Following the discussion in section 2, this this will be true if the matrix $H_{11} = \frac{\partial h_1}{\partial \theta_1}$ is of full column rank. Then θ_1 can be estimated consistently

using only the cross-equation restrictions; the asymptotic covariance matrix of the estimator will be (see Rothenberg (1966) for derivation):

$$M^{(1)} = \left[H_{11}' V(\phi) H_{11}\right]^{-1} \tag{2.60}$$

The efficient estimator on the other hand, has asymptotic covariance matrix given in (2.59) which can be written as

$$M_{11} = \left[M^{(1)} + M^{(2)}\right]^{-1} \tag{2.61}$$

where

$$M^{(2)} = H_{21}' V(\omega) H_{21} - H_{21}' V(\omega) H_{22} \left(H_{22}' V(\omega) H_{22}\right)^{-1} H_{22}' V(\omega) H_{21}$$

Intuitively, if the covariance restrictions contain independent information about θ_1, ignoring them would lead to a less precise estimate. In other words, we should expect that M_{11} is at least as small as $M^{(1)}$. This is indeed the case since the matrix $M^{(2)}$ can be shown to be positive semi-definite. On the other hand, if there is no useful information in the covariance restrictions, or if Ω is completely unrestricted, the matrix $M^{(2)}$ is singular, and the estimator based on the cross-equation restrictions only achieves asymptotic efficiency.

To recapitulate, in addition to providing potentially useful identifying information about the deep parameters, using the covariance restriction in estimation could improve substantially the precision of the estimates. For that we need to have some a priori knowledge about the shocks in the structural model. This is indeed the case in DSGE models, where the shocks are typically assumed to be uncorrelated.

Estimation of models linear in the parameters

In general, the minimum distance problem has to be solved numerically. However, a simple closed-form solution exists for models that a linear in the parameters we want to estimate. More precisely, this is true if

1. we estimate the structural coefficients γ_1 using the cross-equation restrictions only;
2. the model is recursive and thus the matrix $A-B\Phi$ can be written as a triangular matrix.

In both of these circumstances the system of cross-equation and covariance restrictions used for estimation is linear and therefore the minimization problem (2.47) has analytical solution. For example, if only the cross-equation restrictions are used in the estimation, $\hat{\gamma_1}_{MD}$ is obtained by minimizing (see (2.15)):

$$(\hat{\Xi}_2\gamma_1 - \hat{\Xi}_1)'W^*(\hat{\Xi}_2\gamma_1 - \hat{\Xi}_1) \tag{2.62}$$

the solution to which is:

$$\hat{\gamma_1}_{MD} = (\hat{\Xi}_2'W^*\hat{\Xi}_2)^{-1}(\hat{\Xi}_2'W^*\hat{\Xi}_1) \tag{2.63}$$

Once we have $\hat{\gamma_1}$, we can find $\hat{\gamma}$ from (2.14). In a similar fashion we can get the estimator of γ when the covariance restrictions can be written as a linear system, and are also used in the estimation of γ_1.

In the appendix I provide details on how Ξ_1 and Ξ_2 are obtained in the general case. Here I illustrate it in the simple example from section 2. The first thing to note is that the model is recursive. Switching the positions of the two equations, we obtain the following expressions for the structural coefficient matrices:

$$\Gamma_0 = \begin{bmatrix} 1 & 0 \\ -(1-\alpha) & 1 \end{bmatrix}, \Gamma_1 = \begin{bmatrix} 0 & 0 \\ 0 & \alpha \end{bmatrix}, \Gamma_2 = \begin{bmatrix} \beta & 0 \\ 0 & 0 \end{bmatrix}$$

It is straightforward to see that $(\Gamma_0 - \Gamma_1 A)$ is a lower triangular matrix given by:

$$(\Gamma_0 - \Gamma_1 A) = \begin{bmatrix} 1 & 0 \\ -1 + \alpha(1 - a_{21}) & 1 - \alpha a_{22} \end{bmatrix}.$$

Let $\bar{\Omega}$ be such that $\bar{\Omega}\bar{\Omega}' = \Omega$. Since Ω is a covariance matrix, it is positive definite, and therefore $\bar{\Omega}$ is a unique lower triangular matrix. For the same reason there is a unique matrix $K = (\Gamma_0 - \Gamma_1 A)\bar{\Omega}$ such that

$$KK' = \Sigma$$

This implies that

$$K = \bar{\Sigma}$$

where $\bar{\Sigma}$ is the square root of Σ. We end up with the following two sets of linear equations:

$$(\Gamma_0 - \Gamma_1 A)A - \Gamma_2 = 0 \tag{2.64}$$

$$(\Gamma_0 - \Gamma_1 A)\bar{\Omega} - \bar{\Sigma} = 0 \tag{2.65}$$

or, more explicitly:

$$\underbrace{\begin{bmatrix} 0 & -1 & 0 & 0 \\ a_{22}(1 - a_{12}) & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ \bar{\omega}_{11}(1 - a_{12}) & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}}_{\Xi_2} \begin{bmatrix} \alpha \\ \beta \\ \sigma_1 \\ \sigma_2 \end{bmatrix} = \underbrace{\begin{bmatrix} -a_{22} \\ a_{22} - a_{12} \\ -\bar{\omega}_{11} \\ \bar{\omega}_{11} - \bar{\omega}_{21} \\ -\bar{\omega}_{22} \end{bmatrix}}_{\Xi_1} \tag{2.66}$$

where I have used that Φ and $\bar{\Omega}$ are given by:

$$\Phi = \begin{bmatrix} a_{22} & 0 \\ a_{12} & 0 \end{bmatrix}, \quad \text{and} \quad \bar{\Omega} = \begin{bmatrix} \bar{\omega}_{11} & 0 \\ \bar{\omega}_{21} & \bar{\omega}_{22} \end{bmatrix} \tag{2.67}$$

Plugging the estimated vales of Ξ_2 and Ξ_1 in (2.63) yields the minimum distance estimate of $[\alpha,\ \beta,\ \sigma_1,\ \sigma_2]'$.

2.4.3 Inference

The objective function of the minimum distance problem can be used to test the validity of the cross equation and covariance restrictions. Asymptotically, the distribution of J defined below is χ^2 with $m^2 - n - k$ degrees of freedom.

$$J = TS_1(\hat{\theta}, \hat{\tau}, \hat{W}^*) \tag{2.68}$$

Note that the optimal weighting matrix has to be used. For large values of J the null for correctness of the restrictions will be rejected.

The J-test is valid only if the parameters of the model are identified. Alternative statistic, which allows inference when there are identification problems, is suggested in Stock and Wright (2000). It is based on the following asymptotic result:

$$TS_1(\theta_0, \hat{\tau}, W^*(\theta_0)) \xrightarrow{d} \chi^2_{m^2} \tag{2.69}$$

This is true as long as $\hat{\tau}$) is asymptotically normally distributed, and, in particular, does not assume that θ_0 is identifiable (see Theorem 2 in Stock and Wright (2000)). Using this result we can construct asymptotically valid $100(1 - r\%)$ confidence sets as $\big(\theta_0 : TS_1(\theta_0, \hat{\tau}, W^*(\theta_0)) \leq \chi^2_{m^2,\ r}\big)$.

2.5 Monte Carlo experiments

In this section I report the results from several Monte Carlo experiments. The goals of the experiments are: (1) to relate the asymptotic properties of the minimum distance estimators discussed in the previous section to situations when the size of the sample is small; (2) to compare the performance of the minimum distance estimation approach to that of maximum likelihood estimation which is the most widely used approach in the current empirical DSGE literature. I are particularly interested in collecting evidence on the relative performance of the estimators for models with weakly identified structure; (3) to evaluate the practical importance of using covariance restrictions in estimation.

2.5.1 Design of the experiments

The simulated data for all experiments comes from the log-linearized version of a small monetary New Keynesian model; it consists of the following three equations:

$$\pi_t = \alpha_f E_t \pi_{t+1} + \alpha_b \pi_{t-1} + \psi y_t + \varepsilon_t \tag{2.70}$$

$$y_t = \beta_f E_t y_{t+1} + \beta_b y_{t-1} - \lambda \left(r_t - E_t \pi_{t+1} \right) + \nu_t \tag{2.71}$$

$$r_t = \rho_\pi \pi_t + \rho_y y_t + \rho_r r_{t-1} + \zeta_t \tag{2.72}$$

Details of the derivation of these equations are presented in the appendix. Equation (2.70) is a log-linearized aggregate supply equation relating inflation at time t - π_t, to the conditional expectation of inflation in the next period, output gap - y_t, inflation in the last period, and an aggregate supply shock - ε_t. Equation (2.71) is an aggregate demand equation derived from the intertemporal optimization problem of a consumer with external habit formation. It relates the output gap at time t to its expectation for the next period, the real ex-ante interest rate -$(r_t - E_t \pi_{t+1})$, the

lagged output gap, and an aggregate demand shock - ν_t. The last equation (2.72) is a Taylor-type monetary policy rule according to which the nominal interest rate at time t - r_t responds to changes in the inflation, output gap, and also depends on the lagged nominal interest rate. All structural shocks - ε, ν, and ζ are assumed to be mutually and serially uncorrelated white noise precesses.

The vector of structural parameters that will be estimated is $\gamma = [\alpha_f, \alpha_b, \beta_f, \beta_b, \lambda, \rho_\pi, \rho_y, \rho_r, \sigma_\varepsilon, \sigma_\nu, \sigma_\zeta]'$. The monte carlo experiments differ in the values of γ used to generate data for $Z = [\ \pi \quad y \quad r\ ]'$. I present results for the three different parameterizations - P1, P2, and P3, shown in Table 2.2. They were chosen as examples of environments with different degrees of parameter identifiability. The second column in the table shows the condition number of matrix Ξ_2; I use this as an indicator of how well identified is γ. The condition number of Ξ_2 tells us how far will the estimate $\hat{\gamma}_1$ be when $\hat{\Xi}_1$ and $\hat{\Xi}_2$ are different from the true values Ξ_1 and Ξ_2. To compute the condition number I use the Matlab ®command `cond`; larger values of `cond` mean that the identification of the models is weaker. Each experiment consists of generating 1000 samples of size $T = 150$. The samples are obtained by solving the model (2.70)-(2.72) and using the reduced-form matrices to generate 1150 observations of Z; only the last 150 observations are used in estimation to eliminate the influence of the initial conditions.

Parm	Cond	α_f	α_b	γ	β_f	β_b	λ	ρ_π	ρ_y	ρ_r	σ_ε	σ_ν	σ_ζ
P1	9.8	.600	.400	.15	.50	.30	.500	.200	.010	.900	1	1	1
P2	320.1	.627	.000	.01	.48	.52	.005	.254	.095	.845	1	1	1
P3	452.7	.100	.000	.17	.50	.30	.500	.254	.095	.845	1	1	1

Table 2.2: Details on the Monte carlo design. Cond refers to the the condition number of matrix Γ_2, and is a measure for parameter identifiability; high values indicate weak identification.

2.5.2 Estimators Considered

I estimate the parameters of the model in (2.70)-(2.72) using three different estimators - two variants of minimum distance (MD1 and MD2), and a maximum likelihood estimator (ML). Each estimator is applied with and without making use of the covariance restrictions, resulting in six different estimates for γ. When the covariance restrictions are not used in estimation, I first estimate the parameters in γ without the three σ's, and then use the covariance restrictions to solve for the σ's.

- MD1 is obtained by first estimating the reduced-form VAR, and then minimizing the objective function in (2.47). When only the cross-equation restrictions are used, the estimator has a closed-form solution; if the covariance restrictions are also used, the objective function is minimized numerically.
- MD2 also uses the estimates of the reduced-form VAR, but is based on minimizing the function in (2.54). I use Chris Sims' program `gensys` to solve numerically for $h(\theta)$.
- the ML estimator is obtained my directly maximizing the likelihood of reduced form VAR, subject to the restriction implied by the model. The likelihood is computed by assuming that the structural shocks follow a multivariate gaussian distribution.

All estimators except MD1 without using the covariance restrictions, require numerical optimization for which I use Chris Sims' program `csminwel`. Only samples for which all optimizations are successful are counted and used in computing the Monte Carlo statistics.

2.5.3 Monte Carlo Results

The Monte Carlo results are presented in Tables 2.3 to 2.5. For each of the three experiments I report the Mean Bias and the Root Mean Squared Error (RMSE). In the appendix I also provide tables with Mean and Median, Standard deviation (S.D.), and the Mean Absolute Error (MAE) of the estimates. For each estimator,

Table 2.3: Monte Carlo results for P1

	Bias						RMSE					
par.	MD1		MD2		ML		MD1		MD2		ML	
α_f	-0.046	-0.076	0.557	-0.099	-0.106	-0.091	0.835	0.762	1.737	0.782	0.879	0.766
γ	0.038	0.042	0.361	0.082	0.134	0.117	0.462	0.380	1.006	0.357	0.534	0.455
α_b	-0.012	0.027	-0.330	0.020	-0.043	-0.048	0.237	0.198	1.016	0.252	0.493	0.475
β_f	-0.031	0.125	0.456	0.083	0.027	0.092	0.984	0.755	1.514	0.649	0.995	0.667
β_b	-0.024	0.032	-0.060	0.024	0.012	-0.003	0.231	0.169	0.425	0.182	0.285	0.249
λ	0.079	-0.102	-0.013	-0.058	0.023	-0.051	0.835	0.671	0.765	0.545	0.751	0.544
ρ_π	0.045	0.057	-0.930	0.022	-0.022	-0.099	0.641	0.423	2.888	0.435	1.345	1.249
ρ_y	-0.014	-0.013	0.763	0.014	0.064	0.106	0.493	0.305	2.256	0.344	1.033	0.960
ρ_r	0.014	0.080	-0.752	0.081	0.007	0.000	0.434	0.405	2.103	0.434	0.857	0.825
σ_ε	0.608	-1.053	1.715	-0.344	0.691	0.020	2.723	1.583	4.745	1.186	2.815	1.083
σ_ν	0.337	-0.349	3.113	-0.157	0.683	0.089	1.699	1.335	8.210	1.258	3.389	1.304
σ_ξ	-0.024	-0.326	3.271	-0.348	0.537	-0.026	1.351	1.250	10.188	1.393	4.830	1.221

Note: Bias and RMSE were multiplied by 10 to improve readability

Table 2.4: Monte Carlo results for P2

	Bias						RMSE					
par.	MD1		MD2		ML		MD1		MD2		ML	
α_f	-0.332	-0.491	-0.587	-0.463	-0.731	-0.457	2.858	2.521	3.730	2.476	3.635	2.455
γ	0.013	0.015	0.012	0.018	0.024	0.018	0.076	0.071	0.163	0.071	0.095	0.070
α_b	-0.150	-0.066	-0.116	-0.193	-0.178	-0.196	0.946	0.846	0.840	0.849	0.877	0.849
β_f	-0.010	-0.017	0.206	0.023	0.023	0.023	0.144	0.144	0.877	0.296	0.303	0.297
β_b	0.012	0.013	-0.202	-0.031	-0.031	-0.031	0.142	0.139	1.047	0.407	0.417	0.410
λ	0.005	-0.003	-0.156	-0.017	-0.016	-0.017	0.031	0.028	0.676	0.231	0.230	0.232
ρ_π	-1.007	-0.015	-1.523	0.000	-1.581	-0.001	4.615	0.758	5.657	0.730	6.156	0.827
ρ_y	0.033	0.003	0.045	0.003	0.048	0.003	0.140	0.038	0.167	0.037	0.180	0.039
ρ_r	-0.011	-0.003	-0.011	-0.004	-0.012	-0.004	0.069	0.052	0.070	0.052	0.066	0.052
σ_ε	0.331	-0.759	4.010	-0.418	0.805	0.008	1.340	1.452	15.895	1.247	5.730	1.009
σ_ν	0.071	-0.424	0.067	-0.254	0.049	-0.047	1.466	1.466	2.124	1.396	1.461	1.390
σ_ξ	1.872	-0.423	2.947	-0.432	3.535	-0.160	3.156	1.047	5.149	1.047	5.778	0.978

Note: Bias and RMSE were multiplied by 10 to improve readability

the first column reports the results when the covariance restrictions are not used in the estimation, and the second column shows results when all restrictions are used.

Overall, the three estimators perform very well for the data generating processes considered, and the relative performance is quite similar. In terms of precision, in

most cases MD1 has lowest bias when covariance restrictions are not used in estimation, while MLE is least biased when all restrictions are used. MD2 exhibits relatively large bias when only the cross-equation restrictions are used, but is similar to the other two estimators when all restrictions are utilized. The picture is similar when RMSE are compared: MD1 is slightly better for most parameters when the covariance restrictions are not used in estimation, and MLE performs best if the covariance restrictions are used.

Whether the covariance restrictions are used in estimation affects not only the ranking of the estimators but also improves substantially their performance in terms of RMSE. For all estimators and most parameters the RMSE decreases when the covariance restrictions are used. Notable exceptions are the parameters α_f and ψ when estimated with MD1 for data generated by P3; for that case RMSE increases from 4.96 and .69 to 5.1 and .71 respectively. This observation is interesting because the weak identification in parameterizations P3 is mostly due to weak identifiability of the parameters in equation (2.70). This can be seen by comparing the RMSE's for the parameters in that equation for the three parameterizations. This suggests that using only the cross-equation restrictions may be better for the MD1 estimator when some of parameters are weakly identifiable. Apart for the last observation, the results from the experiments do not favor one estimation method over the others in situations with weak identification. As a practical matter, one may find it easier to apply MD2 or MLE when using the covariance restrictions in estimation. The reason is that computing the weighting matrix for MD1 may makes its implementation more complicated. In contrast, the weighting matrix for MD2 is obtained directly from the first stage estimation.

Table 2.5: Monte Carlo results for P3

	Bias						RMSE					
par.	MD1		MD2		ML		MD1		MD2		ML	
α_f	1.644	-0.305	0.279	-0.346	-0.269	-0.368	4.955	5.100	5.958	4.771	5.897	4.763
γ	-0.143	0.097	0.118	0.104	0.121	0.106	0.687	0.709	0.789	0.656	0.776	0.654
α_b	-0.030	-0.060	-0.093	-0.101	-0.104	-0.100	0.888	0.853	0.887	0.839	0.875	0.832
β_f	0.179	0.069	0.264	0.019	0.103	0.022	1.180	0.592	1.387	0.436	1.084	0.425
β_b	-0.038	0.003	-0.011	0.000	-0.022	-0.005	0.422	0.232	0.448	0.228	0.422	0.227
λ	-0.136	-0.117	-0.101	-0.056	-0.077	-0.045	0.775	0.589	0.607	0.414	0.579	0.403
ρ_π	-0.705	0.117	-2.113	0.110	-1.336	0.103	4.711	0.912	7.099	0.897	6.838	0.878
ρ_y	0.134	-0.014	0.401	-0.011	0.257	-0.008	0.898	0.261	1.346	0.259	1.304	0.258
ρ_r	-0.006	0.055	-0.040	0.061	0.036	0.049	0.467	0.439	0.564	0.442	0.470	0.438
σ_ε	0.554	-0.633	0.840	-0.297	0.614	0.035	2.081	1.277	2.524	1.013	2.076	0.981
σ_ν	-0.043	-0.568	0.109	-0.485	-0.085	-0.312	1.440	1.377	1.968	1.303	1.476	1.267
σ_ξ	1.987	-0.342	4.854	-0.327	4.403	-0.116	3.236	1.153	7.193	1.146	6.743	1.105

Note: Bias and RMSE were multiplied by 10 to improve readability

2.6 Conclusion

In this paper, I take as a starting point the cross-equation and covariance restrictions that characterize linearized DSGE models, and showed how they can be used to study the identification, and estimate the parameters of such models. I derived a set of identifiability conditions, and suggested a procedure for a thorough analysis of identification at each point in the parameters space. The procedure can be applied, before DSGE models are estimated, to determine where identification fails, and where it likely to be weak, as well the equations of the model where the identification problems originate. I also examined an estimation approach based on the relationship between the structural and reduced form of DSGE models. In a first step the unrestricted reduced-form of the model is estimated, and in the second step the parameters of interest are obtained so that the model-implied restrictions hold as closely as possible. The approach is asymptotically equivalent to full information maximum likelihood information, and the evidence from a Monte Carlo study showed that it performs well in small samples for the models used there.

A useful extension of the identification results presented here would be to find a systematic way for attributing detected identification issues to specific model parameters. With the results in this paper we can generally do that only at the level of specific equations. For instance, if a rank deficiency, signifying identification failure, or near rank deficiency, suggesting weak identification, is found, we are able to find the particular equations that are responsible, but not the specific parameters in these equations. Since the problems are always caused by a few parameters in these equations, it would be very useful to be able to find out what they are.

A second extension, which I also leave for future work, is to develop a formal framework for evaluation of DSGE models, based on the two step estimation approach used in this paper. Unlike the likelihood approach, which is widely applied for estimating DSGE models, the two-step method used here does not impose the model's restrictions on the data when estimating the parameters. This makes it possible to evaluate different aspects of the estimated model, to detect specification failures, and suggest possible directions for improvement of the theoretical model.

Chapter 3

Evaluating the Information Matrix in linearized DSGE models

3.1 Introduction

Much of the growing literature on empirical evaluation of dynamic stochastic general equilibrium (DSGE) models uses likelihood-based estimation methods. After the equilibrium conditions of the model are linearized, solving it results in a linear state space model the likelihood of which can be evaluated using the Kalman filter. Researchers following the classical tradition estimate their models by maximizing the likelihood function. The Bayesian approach combines the likelihood with the researcher's prior beliefs to find the posterior distribution of the parameters.

Implementing either of the two approaches in practice requires evaluating the second derivative (the Hessian) of the log of the likelihood function. In MLE the asymptotic covariance matrix equals the information matrix, given by the negative of the expected value of the Hessian. Moreover, maximizing the likelihood using gradient based algorithms is most efficient when the inverse of the Hessian is used to direct the search in each iteration. Similarly, Bayesian methods such as Metropolis algorithm (see e.g. Schorfheide (2000)) or Importance sampler (see e.g. DeJong, Ingram, and Whiteman (2000)), use the Hessian to draw from the posterior distribution.

For most DSGE models the mapping from deep to reduced-form parameters is highly non-linear and can only be found numerically. This makes impossible deriving an analytical expression for the Hessian by direct differentiation the likelihood function. Due to this, researchers have been resorting to numerical derivatives. Using numerical differentiation in estimation is known to have important shortcomings. Doing it at each step of optimization procedure leads to large computation cost. More importantly, it can be highly inaccurate when the likelihood function is not well-behaved, and hence lead to unreliable results. It is therefore desirable to use analytical derivatives whenever this is feasible.

In this note I show how the information matrix can be derived analytically for linear DSGE models. My approach is based on the observation that the model is a standard linear state space model with non-linear restrictions on the parameters. Thus, the asymptotic covariance can be expressed in terms of the asymptotic covariance matrix of the unrestricted model, and the gradient of the mapping relating deep and reduced-form parameters. Both of these can be derived analytically. In particular, even though the mapping from deep to reduced-form parameters cannot be found explicitly, there is a simple implicit relationship between the two. A straightforward application of the implicit function theorem provides the necessary gradient.

In the next section I introduce the framework and present my main result. Section 3 provides a simple illustration. Some concluding remarks are given in section 3.

3.2 The information matrix in linearized DSGE models

Linearized DSGE models

A typical DSGE model after linearization can be cast in the following standard form

$$\Gamma_0 Z_t = \Gamma_1 E_t Z_{t+1} + \Gamma_2 Z_{t-1} + \Gamma_3 U_t \tag{3.1}$$

where Z_t is a $m \times 1$ vector of endogenous variables, and the structural errors, U_t, are i.i.d. n-dimensional random vectors with $E[U_t] = 0,\ E[U_t U_t'] = I$. The coefficient matrices $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3 are functions of the $k \times 1$ vector of deep parameters θ.[1]

Under certain conditions, which I assume to hold (see for example Sims (2002) for details), the model in (3.1) has a unique solution given by the following reduced-form model

$$Z_t = AZ_{t-1} + BU_t \tag{3.2}$$

$$Y_t = CZ_t \tag{3.3}$$

where Y_t is a p-dimensional vector of observable variables, and C is $p \times m$ matrix of known constants. Equation (3.2) is the state equation, and equation (3.3) is the measurement equation of the state space system.

Let the $(m^2 + mn)$-dimensional parameter vector τ be composed as

$$\tau = [\mathbf{vec}(A)',\ \mathbf{vec}(B)']' \tag{3.4}$$

Solving the linear DSGE model (3.1) means finding, for a given value of the deep parameters θ, a unique value of the reduced-form parameters τ. I denote the function

[1] I distinguish between structural parameters - the a priori unknown elements of $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3, and deep parameters. See the Appendix for more on that.

mapping θ into τ with h, that is

$$\tau = h(\theta) \tag{3.5}$$

Also, let $H = h_\theta(\theta)$ be the gradient of h. With the exception of very simple models, h is imposable to find analytically. Instead, one has to use numerical algorithms such as Chris Sims' `gensys` (Sims (2002)).

Assuming that U_t is normally distributed, the conditional log likelihood function $l(Y, \theta)$ can be derived using the Kalman filter. I am interested in finding the information matrix $\mathcal{I}_\theta$ defined as

$$\mathcal{I}_\theta = -E[l_{\theta\theta}(Y, \theta)]$$

where $l_{\theta\theta}$ is the second derivative of the log likelihood function with respect the vector of deep parameters.

Note that θ affects $l(Y, \theta)$ through τ only. This implies that $\mathcal{I}_\theta$ can be expressed in the following way

$$\mathcal{I}_\theta = H'\mathcal{I}_\tau H \tag{3.6}$$

where $\mathcal{I}_\tau$ is the information matrix of the unrestricted model.

This result was shown in the seminal work of Rothenberg (1966), and holds in general for estimation subject to equality constraints. The implication for us is that, instead of differentiating $l(Y, \theta)$ with respect to θ directly, we can work with the matrices H and $\mathcal{I}_\tau$. This is useful only if these two matrices are easier to derive.

Several papers have derived analytical expressions for the information matrix of linear state space models. Most recently Klein and Neudecker (2000) use matrix differentiation to obtain the matrix explicitly (see also Andre Klein and Zahaf (2000)). Segal and Weinstein (1989) provide an method for evaluating the information matrix as well as the score and the Hessian of the log likelihood for state space models. Such algorithms are also derived in Zadrozny (1989) and Terceiro (1990).

Since function h is not available analytically, we cannot compute its gradient H by direct differentiation of h. However, as I show next, H is straightforward to find with the help of the Implicit function theorem.

From equation (3.2) we have $E_t Z_{t+1} = AZ_t$. Thus equation (3.1) can be written as

$$\Gamma_0 Z_t = \Gamma_1 A Z_t + \Gamma_2 Z_{t-1} + \Gamma_3 U_t \tag{3.7}$$

or

$$(\Gamma_0 - \Gamma_1 A) Z_t = \Gamma_2 Z_{t-1} + \Gamma_3 U_t \tag{3.8}$$

Comparing the last equation to the one in (3.2) results in the following two sets of equations:

$$(\Gamma_0 - \Gamma_1 A) A - \Gamma_2 = 0 \tag{3.9}$$

$$(\Gamma_0 - \Gamma_1 A) B - \Gamma_3 = 0 \tag{3.10}$$

The system in (3.9)-(3.10) defines an implicit function f from $\mathbb{R}^{k+m^2+mn}$ to $\mathbb{R}^{m^2+mn}$, such that $F(\theta, \tau(\theta)) = \mathbf{0}$. Then an application of the Implicit Function Theorem gives us H as the following lemma shows.

Lemma 3.2.1. *Suppose $f(\theta, \tau)$ is continuously differentiable in an open neighborhood of (θ_0, τ_0), and $f(\theta_0, \tau_0) = 0$. Then if $|f_\tau(\theta_0, \tau_0)| \neq 0$, there exists an open neighborhood $\mathbf{W}$ of (θ_0, τ_0), such that*

$$H \equiv \frac{\partial \tau}{\partial \theta'} = -\big(f_\tau(\theta, \tau)\big)^{-1} f_\theta(\theta, \tau) \tag{3.11}$$

for all $(\theta, \tau) \in \mathbf{W}$

The smoothness condition for $f(\theta, \tau)$ is satisfied for most DSGE models, and fol-

lows from the fact that the matrices Γ_i, A and B are typically smooth functions of θ and τ, respectively. An exception are the DSGE models with regime switching (see Svensson and Williams (2005)). The other condition - that the determinant $|f_\tau(\theta_0, \tau_0)|$ is not zero, is necessary and sufficient for the inverse matrix in (3.11) to exist.

3.3 Illustration

The following simple example illustrates my approach. Suppose that the DSGE model (3.1) is given by

$$Z_t = \alpha E_t Z_{t+1} + (1-\alpha) Z_{t-1} + \sigma U_t \tag{3.12}$$

where Z_t is univariate, $\alpha > .5$ and $U_t \sim \mathbb{N}(0,1)$. The reduced-form solution is

$$Z_t = AZ_{t-1} + BU_t \tag{3.13}$$

In terms of the notation used above, we have $\theta = [\alpha, \sigma]'$, $\Gamma_0 = 1$, $\Gamma_1 = \alpha$, $\Gamma_2 = 1-\alpha$, $\Gamma_3 = \sigma$, and $C = 1$. Here the state variable Z_t is observed (i.e. $Y_t = Z_t$), and therefore the information matrix for the reduced-form parameter vector $\phi = [A, B]'$ is straightforward to compute. Moreover, one can solve by hand for the reduced-form coefficients to find

$$A = \frac{1-\alpha}{\alpha}, \quad B = \frac{\sigma}{\alpha} \tag{3.14}$$

Hence H can be easily computed by direct differentiation

$$H = \frac{\partial \phi(\theta)}{\partial \theta'} = \begin{bmatrix} -1/\alpha^2 & 0 \\ -\sigma/\alpha^2 & 1/\alpha \end{bmatrix} \tag{3.15}$$

The information matrix $\mathcal{I}_\theta$ can then be derived using (3.6). As was mentioned above, however, for a typical DSGE model one cannot solve analytically for A and B. Yet,

as I have shown, H can be computed analytically, using the implicit function defined by (3.9) and (3.10), and appealing to the Implicit function theorem. The implicit function in the example is

$$F(\theta, \phi(\theta)) = \begin{bmatrix} A - \alpha A^2 - 1 + \alpha \\ B - \alpha AB - \sigma \end{bmatrix} = \mathbf{0} \tag{3.16}$$

To apply the Implicit function theorem we need $F_\phi(\theta, \phi(\theta))$ and $F_\theta(\theta, \phi(\theta))$, given by

$$F_\phi(\theta, \phi(\theta)) = \begin{bmatrix} 1 - 2\alpha A & 0 \\ -\alpha B & 1 - \alpha A \end{bmatrix}, \quad F_\theta(\theta, \phi(\theta)) = \begin{bmatrix} -A^2 + 1 & 0 \\ -AB & -1 \end{bmatrix} \tag{3.17}$$

For a general linearized DSGE model, function F and its Jacobians F_ϕ and F_θ are easily derived using software for symbolic computations. The Implicit function theorem gives us H from Lemma (3.11). It is straightforward to show that substituting (3.17) and (3.14) into the formula for H yields the same expression for H as in (3.15).

3.4 Concluding remarks

In this note I show how the information matrix of linearized DSGE models can be evaluated analytically. This is a useful result because it eliminates the need for numerical differentiation when estimating such models by either classical or Bayesian methods, and thus can be expected to improve their speed and reliability. Another important application of this result is for checking local identifiability in DSGE models. Rothenberg (1971) showed that necessary and sufficient condition for this is the information matrix to have a full rank. Canova and Sala (2006) advocate the use of the information matrix approach for studying identification in parameters in DSGE models.

Chapter 4

How much do we learn from the estimation of DSGE models? A case study of identification issues in a New Keynesian business cycle model

4.1 Introduction

In this Chapter I extend the identification approach described in Chapter 2 to situations where the assumption for identifiability of the reduced form does not hold. The necessity for such an extension arises in large-scale DSGE models for which the dimension of τ - the vector of reduced form parameters, is much larger than that of θ. As a result τ is typically not identifiable[1], as a vector, from the likelihood, and it is generally difficult to determine which elements of τ or linear combinations of those elements, are identifiable, and which are not. As a consequence, all identification condition derived in Chapter 2 are only necessary and not sufficient, when

[1]This follows from the observation that for a general linear state space model

$$X_t = CZ_t \tag{4.1}$$
$$Z_t = AZ_{t-1} + BU_t \tag{4.2}$$

$\bar{C} = CD^{-1},\ \bar{A} = DAD^{-1},\ \bar{B} = DB$ is observationally equivalent to $C,\ A,\ B$ for any invertible matrix D of the same dimension as A

the assumption that τ is identifiable does not hold. The conditions are nevertheless useful for detecting lack of identification: if θ cannot be solved for uniquely when τ is known, it cannot be identified when only some of the elements of τ or linear combinations of those elements are known. But an alternative approach is required to obtain a necessary and sufficient condition for identification.

Such a condition is provided by a well-known result of Rothenberg (1971) who showed that the non-singularity of the Information matrix is a necessary and sufficient condition for parameter identification in parametric models.

In Chapter 3 I showed how the information matrix can be evaluated analytically for linearized DSGE models. I factorize the information matrix for θ as a product of two terms: one is the gradient of the mapping from reduced-form parameters τ to deep parameters θ; the second is the information matrix of the reduced-form model. Both factors can be derived and evaluated analytically. This approach not only makes a precise evaluation of the information matrix possible, but also provides a necessary condition for identification of the deep parameters, which does not depend on the data. The condition is that the gradient of the mapping from θ to τ has full rank. This mapping is completely independent from the data used in estimation. Thus, I can detect identification problems that are inherent in the structure of the DSGE model, and not caused by data deficiencies.

Identification problems may arise in the model for two reasons. First, the reduced-form solution of the model, which represents the equilibrium law of motion for the state variables, may be insensitive to changes in a deep parameter. This would make the likelihood surface very flat with respect to that parameter, thus rendering it poorly identified. Second, the changes in the reduced-form model resulting from changes in a deep parameter may be well approximated by changes in a combination of other deep parameters. This would make the first parameter difficult to distinguish from the other deep parameters in the model. The likelihood would again be flat, but this

time in the direction of a linear combination of deep parameters. The decomposition of the Information matrix I propose makes it possible to find out exactly which parameters are poorly identifiable for either of the reasons described above.

Knowing how to evaluate the Information matrix allows us to determine the identifiability of any value of θ in the parameter space of the model. This is useful both for post-estimation and pre-estimation analysis. After the model has been estimated, we may want to know how well identified are the point estimates, and thus how reliable are the estimated standard errors and confidence intervals. If Bayesian techniques are applied for estimation, the Information matrix can be used to assess the importance of priors relative to the data.

More generally, both applied researchers and macroeconomic modelers may be interested in knowing how well-identified is a particular DSGE model, *before* it is taken to the data. Such analysis could reveal, for instance, that there are features of the model that make it unidentifiable, or poorly identified, and this does not depend on particular data used for estimation. Or, we may find that the parameters θ are well-identified in some parts of the parameter space of the model, while in others identification is poor. In order to study the identifiability of the theoretical model, we have to examine the Information matrix everywhere in the parameter space, that is, at all a priori admissible parameter values.

I illustrate both types of identification analysis using a large-scale New Keynesian business cycle model with various real and nominal frictions developed and estimated in Smets and Wouters (2007). [2] In particular, in Section 4 I draw a large number of points from the parameter space of the model, and check the necessary and sufficient

[2]Similar models have been studied, using Bayesian techniques, in Onatski and Williams (2004), DelNegro, Schorfheide, Smets, and Wouters (2005), Justiniano and Primiceri (2006), and Boivin and Giannoni (2006). Previous research suggests that the model fits the data well, in some cases outperforming unrestricted vector autoregressions in out-of-sample forecasting. Schmitt-Grohe and Uribe (2004) and Levin, Onatski, Williams, and Williams (2005) study the design of optimal monetary policy rules in estimated versions of that model (see also Juillard, Karam, Laxton, and Pesenti (2006)).

rank conditions at each one of them. In addition, I evaluate the conditioning of the matrices whose ranks determine identification. A poorly conditioned matrix is one close to being of reduced rank. Thus I determine not only whether the parameters are identifiable in the strict sense, but also how strong is identification. I do this for six parameterizations that differ in which parameters are assumed to be known. Then I turn to the estimation of parameters using quarterly US data. I depart from most of the previous empirical DSGE literature by using maximum likelihood for estimation of the model. This allows me to compare parameter estimates driven by the data only, with those obtained with Bayesian methods, which are determined by both the data and the prior distribution. When the number of observations is large, the two approaches should produce similar results. In small samples, however, the prior distribution could be very influential, especially when identification is weak. This may result in parameter estimates that have little to do with the actual data used for estimation.

On the identification side, I find that Smets and Wouters (2007), who state that three of the deep parameters of the model are not identifiable, are correct only with respect to two of them. The third one - the steady state wage markup parameter, is, identifiable, though generally very weakly so. When I restrict the analysis to identifiable parameterizations, I find that identification is generally quite weak. I show that the problem to a large degree originates in the structure of the model, and thus cannot be resolved by using more informative data. Furthermore, I are able to determine which of the deep parameters are most responsible for the weak identifiability of the model as a whole. The set of worst identifiable parameters varies somewhat across the parameter space, but ten of them are very poorly identified virtually everywhere. These parameters are: elasticity of labor supply, coefficients of price and wage stickiness, steady state wage markup, habit persistence, elasticity of intertemporal substitution, fixed cost of production, and the coefficients of monetary policy

response to output, inflation and lagged interest rates. The problem with these parameters is that their role in the model can be very well approximated by other deep parameters. I also show that this problem is not easily solved by assuming that a few of these parameters are known. For instance, to reduce the parameter interdependence problem for the wage stickiness coefficient, one may have to assume that up to eight other deep parameters are known, instead of estimating them. My analysis thus provides concrete evidence for the notion that models of this scale are severely overparameterized.

On the estimation side, I find that disposing with the strong priors used in previous studies affects substantially the estimates of the parameters in the model. This has some important implications for the behavior of the model, as I show using impulse response and variance decomposition analysis.

This paper is not the first to systematically study parameter identification in DSGE models. An important recent contribution that deals exclusively with these issues is Canova and Sala (2006). There are three important differences between their study and the present paper. First, they approach parameter identification from the perspective of a particular limited information estimation method, namely, impulse response matching (see Rotemberg and Woodford (1998), and Altig, Christiano, Eichenbaum, and Linde (2005) for explanation and illustration of this estimation approach). As they recognize, identification failures of that or other limited information methods do not imply that the problems are generic to all estimation methods. In contrast, if identification fails or is weak when a full information approach is used, as I do here, it will remain a problem for any alternative estimation method. Second, unlike this paper, which evaluates the information matrix analytically, Canova and Sala (2006) use numerical approximation of the Hessian. It is well-known that numerical differentiation could be very imprecise for highly non-linear functions, as is

the case with DSGE models.[3] Moreover, with my approach for computing the Information matrix, I am able to distinguish between the model structure and the data as sources of identification problems. Finally, unlike Canova and Sala (2006), who study identification only in the neighborhood of a particular point in the parameter space, I study the identifiability of a large number of points drawn randomly from everywhere in the space. Thus I am able to characterize parameter identification as a global instead of a local problem of the theoretical model.

Regarding the effect of priors for Bayesian estimation of DSGE models, results similar to mine are reported in Onatski and Williams (2004). They estimate a similar large-scale New Keynesian model, using European data, and find that greater prior uncertainty results in substantially different parameter estimates, compared to those obtained with the tighter priors common in the empirical DSGE literature. They do not address formally the issue of parameter identifiability, as I do in this paper.

The rest of the paper is organized as follows. Section 4 explains my approach to identification. There I show how the Information matrix can be computed analytically, and outline a procedure for studying model identification in general linearized DSGE models. I also discuss the difference between identification in a strict sense, and weak identification as a finite sample phenomenon, and explain, using a simple example, the role of the model and the data in determining the strength of identification. In section 4 I apply the proposed identification analysis procedure to the model of case study. The main results are in 4 where I determine which parameters are not well identified in the model and why. In section 4 I use the data from Smets and Wouters (2007) to find the maximum likelihood estimate of the model, and compare the results to the Bayesian estimates reported in Smets and Wouters (2007). I also compare, using impulse response and variance decomposition analysis, the economic implications of the different parameter estimates. The last section offers some

[3]Hansen, McGrattan, and Sargent (1994) also argue in favor of using analytical derivatives when estimating DSGE models

concluding remarks and directions for future research.

4.2 Identification in DSGE models

4.2.1 Structural and Reduced Form

Currently, most analyses involving either simulation or estimation of DSGE models use linear approximations the original models. Specifically, the model is first expressed in terms of stationary variables, and then linearized or log-linearized around the steady-state values of these variables.

Typically, the linearized system can be written in the form

$$\Gamma_0 Z_t = \Gamma_1 E_t Z_{t+1} + \Gamma_2 Z_{t-1} + \Gamma_3 U_t \tag{4.3}$$

where Z_t is a $m \times 1$ vector of endogenous variables, and the structural errors, U_t, are i.i.d. n-dimensional random vectors with $E[U_t] = 0,\ E[U_t U_t'] = I$. The coefficient matrices $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3 are functions of the $k \times 1$ vector of deep parameters θ.

There are several algorithms for solving linear rational expectations models like (4.3) (see for instance Blanchard and Kahn (1980), Anderson and Moore (1985), Klein (2000), Christiano (2002), Sims (2002)). Depending on the value of θ, there may exist zero, one, or many stable solutions. Assuming that a unique solution exists, it can be cast in the following form

$$Z_t = A Z_{t-1} + B U_t \tag{4.4}$$

where A and B are functions of θ, and are unique for each value of θ. I collect the reduced-form parameters in a τ, defined as

$$\tau = [\mathbf{vec}(A)',\ \mathbf{vec}(B)']'$$

I also define the define the function mapping θ into τ as

$$\tau = h(\theta)$$

The deep parameters of the model cannot be estimated directly from (4.4) as some of the variables in Z are not observed. Instead, I can write the reduced-form system in a state space form, with transition equation given by (4.4), and the following measurement equation

$$X_t = CZ_t \tag{4.5}$$

where X_t is a vector of observed state variables, and C is a known matrix.

Assuming that U_t is normally distributed, the conditional log likelihood function $l(X, \theta)$ can be computed recursively using the Kalman filter (see Hamilton (1994, ch.13)).

4.2.2 Identification of θ

Let Θ be the admissible parameter space of θ, that is, the set of all values of θ which conform to the restrictions postulated by the theoretical model. For each $\theta \in \Theta$, the DSGE model (4.3) is a data generating process for $X = \{X_t\}_{t=1}^{T}$. By the assumption of uniqueness of the solution (4.4) to (4.3), each admissible θ implies a unique joint probability density function $F(X; \theta)$ of the elements of X. Identification of θ requires that the inverse association is also unique. Specifically, $\theta_0 \in \Theta$ is globally identifiable if for any other $\theta_1 \in \Theta$, we have $F(X; \theta_0) \neq F(X; \theta_1)$ for some X with a non-zero probability measure. Local identification of θ_0, on the other hand, requires that the $F(X; \theta_0)$ is unique only in some neighborhood of θ_0. Clearly, local identifiability is necessary for θ to be globally identified. Finally, when all a priori admissible values

$\theta \in \Theta$ are (locally) identifiable, we say that the *model* is (locally) identified.

A well-known result from Rothenberg (1971, Theorem 1) is that a necessary and sufficient condition for local identification of θ_0 is that the information matrix, defined by

$$\mathcal{I}_\theta = E\left[\left\{\frac{\partial l(X;\theta)}{\partial \theta}\right\}'\left\{\frac{\partial l(X;\theta)}{\partial \theta}\right\}\right] = -E\left[\left\{\frac{\partial^2 l(X;\theta)}{\partial \theta \partial \theta'}\right\}\right]$$

has a full rank when evaluated at θ_0. Here $l(X;\theta) = \ln F(X;\theta_0)$ is log-likelihood function. Using this condition we can, in principle, determine the local identifiability of the model as a whole by evaluating the rank of the information matrix at all points of the parameter space.

The problem with applying this result to determine identifiability in DSGE models is that the mapping from θ to the log-likelihood function is, for most models, not available in analytical form. The likelihood function is determined by A and B, which have to be solved for numerically with some of the algorithms mentioned earlier. This makes it impossible to derive analytically the information matrix by direct differentiation of the log-likelihood function. Using numerical differentiation, on the other hand, is computationally very costly, and is known to be very inaccurate for highly non-linear functions which is typically the case for DSGE models. Not only is the function non-linear, but it has to be evaluated numerically in the first place.

In Chapter 2 I showed an alternative approach for evaluating the information matrix. It is based on a result by Rothenberg (1966) who showed that $\mathcal{I}_\theta(\theta, X)$ can be expressed in the following way[4]

$$\mathcal{I}_\theta = H(\theta)'\mathcal{I}_\tau(\theta, X)H(\theta) \tag{4.6}$$

where $\mathcal{I}_\tau(\theta, X)$ is the Information matrix of the unrestricted state space model, and

[4]This follows from a straightforward application of the rule for differentiating composite functions.

$H(\theta)$ is the gradient of h, i.e.

$$H(\theta) = h_\theta(\theta)$$

Both $H(\theta)$ and $\mathcal{I}_\tau(\theta, X)$ can be derived analytically. I outline the derivation of $H(\theta)$ below;

The first step in finding $H(\theta)$ is to realize that even though h cannot be written explicitly, we can find an implicit function relating θ and τ. From (4.3) and (4.4) and the law of iterated expectations we obtain the following two sets of equations (see the Appendix for details):

$$(\Gamma_0 - \Gamma_1 A)A - \Gamma_2 = 0 \tag{4.7}$$

$$(\Gamma_0 - \Gamma_1 A)B - \Gamma_3 = 0 \tag{4.8}$$

A and B depend on θ only through τ, while $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3 are functions of θ only. The expressions in (4.7) and (4.8) define an implicit function $F(\theta, \tau(\theta)) = 0$.[5] Therefore, by the Implicit function theorem,[6]

$$H = \frac{\partial \tau(\theta)}{\partial \theta'} = -\big(F_\tau(\theta, \tau(\theta))\big)^{-1} F_\theta(\theta, \tau(\theta)) \tag{4.9}$$

In practice, it is straightforward to compute F_θ and F_τ using standard packages for symbolic calculus. The computation is further simplified by the fact that F can be factored as[7]

$$F(\theta, \tau(\theta)) = F_1(\tau(\theta))F_2(\theta) \tag{4.10}$$

The approach described above is useful for two reasons. First, it avoids numerical

[5]Evaluating the matrix F proved to be an extremely useful method for detecting and correcting programming errors. See the Appendix for more details on this and a complementary method for doing that.

[6]To apply the implicit function theorem, we need the matrix $F_\tau(\theta, \tau(\theta))$ to be invertible. This was true for all admissible values of θ used in the identification analysis. See below for details.

[7]see the Appendix in Iskrev (2007a)

differentiation, and allows one to accurate evaluate the Information matrix. Second, it can also help in discovering the sources of the identification problems, if such exist. The roots of identification problems may be either in $\mathcal{I}_\tau(\theta, X)$, or $H(\theta)$, or both. The first matrix measures how well the reduced form parameters τ are identified, and depends, in part, on the properties of the data, as X is used in its calculation. $H(\theta)$, on the other hand, tells us how well-identified are the deep parameters θ given τ, and does not depend on the data. Therefore, finding a rank deficient, or poorly conditioned $H(\theta)$, means that θ is not identifiable, or is weakly identifiable, due to reasons inherent in the structure of the model. The distinction between the model and the data as causes for identification problems is relevant only as far as the strength of identification is concerned. For fully articulated economic models, such as DSGE models, the identifiability of parameters is completely determined by the structure of the model. This is because every aspect of the data generating process, and the likelihood of the model, can be traced back to the underlying deep parameters and structural relationships. As we know from the literature on weak instruments, however, how strong identification is has important implications for the small sample properties of estimators, as well as for inference. I find it useful, therefore, to distinguish between the role of $H(\theta)$, which depends on θ only, and $\mathcal{I}_\tau(\theta, X)$, which depends on both θ and the data.

The following simple example helps clarify the distinction between the model and the data as sources of identification problems.

4.2.3 Model vs. Data as Sources of Identification problems: Example

Suppose that the model (4.3) is given by

$$Z_t = \theta_1 E_t Z_{t+1} + (1 - \theta_1) Z_{t-1} + \theta_2 U_t \tag{4.11}$$

where Z_t is univariate, $\theta_1 > .5$ and $U_t \sim \mathbb{N}(0,1)$. The reduced-form solution is

$$Z_t = \tau_1 Z_{t-1} + \tau_2 U_t \tag{4.12}$$

In terms of the notation used above we have $\Gamma_0 = 1$, $\Gamma_1 = \theta_1$, $\Gamma_2 = 1 - \theta_1$, $\Gamma_3 = \theta_2$, and $C = 1$. Here the state variable Z_t is observed (i.e. $X_t = Z_t$), and therefore the information matrix $\mathfrak{I}_\tau$ for the reduced-form parameter vector τ is straightforward to compute. Moreover, one can solve by hand for the reduced-form coefficients to find

$$\tau_1 = \frac{1-\theta_1}{\theta_1}, \quad \tau_2 = \frac{\theta_2}{\theta_1} \tag{4.13}$$

We can view the estimation of the deep parameters θ as a two-step procedure: first, estimate the reduced form parameters τ; second, given the estimated $\hat{\tau}$, solve for $\hat{\theta}$.

Therefore the following two conditions must be satisfied for θ to be well-identified:

1. τ can be precisely estimated;
2. small errors in $\hat{\tau}$ result in small errors in $\hat{\theta}$.

The first condition is determined by how informative is the particular realization of the data we observe. If, for example, the data is very noisy, or the sample very short, the standard errors of $\hat{\tau}$, and therefore of $\hat{\theta}$, will be large. The second condition is determined solely by the features of the model, and, more precisely, by the mapping in (4.13). If that mapping is poorly conditioned, small errors in τ would result in large errors in θ. In that case θ would be poorly identified for reasons particular to the structure of the model and not because of the data.

The intuition from this simple example extends to the general model. One can show that, when the parameters τ of the reduced form model are identified, the two step procedure described above is asymptotically equivalent to full information maximum likelihood estimation (see Iskrev (2007b)). Therefore features of the data

sample used in estimation, which reduce the quality of the reduced form estimates, will also cause poor identifiability of θ. Another such factor, in addition to the short sample size and noisiness of the data mentioned above, would be strong collinearity among the observed data series. This is known to cause problems with identification in the standard linear model, and has the same effect on the estimates of τ.

4.2.4 Identification vs. Weak Identification

Given a parameter value θ_0, computing the rank of $\mathcal{I}_\theta(\theta_0)$ would tell us whether θ_0 is identifiable or not. The matrix must be non-singular for a consistent estimator of θ to exist.

Verifying that the rank condition is satisfied, however, tells us nothing about the strength of identification. The is important in practice because much of the standard econometric theory is based on approximations that are valid only asymptotically. There is now a considerable body of literature showing that the quality of these approximations in finite samples, depends crucially on the strength of identification (see Dufour (2003) and the references therein). In particular, when the identification of the parameters is weak, the small sample properties of the estimators may be very poor, and the standard methods for constructing confidence intervals, and for testing hypothesis, are prone to be very inaccurate.

What do I mean by weak identification? To answer this question it helps to first analyze the causes for lack of identification. There are two possible reasons why a parameter θ_i may be unidentifiable:

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

$$\frac{\partial l}{\partial \theta_i} = 0, \quad \text{for all } X \tag{4.14}$$

(b) The change in the likelihood caused by changing θ_i can be offset by changing

other parameters in θ, i.e.

$$\frac{\partial l}{\partial \theta_i} = \sum_{j \neq i} a_j \frac{\partial l}{\partial \theta_j}, \quad \text{for all } X \tag{4.15}$$

where a_j, $j \neq i$ is a scalar.

In the first case row i and column i of the Information matrix will be vectors of zeros; in the second they will be equal to a linear combination of the other rows/columns of the Information matrix. The likelihood will be flat with respect to θ_i in the first case, and with respect to a linear combination of several parameters in θ - in the second.

If "=" in (4.14) and (4.15) is replaced by "$\approx$", θ_i will be weakly identified - the likelihood will be almost though not completely flat with respect to one or a combination of deep parameters. In the first case the value of θ_i is difficult to pin down from the likelihood, and its estimate will be very sensitive to random variation is the data; in the second case the estimates of several of the deep parameters will be highly correlated and again small changes in the data may result in substantial changes in the point estimates.

Note that the weak identification version of (a) is equivalent to a very small variance of i-th component of the score vector; likewise, (b) is equivalent to a strong linear dependence, or collinearity, among the components of the score. To separate these two causes for poor identification of θ I use the following factorization of the Information matrix

$$\mathcal{I}_\theta = D^{\frac{1}{2}} \tilde{\mathcal{I}}_\theta D^{\frac{1}{2}}, \quad D = \text{diag}(\mathcal{I}_\theta) \tag{4.16}$$

i.e. D is a diagonal matrix with the variance of $\frac{\partial l}{\partial \theta_i}$ in the (i,i)-th position, and $\tilde{\mathcal{I}}_\theta$ is the normalized Information matrix whose (i,j)-th element contains the correlation between the i-th and j-th component of the score vector.

As with the singularity of the Information matrix, which captures both of the

possible causes for lack of identification, weak identification in either of the two forms results in having an Information matrix which has full rank, but is close to being singular.[8] Unlike singularity, however, which is unambiguously determined by the rank of the matrix, near-singularity and therefore weak identification is harder to characterize.

An obvious candidate for a measure of identification strength is the condition number of the Information matrix. It can be shown that the condition number of a non-singular matrix A, defined by $\text{cond}(A) = \|A\|\|A^{-1}\|$, is equal to the inverse of distance of that matrix to the set of singular matrices.[9] Specifically

$$\frac{1}{\text{cond}(A)} = \min \left\{ \frac{\|\triangle A\|}{\|A\|} : A + \triangle A \text{ is singular} \right\} \tag{4.19}$$

Note that cond(.) depends on the underlying norm; if the Euclidean norm is used the condition number of a matrix is equal to the ratio of the largest to the smallest singular values of that matrix. From 4.19 it follows that the smaller is the condition number of the Information matrix, the further it is from singularity, and therefore the stronger is the identification of parameters. [10]

The condition number of $\mathcal{I}_\theta$ is an indicator of how informative the likelihood is for θ as a whole. It plays a role, in the multivariate case, similar to that of the value the Information matrix when θ is a scalar. In the univariate case, $\mathcal{I}_\theta = 0$ indicates that

[8]This is consistent with the notion of weakness in the weak instruments literature. For instance, in Moreira (2003) the structural parameters β in

$$y_{n\times 1} = Y_{n\times l}\beta + u \tag{4.17}$$
$$Y = Z_{n\times k}\Pi_{k\times l} + \epsilon \tag{4.18}$$

are said to be "almost unidentified when Π is in a small neighborhood around a matrix with rank less than l." (footnote 3). In a fully parametric setting this is equivalent to the Information matrix for β being close to singularity.

[9]This is known as the Eckart-Young theorem (see e.g. Demmel (1987)).

[10]In the standard linear regression model $y = X\beta + u$, the Information matrix is proportionate to $X'X$, and identification problems are caused by strong linear dependence among the columns of X. Belsley, Kuh, and Welsch (1980) suggest the use of the condition number of $X'X$ for detection of collinearity problems in this setting.

the likelihood does not change as we vary θ, i.e. the likelihood function is completely flat and θ is unidentifiable. When $\mathcal{I}_\theta > 0$ but is very small, the likelihood is almost flat, and thus θ is weakly identified. Similarly, in the multivariate case when the Information matrix is exactly singular, the condition number is infinity, and the likelihood function is absolutely flat in some directions, and is thus completely uninformative with respect to one or more parameters. An almost singular Information matrix, on the other hand, has a large condition number, and implies that the likelihood is nearly flat in some directions, and thus provides very little information for some parameters. We say that a matrix with low condition number is well-conditioned, and if the condition number is high, the matrix is poorly conditioned.

Because of the factorization in (4.16), the Information matrix will be poorly conditioned if either D or $\tilde{\mathcal{I}}_\theta$ or both are poorly conditioned. Large condition number of $\tilde{\mathcal{I}}_\theta$ indicates identification problems due to strong collinearity (see (4.15)). Note, however, that unlike the correlation matrix $\tilde{\mathcal{I}}_\theta$, D is not scale invariant. Its conditioning depends on magnitude of smallest and largest component of $\frac{\partial l}{\partial \theta}$, and therefore depends on the units with which the parameters in θ are measured. A unit-free measure is $\frac{\theta_i \partial l}{\theta_i}$ - the percentage change in the likelihood due to a 1% change in θ_i. Therefore, instead of matrix D I will check the conditioning of the matrix $\tilde{D}$ defined by

$$\tilde{D}(i,i) = \theta(i)^2 D(i,i) \tag{4.20}$$

A poorly conditioned $\tilde{D}$ indicates identification problems due to a low sensitivity of the likelihood with respect to some deep parameters; a poorly conditioned $\tilde{\mathcal{I}}_\theta$ indicates identification problems caused by a strong parameter interdependence with respect to the likelihood function.

The factorization of the information matrix into two scale-free component - $\tilde{D}$ and $\tilde{\mathcal{I}}_\theta$ may be used to shed light on the question of what constitutes a large condition number. In the Appendix I show that the asymptotic variance of the estimate of θ_i

can be expresses as

$$\text{var}(\hat{\theta}_i) = D(i,i)^{-1}\left(\frac{1}{1-\cos^2(i,-i)}\right)$$

where $\cos(i,-i)$ is the cosine of the angle between the i-th element of the score, and the space spanned by the other elements of the score. A large value of $\cos(i,-i)$ indicates strong parameter interdependence problem for θ_i. Dividing both sides by θ_i^2, we have

$$\frac{\text{var}(\hat{\theta}_i)}{\theta_i^2} = \left(\frac{1}{\tilde{D}(i,i)}\right)\left(\frac{1}{1-\cos^2(i,-i)}\right) \tag{4.21}$$

From (4.21) it follows that the normalized asymptotic variance of $\hat{\theta}_i$ will be large if either $\tilde{D}(i,i)$ is small, or if $\cos(i,-i)$ is close to one. Suppose first that $\cos(i,-i) = 0$, i.e. there is no parameter interdependence problem for θ_i. Let $\tilde{D}(m,m) = \min_i \tilde{D}(i,i)$ and $\tilde{D}(M,M) = \max_i \tilde{D}(i,i)$. Then

$$\frac{\text{var}(\hat{\theta}_M)}{\theta_M^2} = \text{cond}(\tilde{D})^2 \frac{\text{var}(\hat{\theta}_m)}{\theta_m^2}$$

That is, $\text{cond}(\tilde{D})$ is large, the normalized asymptotic variance of $\hat{\theta}_M$ is much larger than that of $\hat{\theta}_m$.

Next, consider the effect of the second term on the left hand side of (4.21). It tells us how much the intrinsic uncertainty in $\hat{\theta}_i$, represented by $\frac{1}{\tilde{D}(i,i)}$, is magnified because of parameter interdependence. For instance, if $|\cos(i,-i)| = .9$, $\frac{1}{1-\cos^2(i,-i)} = 5.26$. That is, because of the parameter interdependence, the normalized asymptotic variance of $\hat{\theta}_i$ is more than 5 times as large as what it would have been if there was no parameter interdependence. The condition number of $\tilde{\mathcal{I}}_\theta$ provides a bound on the values of $\cos(i,-i)$, namely[11]

[11]This result follows directly from the Kantorovich-Wielandt inequality, see Horn and Johnson (1985) for details

$$\max_i |\cos(i, -i)| \leq |\cos\left(2\cot^{-1}(\sqrt{\text{cond}(\tilde{\mathcal{I}}_\theta)})\right)|$$

For instance, if $\text{cond}(\tilde{\mathcal{I}}_\theta) = 100$, a value which is frequently used as an indicator of severe multicollinearity in linear models, the value of the bound is .98. It implies that the the second factor in (4.21), measuring the parameter interdependence effect on the asymptotic variance, may be as large as 25.

I should make it clear from the outset that the Information matrix approach to identification is for local analysis only. In general, global identification analysis for models that are non-linear in the parameters is not feasible.[12] In Chapter 2 I derived conditions for global identification of the *structural* parameters in linearized DSGE models, i.e. parameters in which the structural equations are linear.[13] However, the goal in the empirical DSGE research is usually to estimate the deep parameters, for which identification can be analyzed only locally.

4.2.5 Identification analysis procedure

In the previous section I outlined how the information matrix $\mathcal{I}_\theta(\theta, X)$ can be evaluated. By computing the rank of $\mathcal{I}_\theta(\theta, X)$ I can determine whether the value of θ, at which the matrix is evaluated, is identifiable or not. The model as a whole is identified if all points from the parameter space Θ are identifiable. It is clearly not feasible to verify that the rank condition is satisfied at all points in Θ, and instead I will check many randomly drawn points from Θ.[14] The following steps outline the

[12] See Rothenberg (1971) for more details.

[13] I distinguish between *deep* and *structural* parameters. For instance, if one of the equations in the linearized DSGE model is the New Keynesian Phillips curve

$$\pi_t = \frac{\beta}{1+\varpi\beta}E_t\pi_{t+1} + \frac{(\psi+\nu)(1-\zeta\beta)(1-\zeta)}{(1+\varpi\beta)\zeta}y_t + \frac{\varpi}{1+\varpi\beta}\pi_{t-1} + e_t$$

I call $\beta,\ \varpi,\ \psi,\ \nu$ and ζ deep parameters, and $\gamma_1 = \frac{\beta}{1+\varpi\beta}$, $\gamma_2 = \frac{(\psi+\nu)(1-\zeta\beta)(1-\zeta)}{(1+\varpi\beta)\zeta}$ and $\gamma_3 = \frac{\varpi}{1+\varpi\beta}$ - structural parameters.

[14] Boswijk and Doornik (2003) suggest this approach for checking identification of cointegration relationships.

identification analysis procedure I propose:

1. Draw randomly a point θ^j from Θ.
2. Check whether the reduced-form solution of the linearized structural model exists and is unique. If both of these conditions are not satisfied, go back to (1).
3. Compute the rank and the condition number of $H(\theta)'H(\theta)$. If it is of less then full rank, go back to (1).
4. Compute the rank and the condition number of $\mathcal{I}_\theta$, $\tilde{D}$ and $\tilde{\mathcal{I}}_\theta$.

In Step (1) I take one a priori admissible value of θ, which I then treat as the true parameter value in steps (2) to (4). Upon completion of the procedure, I will know if that value of θ is identifiable, and how strong identification is. Step (2) is necessary to ensure that there exists an unique likelihood function at θ^j. Conditions for existence and uniqueness of the solution can be found in Sims (2002), and are automatically checked by most computer algorithms for solving linear rational expectations models. I call *admissible* the values of θ for which these conditions are satisfied. In Step (3) I check the necessary condition for identification. Finding that $H(\theta)H(\theta)'$ is rank deficient, or poorly conditioned at θ^j, tells me that this particular point of the parameter space is either not identifiable, or is weakly identifiable for structural reasons, i.e. irrespectively of the data. To complete step (4) I need to evaluate $\mathcal{I}_\tau(\theta, X)$, which depends on the data as well as on θ^j. Therefore I need to first generate data X, assuming that θ^j is the true parameter value. To account for sampling variability, in practice I generate many replicas of X, and compute the reduced form Information matrix as the average Information matrix. From the rank of $\mathcal{I}_\theta(\theta, X)$, and conditioning of $\tilde{D}$ and $\tilde{\mathcal{I}}_\theta(\theta, X)$ I then determine whether θ^j is identified or not, and whether identification, from both the model and the data, is strong or weak. Based on the analysis in section 4, the value of 100 for the condition number of $\tilde{\mathcal{I}}_\theta$ will be used as an indicator of severe parameter interdependence problem, and therefore weak identification of at least some parameters in θ.

4.3 Case Study: Identification

4.3.1 The Smets Wouters (2007) model

The model in Smets and Wouters (2007) (see also Christiano, Eichenbaum, and Evans (2005)) is an extension of the standard RBC model featuring a number of nominal frictions, such as price and wage stickiness, and real rigidities - habit formation in consumption, investment adjustment cost, monopolistic competition, and variable cost of adjusting capital utilization. In addition, it contains a large number of serially correlated structural shocks. In this section I present a brief outline of the main components of the model. For details see the appendix accompanying Smets and Wouters (2007).

Households

There is a continuum of households indexed by j, each having the following utility function

$$E_t\left[\sum_{s=0}^{\infty}\beta^s\frac{1}{1-\sigma_C}\left((C_{t+s}(j)-\lambda C_{t+s-1}(j))^{1-\sigma_C}\right)\exp\left(\frac{\sigma_c-1}{1+\sigma_l}L_{t+s}(j)^{1+\sigma_l}\right)\right], \quad (4.22)$$

where $C_{t+s}(j)$ is consumption, $L_{t+s}(j)$ is hours worked; λ is an external habit persistence parameter.

Each household supplies differentiated labor services monopolistically to a continuum of labor markets charging nominal wage denoted with $W_t(j)$; W_t is an index of the nominal wage in the economy.

Households supply homogeneous labor to labor unions (indexed by l), who then sell it to labor packers. Labor services are differentiated by a union, who therefore have market power. Wage setting by unions (as well as price setting by firms discussed below) is subject to nominal rigidities with a Calvo mechanism whereby each

period a union can set the nominal wage to the optimal level with constant probability equal to $1-\xi_w$. Unions that cannot adjust their nominal wage optimally, change it according to the following indexation rule

$$W_{t+s}(l) = \gamma W_{t-1}(l)\pi_{t-1}^{\iota_w}\pi_*^{(1-\iota_w)}, \tag{4.23}$$

where γ is the deterministic growth rate, ι_w measures the degree of wage indexation to past inflation, and π_* is the steady state rate of inflation.

Labor packers buy differentiated labor services $L_t(l)$ from unions, package and sell composite labor L_t, defined implicitly by

$$\int_0^1 \mathcal{H}\left(\frac{L_t(l)}{L_t}; \lambda_{w,t}\right) dl = 1, \tag{4.24}$$

to the intermediate good sector firms. The function $\mathcal{H}$ is increasing, concave, and satisfies $\mathcal{H}(1) = 1$; $\lambda_{w,t}$ is a stochastic exogenous process changing the elasticity of demand, and the wage markup over the marginal disutility from work.

In addition to supplying labor at wage W_t, households rent capital to the firms producing intermediate goods, and earn rent at rate $R_t^K(j)$. Households accumulate physical capital according to the following law of motion:

$$\bar{K}_t(j) = (1-\delta)\bar{K}_{t-1}(j) + \varepsilon_t^i\left[1 - \mathcal{S}\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right] I_t(j), \tag{4.25}$$

where δ is the rate of depreciation, I_t is gross investment, and the investment adjustment cost function $\mathcal{S}$ satisfies $\mathcal{S}' > 0, \quad \mathcal{S}'' > 0$, and in steady state $\mathcal{S} = 0, \quad \mathcal{S}' = 0$. ε_t^I represents the current state of technology for producing capital, and is interpreted as investment-specific technological progress (Greenwood, Hercowitz, and Krusell (2000)).

Households control the utilization rate $Z_t(j)$ of the physical capital they own,

and pay $P_t a(Z_t(j))\bar{K}_{t-1}(j)$ in terms of consumption good when the capital intensity is $Z_t(j)$. The income from renting capital to firms is $R_t^k K_t(j)$, where $K_t(j) = Z_t(j)\bar{K}_{t-1}(j)$ is the flow of capital services provided by the existing stock of physical capital $\bar{K}_{t-1}(j)$. The utility function (4.22) is maximized with respect to consumption, hours, investment, and capital utilization, subject to the capital accumulation equation (4.25), and the following the per-period budget constraint

$$C_{t+s}(j) + I_{t+s}(j) + \frac{B_{t+s}(j)}{\varepsilon_{t+s}^b R_{t+s} P_{t+s}} - T_{t+s} = \frac{W_{t+s}(j)}{P_{t+s}} L_{t+s}(j) + \left(\frac{R_{t+s}^k Z_{t+s}(j)}{P_{t+s}} - a(Z_{t+s}(j)) \right) \bar{K}_{t+s-1}(j) + \frac{B_{t+s-1}(j)}{P_{t+s}} + \frac{\Pi_{t+s}(j)}{P_{t+s}}, \tag{4.26}$$

where B_{t+s} is a one-period nominal bond expressed on a discount basis. ε_t^b is an exogenous premium on the bond return, T_{t+s} is lump-sum taxes or subsidies, and Π_{t+s} is profit distributed by the labor union.

Firms

A perfectly competitive sector produces a single final good used for consumption and investment. The final good is produced from intermediate inputs $Y_t(i)$ using technology defined implicitly by

$$\int_0^1 \mathcal{G}\left(\frac{Y_t(i)}{Y_t}; \lambda_{p,t} \right) di = 1, \tag{4.27}$$

where $\mathcal{G}$ is increasing, concave, and $\mathcal{G}(1) = 1$; $\lambda_{p,t}$ is an exogenous stochastic process affecting the elasticity of substitution between different intermediate goods, also corresponding to markup over marginal cost for intermediate good firms.

Firms maximize profits given by

$$P_t Y_t - \int_0^1 P_t(i) Y_t(i) di, \tag{4.28}$$

where $P_t(i)$ is the price of intermediate good $Y_t(i)$.

Intermediate goods are produced in a monopolistically competitive sector. Each variety i is produced by a single firm using the technology

$$Y_t(i) = \varepsilon_t^a K_t(i)^\alpha (\gamma^t L_t(i))^{1-\alpha} - \Phi\gamma^t, \tag{4.29}$$

where Φ is a fixed cost, ε_t^a denotes total factor productivity, and γ is the deterministic growth rate of labor productivity.

As with wages, every period only a fraction $1 - \xi_P$ of intermediate firms can set optimally the price of the good they produce. The remaining ξ_p firms index their prices to past inflation according to

$$P_t(t) = \gamma P_{t-1}(i)\pi_{t-1}^{\iota_p}\pi_*^{(1-\iota_p)}, \tag{4.30}$$

where ι_p measures the degree of price indexation to past inflation.

The Government

The government's budget constraint is simply

$$P_t G_t + B_{t-1} = T_t + \frac{B_t}{R_t}, \tag{4.31}$$

where G_t is government consumption in terms of final good.

The central bank sets the nominal interest rate according to the following rule

$$\frac{R_t}{R^*} = r_t \left(\frac{R_{t-1}}{R^*}\right)^\rho \left[\left(\frac{\pi_t}{\pi^*}\right)^{r_\pi} \left(\frac{Y_t}{Y_t^*}\right)^{r_y}\right]^{1-\rho} \left(\frac{Y_t/Y_{t-1}}{Y_t^*/Y_{t-1}^*}\right)^{r_{\triangle y}} \tag{4.32}$$

where R^* is the steady state level of the gross nominal interest rate, r_t is a monetary policy shock, and Y^* is potential output, defined as the output in a flexible price and

wage economy.

Shocks

There are seven exogenous shocks in the model. Five of the shocks - the risk premium, TFP, investment-specific technology, government purchases, and monetary policy - follow AR(1) processes

$$\ln \varepsilon_t^b = \rho_b \ln \varepsilon_{t-1}^b + \eta_t^b \tag{4.33}$$

$$\ln \varepsilon_t^a = \rho_a \ln \varepsilon_{t-1}^a + \eta_t^a \tag{4.34}$$

$$\ln \varepsilon_t^i = \rho_i \ln \varepsilon_{t-1}^i + \eta_t^i \tag{4.35}$$

$$\ln \varepsilon_t^g = \rho_g \ln \varepsilon_{t-1}^g + \rho_{ga} \eta_t^a + \eta_t^g \tag{4.36}$$

$$\ln \varepsilon_t^r = \rho_r \ln \varepsilon_{t-1}^r + \eta_t^r \tag{4.37}$$

The remaining two shocks - wage and price markup shocks - follow ARMA$(1,1)$ processes

$$\ln \lambda_{w,t} = (1-\rho_w) \ln \lambda_w + \rho_w \ln \lambda_{w,t-1} + \eta_t^w + \mu_w \eta_{t-1}^w \tag{4.38}$$

$$\ln \lambda_{p,t} = (1-\rho_p) \ln \lambda_p + \rho_p \ln \lambda_{p,t-1} + \eta_t^p + \mu_p \eta_{t-1}^p \tag{4.39}$$

Model Solution

The economy in the model is assumed to evolve along a deterministic growth path, with γ being the gross rate of growth. To solve the model, I first detrend all growing variables - consumption, investment, capital, real wages, output and government spending, and then all equilibrium conditions are log-linearized around the deterministic steady state of the detrended variables. A detailed discussion of all log-linear equations can be found in Smets and Wouters (2007)

The linearized version of the model can be written as in 4.3 with Z_t being a 33×1 vector given by $Z_t = \left[Z_t^{'f}, Z_t^{'s} \right]'$, where Z_t^f and Z_t^s are defined as

$$Z_t^f = \left[c_t^f, l_t^f, w_t^f, q_t^f, i_t^f, r_t^{kf}, r_t^f, k_t^f, \bar{k}_{t-1}^f, y_t^f, z_t^f \right]'$$

and

$$Z_t^s = \left[c_t^s, l_t^s, \pi_t, w_t^s, q_t^s, i_t^s, r_t^{ks}, r_t^s, k_t^s, \bar{k}_{t-1}^s, y_t^s, z_t^s, \mathrm{mc}_t, \varepsilon_t^b, \varepsilon_t^i, \varepsilon_t^a, \varepsilon_t^g, \varepsilon_t^p, \varepsilon_t^w, \varepsilon_t^r, \eta_t^p, \eta_t^w \right]'$$

Here I use small letters to represent the percent deviation of the variables from their steady state levels[15]. Z^f is a vector collecting the variables in the flexible price and wage version of the economy, and Z^s collects the variables from the sticky price and wage economy. U_t is a vector of the seven structural shocks:

$$U_t = [\eta_t^a, \eta_t^b, \eta_t^I, \eta_t^w, \eta_t^p, \eta_t^g, \eta_t^r]'$$

The coefficient matrices $\Gamma_0,\ \Gamma_1,\ \Gamma_2$ and Γ_3 in the canonical form 4.3 are functions of a 39×1 vector of deep parameters θ, defined by

$$\begin{aligned} \theta = \quad & [\delta, \lambda_w, g_y, \varepsilon_p, \varepsilon_w, \rho_{ga}, \beta, \mu_w, \mu_p, \alpha, \psi, \varphi, \sigma_c, \lambda, \Phi, \iota_w, \xi_w, \iota_p, \xi_p, \sigma_l, r_\pi, r_{\triangle y}, r_y, \rho, \rho_a, \rho_b, \\ & \rho_g, \rho_I, \rho_r, \rho_p, \rho_w, \gamma, \sigma_a, \sigma_b, \sigma_g, \sigma_I, \sigma_r, \sigma_p, \sigma_w]' \end{aligned} \tag{4.40}$$

As in Smets and Wouters (2007), I assume that the only observed variables are consumption, investment, output, wages, hours, inflation, and the nominal interest rate. Thus X_t is given by

$$X_t = \begin{bmatrix} c_t & l_t & \pi_t & w_t & i_t & r_t & y_t \end{bmatrix}$$

[15] q denotes the percent deviation of real value of capital from the steady state level of one.

and the remaining $39 - 7 = 32$ variables in Z are treated as latent. Finally, matrix C is the measurement equation (4.5) is a 7×32 matrix constructed from the rows of 32×32 identity matrix.

4.3.2 Identification of the Smets Wouters (2007) model

Now I apply the procedure from section 4 to the model described above. I take the parameter space Θ to be the one defined by the prior distribution of θ, as specified in Smets and Wouters (2007). A summary of that distribution is provided in Table B.1 of Appendix B. This prior distribution is very common in the recent studies using Bayesian methods to estimate similar New Keynesian DSGE models. An alternative approach would be to treat all a priori admissible parameter values as equally likely, that is, to assume uniform priors. The benefit of our approach is that it provides a better coverage of the parts of the space that are considered in the literature as more plausible. For instance, the discount factor β could, theoretically, lie anywhere between 0 and 1. However, values close to .99 are considered to be much more likely than values close to 0. This type of considerations are reflected by the choice of shape and parameters of the prior distribution.

In their estimation procedure Smets and Wouters (2007) treat five deep parameters as known. These are: discount rate δ, share of government spending in GDP g_y, steady state markup in the labor market λ_w, and the two curvature parameters of the aggregation functions in the labor and final good sectors - ϵ_p and ϵ_w. [16] For the first two parameters the reason is that they are difficult to estimate with the data used in estimation. The markup and the two curvature parameters, on the other hand,

[16] These parameters are defined as $\epsilon_p = \frac{\partial \ln(\kappa_p(1))}{\partial \ln(\tilde{P})}$, $\epsilon_w = \frac{\partial \ln(\kappa_w(1))}{\partial \ln(\tilde{W})}$, where $\kappa_p(x) = -\frac{\mathcal{G}'(x)}{x\mathcal{G}''(x)}$, $\kappa_w(y) = -\frac{\mathcal{H}'(y)}{y\mathcal{H}''(y)}$ are elasticities of demand for goods and labor services, and $\tilde{P}$ and $\tilde{W}$ are the relative price and wage. They measure the percent change in the elasticity of demand for goods and labor due to one percent change in the relative price/wage, evaluated in steady state. In the simple case, where the aggregator functions $\mathcal{H}$ and $\mathcal{G}$ have the Dixit-Stiglitz functional form, both parameters are equal to zero (see Eichenbaum and Fisher (2007))

are asserted to be unidentifiable. The second claim is easier to check, so I examine it first.

The easiest way to detect lack of identification of one or more deep parameters is to examine matrix $H(\theta) = \frac{\partial \tau(\theta)}{\partial \theta'}$. It must have full column rank for θ to be identified. Moreover, if a parameter is generally unidentifiable, it would not matter at what admissible value of θ I compute $H(\theta)$, as it will be with reduced rank for any $\theta \in \Theta$. In what follows I use the posterior mode of θ reported in Smets and Wouters (2007). When θ includes all 39 parameters listed in (4.40), the rank of $H(\theta)$ is 36. One of these parameters, however, is the trend growth rate γ for which there is additional information in the trending observed variables that I have not taken into account. Treating γ as known, and computing $H(\theta)$ for the remaining 38 deep parameters, I conclude that two of them are not identifiable. Closer inspection of $H(\theta)$ (see section 4 for more details) shows us that ϵ_p and ϵ_w are indistinguishable from the Calvo probability parameters ξ_p and ξ_w. In other words, one can identify either ϵ_p or ξ_p but not both simultaneously, and similarly for ϵ_w and ξ_w.

The lack of separate identification of these parameters is due to the role they play in the model. A high value of ϵ_p, for instance, implies that the elasticity of demand increases rapidly when a firm's relative price increases. This implies that it is optimal for the firm to increase its price by a smaller amount, compared to a case when ϵ_p is low. As a result prices are adjusted less rapidly. The same outcome is observed when ξ_p - the probability of a firm not being able to adjust its price to the optimal level, is large. Similar relationship exist between the wage parameters ϵ_w and ξ_w.[17] No such problem was detected regarding λ_w, and when I compute $H(\theta)$ after ϵ_p and ϵ_w are removed from θ, it has full rank. I conclude, therefore, that there is nothing in the model that makes the wage markup parameter λ_w unidentified. Computing

[17]Note, however, that although they play similar roles, these two pairs of parameters are not necessarily indistinguishable in the non-linear version of the model. Linearization in general make parameters harder to identify (see McManus (1992)).

the full information matrix $\mathcal{I}_\theta(\theta, X) = H(\theta)'\mathcal{I}_\tau(\theta, X)H(\theta)$ confirms that λ_w is indeed identified at the posterior mode of θ.

As I mentioned above, having γ among the parameters with respect to which $H(\theta)$ is computed causes additional identification problems. It may be useful to know what the source of these problems is, and whether it would be possible to estimate γ from the stationary version of the model using detrended data. To answer these questions I computed $H(\theta)$ for θ that includes γ, and sequentially exclude one of the remaining deep parameters. I find that $H(\theta)$ has reduced rank when δ, β, φ, λ and γ are all included, and is with full rank whenever one of these five parameters is excluded. This implies that γ can be identified, using detrended data only, if either δ, β, φ or λ is kept fixed instead of estimated. This is true, for instance, for the parametrization estimated in Smets and Wouters (2007), where it is assumed that δ is known.

I study the identifiability of the model for six parameterizations that differ in the parameters assumed to be known. The parameters are those assumed known in Smets and Wouters (2007) plus γ. The values of the fixed parameters, reported in Table 4.1 below, are also taken from that paper. The trend parameter γ is held fixed in all cases except parametrization 5. In parametrization 1 all other parameters are left free. In parameterizations 2 to 4 one of the other three parameters - δ, λ_w and g_y respectively, is also assumed known. Considering these cases allows us to compare the strength of these parameters' identifiability. In parametrization 5 all parameters except γ are fixed. Since δ is one of them, as I explained above, γ is identified from the stationary model. In parametrization 6 all parameters are assumed known and thus it is closest to the parametrization estimated in Smets and Wouters (2007).[18] The number of free parameters in θ, for each parametrization, is given in Column 6 of Table 4.1.

I draw $1,000,000$ points from Θ and perform steps (1) to (3) described in section

[18]The difference is that in Smets and Wouters (2007) γ is estimated using trending data, while in parametrization 6 γ is assumed known.

Table 4.1: Parameterizations

param.	δ	λ_w	g_y	γ	dim(θ)	cond(H)	cond($\mathfrak{I}_\theta$)
1	free	free	free	.431	36	6.0e2	1.8e7
2	.025	free	free	.431	35	5.8e2	3.7e6
3	free	1.5	free	.431	35	3.4e2	1.2e7
4	free	free	.18	.431	35	5.9e2	1.8e7
5	.025	1.5	.18	free	34	7.4e2	2.0e7
6	.025	1.5	.18	.431	33	3.1e2	1.9e6

Note: Column 6 shows the number of free parameters in θ. Columns 7 and 8 show the median condition numbers of the Jacobian $H = \frac{\partial \tau}{\partial \theta'}$ and the full Information matrix, respectively. Values
cond(X) $< 4.5e15 = 4.5 \times 10^{15}$ indicate that matrix X is full rank.

4 for each one of them. The distributions of the actual draws are shown in Figure B.1 in Appendix B). I sort the admissible draws and divide them into 10 groups; then I perform step (4) for 100 points from each group. Thus I compute the full information matrix $\mathcal{I}_\theta(\theta, X)$ for 1,000 admissible points from Θ. I did not evaluate that matrix for all admissible draws because with the routine I use for evaluation of $\mathcal{I}_\tau(\theta, X)$, it takes very long to compute that matrix.

Between 96% and 98% of the draws were admissible (see table B.2 in Appendix B). There was no stable solution for about .1% to .3% of them, and for about 2% to 4% there were multiple solutions.

Matrix $H(\theta)$ had a full column rank for all of the admissible draws. Thus the necessary condition for identification was satisfied everywhere in the parameter space. Columns 7 Table 4.1 show, for each of the six parameterizations, the median condition number of the Jacobian $H(\theta)$. The other deciles of the distribution of the condition numbers of $H(\theta)$ are reported in Table B.3 of Appendix B.

The significance of the quite large condition numbers of $H(\theta)$ is twofold. First,

being the part of the full Information matrix that depends only of the model, a poorly conditioned H makes $\mathcal{I}_\theta$ close to singular, even when the date is very informative, i.e. $\mathcal{I}_\tau$ is very well conditioned. Second, $H(\theta)$ gives the sensitivity of the likelihood function to small perturbations in θ. A high condition number of $H(\theta)$, therefore, implies that the likelihood responds very slightly to relatively large perturbations in some of the components of θ. Both implication of a poorly conditioned $H(\theta)$ in turn indicate that θ is not well identified for model-related reasons.

The information matrix $\mathcal{I}_\theta(\theta, X)$ failed to be of full rank for about .8% of the 1000 draws for which it was evaluated. Columns 8 Table 4.1 show, for each of the six parameterizations, the median condition number of the full Information matrix for θ. Table B.5 in Appendix B shows the ten deciles of the distribution of the condition numbers of $\mathcal{I}_\theta(\theta, X)$. We see that even though it has a full rank for almost all of the draws, its condition numbers is extremely high, which implies that the matrix is poorly conditioned virtually everywhere in the parameter space.

Table 4.2: Cross-correlations

	λ_w	g_y	μ_p	φ	σ_c	h	Φ	ι_w	ξ_w	ι_p	σ_l	$r_{\triangle y}$	r_y
β	.42	**.98**	-.07	.28	.44	-.40	-.85	-.24	-.41	.26	-.36	.42	-.30
φ	**.95**	.26	**-.92**	1	**.90**	-.78	-.24	-.87	**-.95**	**.90**	-.73	**.96**	**-.98**
σ_c	**.99**	.40	-.78	**.90**	1	**-.96**	-.54	**-.91**	**-.99**	.84	-.88	.86	-.88
h	**-.91**	-.36	.69	-.78	**-.96**	1	.56	**.90**	**.92**	-.67	**.95**	-.72	.78
ι_w	**-.90**	-.21	.75	-.87	**-.91**	**.90**	.29	1	**.90**	-.71	.88	-.79	.87
ξ_w	**-.99**	-.38	.84	**-.95**	**-.99**	**.92**	.49	**.90**	1	-.89	.83	**-.91**	**.92**
ι_p	**.90**	.22	-.76	**.90**	.84	-.67	-.33	-.71	-.89	1	-.51	.89	-.81
ξ_p	.52	.89	-.11	.30	.58	-.61	**-.98**	-.34	-.52	.31	-.53	.38	-.31
σ_l	-.82	-.30	.69	-.73	-.88	**.95**	.45	.88	.83	-.51	1	-.66	.75
$r_{\triangle y}$	**.92**	.37	-.84	**.96**	.86	-.72	-.32	-.79	**-.91**	.89	-.66	1	**-.93**
r_y	**-.92**	-.29	.93	**-.98**	-.88	.78	.22	.87	**.92**	-.81	.75	**-.93**	1
ρ	.89	.32	-.74	.88	.84	-.68	-.40	-.69	-.88	**.95**	-.57	**.91**	-.78
ρ_I	.81	.50	-.56	.65	.87	**-.94**	-.68	-.79	-.82	.51	**-.93**	.61	-.68
σ_I	**-.96**	-.45	.81	**-.91**	**-.97**	**.93**	.50	**.91**	**.97**	-.76	**.91**	-.88	**.92**
σ_p	-.81	-.04	**.99**	**-.91**	-.76	.67	0	.74	.82	-.74	.67	-.83	**.92**

Note: Pairwise correlation coefficients $corr(\hat{\theta}_i, \hat{\theta}_j)$ exceeding .95 in absolute value. The values are obtained by inverting and normalizing the information matrix evaluated at θ for which the condition number of the matrix is equal to the median value from Table B.3. High correlation between the estimates of two deep parameters indicates that they are difficult to identify.

The poor conditioning of the information matrix suggests that some of its columns are nearly linearly dependent. Since the information matrix is equal to the inverse of the asymptotic covariance matrix for the estimate of θ, this in turn implies that there exists a strong degree of interdependence among the estimates of some of the deep parameters. This creates identification problems as these parameters' separate effects on the likelihood are difficult to isolate.[19] I can measure the degree of linear dependence by computing the correlations between the columns of the covariance matrix. The complete set of pairwise correlation coefficients may be obtained by inverting and normalizing the information matrix.[20] Table 4.2 shows all pairs of parameters whose estimates have correlation exceeding .95 in absolute value. The correlation coefficients were computed at the value of θ where the condition number of $\mathcal{I}_\theta(\theta, X)$ equals the median of all points at which the information matrix was evaluated. We see, for instance, that the estimate of the wage markup parameter λ_w is extremely highly correlated with ξ_w and σ_c. This partially confirms the claim in Smets and Wouters (2007) that this parameter is difficult to identify in their model, although, as I discussed above, they are mistaken in asserting that λ_w is not identified. Other parameters that would be very difficult to identify at this particular value of θ are σ_c, ξ_w, h and σ_l as well as the policy rule coefficients ρ, ρ_y and $\rho_{\triangle y}$. Although these observations are made on the basis of single point θ from the parameter space, they remained valid for many other parameter values I tried. In addition, as can be seen from Table B.6, very high degree of linear dependence can also be found for other pairs of parameters, such as σ_w, ξ_w, h and λ_2, or r_π, ρ, ρ_I and σ_I. The correlation coefficients reported in Table B.6 were computed at θ equal to the value where the

[19]This is easy to see for the linear regression model $y = X\beta + \epsilon$. When two of the regressors, X_i and X_j are nearly collinear, the corresponding coefficients, β_i and β_j will be difficult to identify. Also, since the covariance matrix of the estimate $\hat{\beta}$ is proportionate to $\mathbb{E}X'X$, high collinearity between the regressors implies high correlation between the corresponding elements of $\hat{\beta}$.

[20]That is, I divide each i, j covariance term of the matrix by the product of the standard deviations of variables i and j. Neely, Roy, and Whiteman (2001) also use the correlation matrix of the parameter estimates to determine the sources of identification problems

condition number of $\mathcal{I}_\theta(\theta, X)$ equals the 7-th percentile of all points at which the information matrix was evaluated. Since the condition number of matrix is higher - 6.4×10^8 vs. 1.8×10^7, the linear dependencies shown in Table B.6 are substantially stronger than those reported in Table 4.2.

I draw the following three conclusions from this exercise. First, although the necessary and sufficient condition for identification is generally satisfied, the conditioning of the information matrix is very poor, indicating that θ is very weakly identified in most of the parameter space. Second, the reasons for weak identification are mainly in $H(\theta)$, which is entirely determined by the structure of the model, and not affected by the data. To see that, remember the relationship between the information matrix $\mathcal{I}_\theta(\theta, X)$ and $H(\theta)$ (see equation (4.6)). Even when $\mathcal{I}_\tau(\theta, X)$ is very well conditioned, poor conditioning of $H(\theta)$ will result in poorly conditioned $\mathcal{I}_\theta(\theta, X)$. For instance, suppose that there is very small amount of uncertainty in the estimate of τ, and $\mathcal{I}_\tau(\theta, X)$ has a condition number equal to one. In particular, I let $\mathcal{I}_\tau(\theta, X) = \mathcal{I}_\tau^*$ be a diagonal matrix whose inverse - the covariance matrix for τ, has non-zero elements equal to 1% of the true values of τ. The deciles of the distribution of the condition numbers of $\mathcal{I}_\theta = H'\mathcal{I}_\tau^* H$ are shown in table B.4. If, for instance, the condition number of $H(\theta)$ is $6e2$ - the median for parametrization 1, I find that the condition number of $\mathcal{I}_\theta(\theta, X)$ is about $3.7e5$. Thus, even though $\mathcal{I}_\theta(\theta, X)$ was computed in relatively small number of points from Θ, our findings regarding $H(\theta)$ suggest that the identification of θ is generally weak. Third, the strength of identification improves only a little when δ, λ_w and g_y are kept fixed. We see that by comparing the conditioning of $H(\theta)$ and $\mathcal{I}_\theta(\theta, X)$ for parametrization 1 and 6. The difference is relatively small. Moreover, the improvement seen in parametrization 6 is, at least partly, due to the smaller number of free parameters, and not only because the identifiability of the fixed parameters is much weaker. Of these three parameters, g_y appears to be be the worst identified one. This can be deduced by comparing the conditioning of parametrization 4 with

that of parameterizations 2 and 3.

4.3.3 Discussion

The analysis of parameter identification in the Smets and Wouters model suggests that weak identifiability, and not complete failure of identification, is likely to be the more serious problem for DSGE models in general. Even when some parameters are not identifiable, as is the case with ε_p, ξ_p, ε_w, and ξ_w in the model I consider here, this is easy to detect - by computing the rank of Jacobian matrix H, and straightforward to deal with - by fixing instead of estimating parameters that lack identification.

In the previous section I used the condition number of the Information matrix to measure the strength of identification. As I discussed in section 4 , the condition number of a matrix measures the distance from singularity of the matrix. In the econometrics literature Forchini and Hillier (2005) also propose the condition number of the information matrix as a measure of the strength of identification in parametric models, and show that it is closely related to the concentration parameter, suggested by Stock, Wright, and Yogo (2002) as a measure for the strength of identification in linear models.

A well known property of the condition number, usually emphasized in the context of linear regression models $y_t = X_t'\theta + \varepsilon_t$, is that it measures the sensitivity of the estimator of θ to errors in the estimates of $\mathbb{E}X_tX_t'$ and $\mathbb{E}X_ty_t$. Since the true parameter value θ_0 solves the population equation

$$\mathbb{E}X_ty_t = \mathbb{E}X_tX_t'\theta$$

and with finite data $\mathbb{E}X_tX_t'$ and $\mathbb{E}X_ty_t$ are estimated with error, the estimate $\hat{\theta}$ differs from the true θ_0. In this context the identification of θ is poor if small errors $\triangle\mathbb{E}X_tX_t'$ and $\triangle\mathbb{E}X_ty_t'$ in the estimates of $\mathbb{E}X_tX_t'$ and $\mathbb{E}X_ty_t$ lead to a large error $\triangle\theta = \hat{\theta} - \theta_0$.

This sensitivity is quantified by $\text{cond}(\mathbb{E}X_tX_t')$ from[21]

$$\max_{\triangle\mathbb{E}X_tX_t',\ \triangle\mathbb{E}X_ty_t'} \frac{\|\hat{\theta}-\theta_0\|}{\|\theta_0\|} = \text{cond}(\mathbb{E}X_tX_t')\left(\frac{\|\triangle\mathbb{E}X_tX_t'\|}{\|\mathbb{E}X_tX_t'\|}+\frac{\|\triangle\mathbb{E}X_ty_t'\|}{\|\mathbb{E}X_ty_t'\|}\right) \tag{4.41}$$

That is, the larger is the condition number of $\mathbb{E}X_tX_t'$, the more sensitive is $\hat{\theta}$ to errors in the estimates of $\mathbb{E}X_tX_t'$ and $\mathbb{E}X_ty_t$.

Another useful property of the condition number is that it measures the sensitivity of the estimate of the inverse of a matrix to errors in the estimate of the matrix itself. Specifically,

$$\max_{\triangle A} \frac{\|(A+\triangle A)^{-1}-A^{-1}\|}{\|A^{-1}\|} = \text{cond}(A)\left(\frac{\|\triangle A\|}{\|A\|}\right) \tag{4.42}$$

Therefore, if a matrix is poorly conditioned, small errors in the estimate of the matrix lead to large errors in the estimate of its inverse.

For correctly specified parametric models the asymptotic covariance matrix of the estimators is equal to the inverse of the Information matrix. The implication of (4.42) is that when $\text{cond}(\mathcal{I}_\theta)$ is large, even small errors in the estimate $\hat{\mathcal{I}}_\theta$ of $\mathcal{I}_\theta$, may cause large errors in the estimate of the covariance matrix $\mathbf{V}_\theta$. In particular, the standard errors for $\hat{\theta}$ - the diagonal elements of $\mathbf{V}_\theta$, could be very imprecisely estimated when $\text{cond}(\mathcal{I}_\theta)$ is large. To see how large these errors could be in Smets and Wouters (2007) model, I carried out the following Monte Carlo simulation exercise. For each of the ten deciles shown in Table B.5, I assumed that the corresponding matrix $\mathcal{I}_\theta$ is the true information matrix. I then added small errors to the diagonal elements of $\mathcal{I}_\theta$, drawn from standard normal distribution with variance equal to 1% of the true value. The resulting matrix $\tilde{\mathcal{I}}_\theta$ is then inverted and the percentage error in the diagonal elements of $\tilde{\mathbf{V}}_\theta$ recorded. Table B.7 in Appendix B shows the results from 1000 repetitions.

[21] see Horn and Johnson (1985) for a proof

The reported numbers are percent error in the standard errors of θ for 1 percent error in the corresponding diagonal element of $\mathcal{I}_\theta$. The results demonstrate that the estimated covariance matrix is very sensitive to even small errors in the estimate of the information matrix, and the higher the condition number of $\mathcal{I}_\theta$ is, the larger the errors in the estimate of $\mathbf{V}_\theta$ tend to be. This shows us that the standard errors obtained by inverting the information matrix are practically meaningless.

In addition to the implications for the validity of the estimated covariance matrix and confidence intervals, the conditioning of the information matrix also affects the speed with which the estimator $\hat{\theta}_t$ converges to θ_0 as t increases. To see that, consider again the linear regression model $y_t = X_t'\theta + \varepsilon_t$. It can be shown that the optimal estimate with $T+1$ observations is

$$\hat{\theta}_{T+1} = \hat{\theta}_T + \left[\hat{E}_{T+1} X_t X_t'\right]^{-1} X_{T+1} \left[y_{T+1} - X_{T+1}'\hat{\theta}_T\right] \tag{4.43}$$

where $\hat{\theta}_T$ and $\hat{E}_T X_t X_t'$ are the estimates of θ and information matrix EX_tX_t' with T observations. From (4.43) it is clear that the rate with which $\hat{\theta}_T$ converges to θ_0 depends on the convergence of $(\hat{E}_T X_t X_t')^{-1}$ to $(EX_tX_t')^{-1}$. However, from (4.42) it can be deduced that when $\text{cond}(EX_tX_t')$ is large, the convergence of $(\hat{E}_T X_t X_t')^{-1}$ will be slow.

4.3.4 Why is identification weak?

The results so far suggests that the model as a whole is poorly identified. Moreover, what I find regarding the conditioning of $H(\theta)$ indicate that the cause for this is in the structure of the (linearized) model. This is because poor conditioning of $H(\theta)$ translates into poor conditioning of the information matrix, and consequently, weak identification of θ. Specifically, note that applying the chain rule for differentiation I

can express the derivative of the log-likelihood with respect to θ_i as

$$\frac{\partial l}{\partial \theta_i} = \frac{\partial l}{\partial \tau'} \frac{\partial \tau}{\partial \theta_i} \tag{4.44}$$

Then it follows that

$$\text{if} \quad \frac{\partial \tau}{\partial \theta_i} \approx \mathbf{0} \quad \Rightarrow \quad \frac{\partial l}{\partial \theta_i} \approx 0, \quad \text{and} \tag{4.45}$$

$$\text{if} \quad \frac{\partial \tau}{\partial \theta_i} \approx \sum_{j \neq i} a_j \frac{\partial \tau}{\partial \theta_j} \quad \Rightarrow \quad \frac{\partial l}{\partial \theta_i} \approx \sum_{j \neq i} a_j \frac{\partial l}{\partial \theta_j} \tag{4.46}$$

In words, if θ_i is poorly identified in the model, it will be poorly identified when the model is taken to the data. Poor identification in the model results either from the reduced-form parameters τ being almost insensitive to θ_i, or from the effect of θ_i on τ being well approximated by that of a linear combination of other deep parameters. When τ is completely insensitive to θ_i, or the effect of θ_i can be exactly replicated by that of other elements of the vector θ, θ_i will be unidentifiable in the model, and, therefore, from the data.

The condition number of H is a simple overall indicator of the existence of problems of that nature. Both the low sensitivity of reduced-form parameters τ to a deep parameter, and the strong interdependence among the effect of multiple deep parameters in the model, will result in at least one very small singular value of H. According to the condition number, a singular value is small when it is much smaller relative to the largest singular value. Like the rank of a matrix, which indicates whether or not there is an exact linear dependence among the columns, a large condition number indicates only that there is at least one near linear dependence among the columns of the matrix.

More information about the properties of H may be obtained by considering all singular values of the matrix, instead of only the smallest and the largest ones. Figure 4.1 shows the singular values of $H(\theta)$ evaluated at the posterior mean in Smets

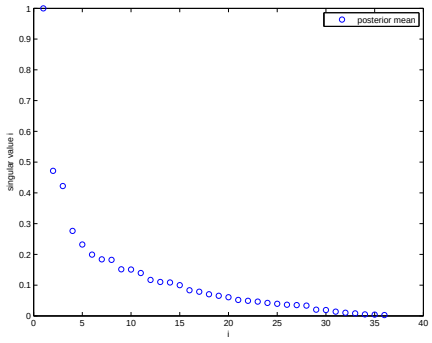

Figure 4.1: Singular values of $H(\theta)$ at the posterior mean reported in Smets and Wouters (2007)

and Wouters (2007). On the figure each singular value is divided by the largest one. We see that the last six singular values are all very close to zero, and to each other. Therefore there are at least six independent combinations of deep parameters which are nearly linearly dependent. This conclusion can only be made with respect to the particular point of the parameter space where H was evaluated. However, similar plots of the singular values of H evaluated at different points in the parameter space lead to similar conclusions. Small singular values imply that the original matrix H may be very well approximated with a matrix of lower rank. For example, if rank of H is 36, but the the smallest six singular values are nearly zero, there exist a matrix $\check{H}(30)$ with rank 30, such that the distance $d(H, \check{H}(30)) = \frac{\|H-\check{H}(30)\|}{\|H\|}$ is small. Using the Frobenius norm[22], Figure 4.2 plots the distance $d(H, \check{H}(i))$ for $i = 1 : 36$. At

[22]The Frobenius norm for a non-square matrix $A_{m\times n} = \{a_{ij}\}$ is

$$\|A\|_F = \sqrt{\sum_i^m \sum_j^n a_{ij}^2}$$

$i = 30$ we have $d(H, \check{H}(i)) = .0041$. Again, we can conclude that a matrix with a rank of 30 provides an extremely close approximation of the rank 36 matrix H.

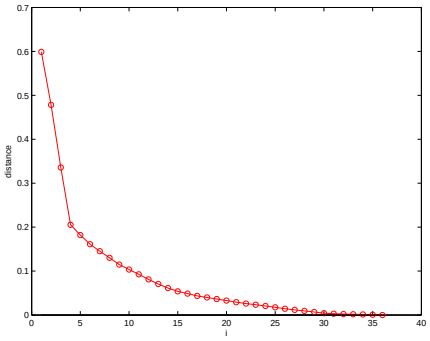

Figure 4.2: Distance $d(H, \check{H}(i)) = \frac{\|H - \check{H}(i)\|}{\|H\|}$ between H and best rank i approximation of H at the posterior mean reported in Smets and Wouters (2007)

From a modeler's perspective it important to know what parameters are involved in the near linear dependencies indicated by the condition number and the singular values plots. Each deep parameter represents some feature of the model, and it is useful to know what features are either unimportant or almost redundant, given the other features of the model. As with the full Information matrix, either one of the possible causes for problems with identification - low sensitivity and parameter interdependence, result in the poor conditioning of matrix $H(\theta) \equiv \{\frac{\partial \tau}{\partial \theta_i}\}_i$. Therefore I can determine which parameters are not well identified in the model and why, by studying the columns of $H(\theta)$ and the relationships among them. This will be my objective in the remaining of this section.

Sensitivity

The sensitivity of the reduced-form parameters with respect to the deep parameters can be measured by

$$s_i(\theta) = \sqrt{\frac{1}{J}\sum_{j}^{J}\left(\theta_i \frac{\partial \tau_j}{\partial \theta_i}\right)^2} \tag{4.47}$$

where J is the dimension of τ. This gives us the root mean squared change in τ due to a 1% change in the deep θ_i. Large value of $s_i(\theta)$ implies that θ_i plays an important role in the model, while smaller value means that θ_i is relatively less important.

Table 4.3 shows the results for the relative importance of the parameters in the model when θ is evaluated at the posterior mean value from Smets and Wouters (2007). The most important parameters, according to this measure, are the autocorrelation coefficients of the wage and price markup shocks, the wage and price stickiness parameters, and two of the policy rule parameters - smoothing coefficient and the coefficient of the response to inflation. Quite important are also the steady state wage markup and the habit persistence parameters. Among the least important parameters are discount factor, and all of the standard deviations of the structural shocks in the model. The dispersion of the sensitivity values is quite striking. On average, the five most important parameters are more than 120 times more important than the five least important parameters.

Table B.8 in the Appendix shows the values $s_i(\theta)$ computed at values of θ corresponding to the minimum and the 10 deciles of the distribution of cond$(H'H)$ based on the 1 million draws from Θ. There we see that the relative ranking of the parameters varies somewhat depending on the value of θ. For instance the habit persistence parameter or the elasticity of the investment adjustment cost function are sometimes among the most important parameters in the model. On the other hand the autocorrelation coefficient of the wage markup shock is frequently among the least important

parameters in the model. Nevertheless, the sets of the most and the least important parameters remain generally stable.

Parameter interdependence

The problem of parameter interdependence arises when different parameters play very similar role in the model. When this is true, a deep parameter will be poorly identified even if it is important in the model, in the sense of the reduced-form parameters being very sensitive to changes in that parameter. As a result the likelihood will be almost flat with respect to a linear combination of parameters, even if it is not flat with respect to the individual parameters.

Locally, the degree of parameter interdependence between θ_i and the other deep parameters can be measured as the angle between $\frac{\partial \tau}{\partial \theta_i}$ and its projection onto the space spanned by $\{\frac{\partial \tau}{\partial \theta_1}, \frac{\partial \tau}{\partial \theta_2}, ..., \frac{\partial \tau}{\partial \theta_{i-1}}, \frac{\partial \tau}{\partial \theta_{i+1}}, ..., \frac{\partial \tau}{\partial \theta_J}\}$. For ease of notation I will use H_i to denote $\frac{\partial \tau}{\partial \theta_i}$, and H_{-i} to denote its projection. The cosine of the angle between H_i and H_{-i} is given by

$$\cos(i, -i) = \frac{H_i' H_{-i}}{(H_i' H_i)^{1/2} (H_{-i}' H_{-i})^{1/2}} \tag{4.48}$$

I will call this angle the degree of *multiple collinearity*, and use it to measure how well the effect of θ_i in the model can be mimicked by the other deep parameters. Values close to -1 or 1 imply that there is a very strong collinearity problem for θ_i. Values close to 0 on the other hand suggest that the role θ_i plays in the model cannot be approximated well by other deep parameters.

I can similarly measure the degree of interdependence between any two deep parameters θ_i and θ_j as the angle between H_i and H_j. I will call this the degree of *pairwise collinearity* and use it to measure how closely related or substitutable are these two parameters in the model. For instance, the degree of pairwise collinearity

between the Calvo parameter for wages ξ_w and the elasticity parameter ϵ_w in the Smets and Wouters (2007) model is 1, as is that between the price parameters ξ_p and ϵ_p . Therefore these two pairs of parameters are completely substitutable and cannot be identified separately.

Table 4.4 shows the largest values of the pairwise collinearity measure for each of the deep parameters when θ is evaluated at the posterior mean in Smets and Wouters (2007). Columns 3 to 8 correspond to each of the six parameterizations I consider. The results suggest severe collinearity problems for the wage markup and wage stickiness parameters (λ_w and ξ_w) as well as for the policy rule parameters (ρ, $r_{\triangle y}$, r_π). The degree of pairwise collinearity for these parameters is .99 (.98 for $r_{\triangle y}$) which implies that, at least locally, the effect of changing one of these parameters can be almost completely offset by changing another deep parameter. Fixing λ_w does not resolve the problem since ξ_w remains highly collinear with elasticity of labor supply (σ_l). Other parameters with high degree of pairwise collinearity are: price stickiness parameter (ξ_p), policy response to output (r_y), elasticity of intertemporal substitution parameter (σ_c), price indexation (ι_p), habit persistence (λ), and the autocorrelation coefficients of the price and wage markup shocks (ρ_w and ρ_p).

Table 4.3: Parameter Importance

	Parameter	$s_i(\theta)$
ρ_w	autocorr. wage markup shock	5.436
ρ	policy smoothing	2.928
ρ_p	autocorr. price markup shock	1.961
ξ_w	Calvo wages	1.778
ξ_p	Calvo prices	1.040
r_π	policy inflation	0.989
λ_w	wage markup	0.914
λ	habit	0.862
Φ	fixed cost	0.683
ρ_g	autocorr. gov. spending	0.569
ρ_a	autocorr. TFP	0.486
σ_c	elast.inter.subst.	0.386
$r_{\triangle y}$	policy output growth	0.367
ρ_I	autocorr. investment	0.310
μ_p	MA price markup shock	0.283
μ_w	MA wage markup shock	0.268
ψ	cap. utilization cost	0.266
α	share capital	0.204
φ	invest. adj. cost	0.172
σ_l	elast. hours	0.164
ι_w	indexation wages	0.143
δ	depreciation rate	0.119
ι_p	indexation prices	0.114
r_y	policy output	0.111
ρ_b	autocorr. risk premium	0.096
σ_b	std. dev. risk premium	0.092
σ_I	std. dev. investment	0.080
σ_a	std. dev. TFP	0.054
g_y	G/Y	0.049
σ_g	std. dev. gov. spending	0.036
σ_r	std. dev. policy	0.027
σ_w	std. dev. wage markup shock	0.024
σ_p	std. dev. price markup shock	0.021
ρ_r	autocorr. policy	0.019
β	discount factor	0.008

Note: The table shows the values of the sensitivity statistic $s_i(\theta) = \sqrt{\frac{1}{J}\sum_j^J \left(\theta_i \frac{\partial \tau_j}{\partial \theta_i}\right)^2}$ evaluated at the posterior mean of θ reported in Smets and Wouters (2007)

Table 4.4: Maximum pairwise collinearity (posterior mean)

		parametrization						
i		1	2	3	4	5	6	k
λ_w	wage markup	.99	.99	*fixed*	.99	*fixed*	*fixed*	ξ_w
ξ_w	Calvo wages	.99	.99	-.90	.99	-.90	-.90	λ_w (σ_l)
ξ_p	Calvo prices	.83	.83	.83	.83	.83	.83	ι_p
ι_w	indexation wages	.51	.51	.51	.51	.51	.51	ρ_p
ι_p	indexation prices	.83	.83	.83	.83	.83	.83	ξ_p
μ_w	MA wage shock	.55	.55	.55	.55	.55	.55	ξ_w
μ_p	MA price shock	.48	.48	.48	.48	.48	.48	ξ_p
α	capital share	.56	.56	.56	.56	.56	.56	$r_{\triangle y}$
ψ	cap. utilization cost	.32	.32	.32	.32	.32	.32	β
φ	invest. adj. cost	-.62	-.62	-.62	-.62	-.62	-.62	Φ
σ_c	elast.inter.subst.	.84	.84	.84	.84	.84	.84	λ
λ	habit	.84	.84	.84	.84	.84	.84	σ_c
Φ	fixed cost	.80	.80	.80	.80	.80	.80	ξ_p
σ_l	elast. hours	-.92	-.92	-.90	-.92	-.90	-.90	λ_w (ξ_w)
r_π	policy inflation	-.99	-.99	-.99	-.99	-.99	-.99	ρ
$r_{\triangle y}$	policy output growth	-.98	-.98	-.98	-.98	-.98	-.98	r_π
r_y	policy output	-.85	-.85	-.85	-.85	-.85	-.85	r_π
ρ	policy smoothing	-.99	-.99	-.99	-.99	-.99	-.99	r_π
δ	depreciation rate	-.74	*fixed*	-.74	-.74	*fixed*	*fixed*	λ
g_y	G/Y	.78	.78	.78	*fixed*	-	*fixed*	r_π
γ	trend	*fixed*	*fixed*	*fixed*	*fixed*	-.78	*fixed*	ψ
β	discount factor	-.45	-.45	-.45	-.45	.64	-.45	α
ρ_a	autocorr. TFP	-.37	-.37	-.37	-.37	-.37	-.37	σ_a
ρ_b	autocorr. risk premium	.29	.29	.29	.29	.29	.29	λ
ρ_g	autocorr. gov. spending	-.24	-.24	-.24	-.24	-.24	-.24	σ_b
ρ_I	autocorr. investment	.32	.32	.32	.32	.32	.32	σ_g
ρ_r	autocorr. policy	.18	.18	.18	.18	.18	.18	σ_I
ρ_p	autocorr. price shock	.82	.82	.82	.82	.82	.82	ξ_p
ρ_w	autocorr. wage shock	.85	.85	.84	.85	.84	.84	λ_w (ξ_w)
σ_a	std. dev. TFP	-.37	-.37	-.37	-.37	-.37	-.37	ρ_a
σ_b	std. dev. risk premium	.29	.29	.29	.29	.29	.29	λ
σ_g	std. dev. gov. spending	-.24	-.24	-.24	-.24	-.24	-.24	ρ_g
σ_I	std. dev. investment	.32	.32	.32	.32	.32	.32	ρ_I
σ_r	std. dev. policy	.18	.18	.18	.18	.18	.18	ρ_r
σ_p	std. dev. price shock	-.12	-.12	-.12	-.12	-.12	-.12	μ_p
σ_w	std. dev. wage shock	-.13	-.13	-.13	-.13	-.13	-.13	μ_w

Note: The table shows for each deep parameter θ_i the value of $\max_{j \neq i}\left(\frac{H_i'H_j}{(H_i'H_i)^{1/2}(H_j'H_j)^{1/2}}\right)$, where $H_l = \frac{\partial \tau}{\partial \theta_l}$ gives the effect on the reduced-form model of changes in θ_l. Values close to 1 or -1 indicate that H_i and H_k are nearly collinear. θ is evaluated at the posterior mean in SmetsWouters(2007)

Table 4.5: Multiple collinearity (posterior mean)

		parametrization					
		1	2	3	4	5	6
λ_w	wage markup	.999	.999	*fixed*	.999	*fixed*	*fixed*
ξ_w	Calvo wages	.999	.999	.996	.999	.996	.995
ξ_p	Calvo prices	.997	.997	.995	.997	.995	.995
ι_w	indexation wages	.976	.976	.976	.976	.976	.976
ι_p	indexation prices	.975	.974	.969	.974	.968	.967
μ_w	MA wage shock	.719	.698	.719	.719	.719	.697
μ_p	MA price shock	.877	.875	.869	.876	.867	.863
α	capital share	.983	.982	.980	.983	.980	.977
ψ	cap. utilization cost	.420	.417	.420	.420	.420	.417
φ	invest. adj. cost	.928	.923	.925	.927	.933	.919
σ_c	elast.inter.subst.	.997	.996	.996	.997	.996	.995
λ	habit	.993	.992	.993	.993	.994	.992
Φ	fixed cost	.991	.991	.990	.989	.987	.987
σ_l	elast. hours	.993	.992	.993	.993	.993	.992
r_π	policy inflation	.999	.999	.999	.999	.999	.999
$r_{\triangle y}$	policy output growth	.995	.995	.995	.995	.995	.995
r_y	policy output	.996	.995	.996	.996	.996	.995
ρ	policy smoothing	.999	.999	.999	.999	.999	.999
δ	depreciation rate	.990	*fixed*	.989	.990	*fixed*	*fixed*
β	discount factor	.983	.983	.978	.982	.983	.977
g_y	G/Y	.909	.908	.908	*fixed*	*fixed*	*fixed*
γ	trend	*fixed*	*fixed*	*fixed*	*fixed*	.994	*fixed*
ρ_a	autocorr. TFP	.922	.875	.918	.922	.918	.869
ρ_b	autocorr. risk premium	.843	.841	.790	.839	.786	.786
ρ_g	autocorr. gov. spending	.731	.642	.711	.729	.707	.619
ρ_I	autocorr. investment	.554	.551	.537	.546	.530	.528
ρ_r	autocorr. policy	.230	.227	.229	.230	.229	.226
ρ_p	autocorr. price shock	.996	.996	.996	.996	.996	.996
ρ_w	autocorr. wage shock	.997	.995	.997	.997	.997	.995
σ_a	std. dev. TFP	.447	.446	.445	.446	.444	.444
σ_b	std. dev. risk premium	.838	.836	.785	.834	.781	.781
σ_g	std. dev. gov. spending	.285	.284	.285	.274	.274	.274
σ_I	std. dev. investment	.370	.370	.367	.368	.365	.365
σ_r	std. dev. policy	.264	.259	.262	.264	.262	.258
σ_p	std. dev. price shock	.212	.211	.212	.211	.211	.209
σ_w	std. dev. wage shock	.191	.189	.191	.191	.191	.189

Note: The table shows for each deep parameter θ_i the value of $\left(\frac{H_i'H_{-j}}{(H_i'H_i)^{1/2}(H_{-j}'H_{-j})^{1/2}}\right)$, where $H_i = \frac{\partial \tau}{\partial \theta_i}$ gives the effect on the reduced-form model of changes in θ_i, and H_{-i} is the projection of H_i onto the space spanned by the other columns of H. Values close to 1 or -1 indicate that H_i and H_{-i} are nearly collinear. θ is evaluated at the posterior mean in SmetsWouters(2007)

Table 4.5 shows the values of the multiple collinearity measure for each of the deep parameters when θ is evaluated at the posterior mean in Smets and Wouters (2007). For 21 out of 35 parameters the degree of multiple collinearity exceeds .9, and for 14 of them it is greater than .99. Virtually all of the behavioral and technology parameters are very poorly identified according to this measure. Apart from the elasticity of the capacity utilization cost function (ψ), the only parameters in the model that do not suffer from severe interdependence problem are the stochastic shock parameters, with the exception of the autocorrelation coefficients of the sector-neutral technology, price and wage markup shocks (ρ_a, ρ_p, ρ_w), and the standard deviation of the risk premium shock (σ_b). Fixing some of the parameters in the other 5 parameterizations leads to only marginal improvements.

These results suggest that most of the parameters estimated in Smets and Wouters (2007) are very poorly identified in the (linearized) theoretical model, which, as I explained before, implies that they will be poorly identified when the model is estimated. I should emphasize however, that the results in Tables 4.4 and 4.5 are conditional on the value of θ that I used, namely the posterior mean reported in Smets and Wouters (2007). It is possible that in other points in the parameter space the identification is much better. To examine if such is the case I used the admissible points from Θ that were drawn in the previous section. These points were ordered according to the condition number of $H(\theta)$, and eleven of them were selected - those yielding the smallest value and the 10 deciles of the distribution of cond($H(\theta)$). For each of these eleven values of θ I computed the pairwise and multiple collinearity values as I did in Tables 4.4 and 4.5. The results are shown in Tables B.9 and B.10 in the Appendix. We see that the poor parameter identification in the model is not a problem only at the particular point I studied before. Even though there is some variability in the degrees of pairwise or multiple collinearity, the parameters I found before to be poorly identified remain so for all values of θ I checked.

Table 4.6: Worst identified parameters

		Multiple collinearity		
		all	subset	most important
λ_w	wage markup	.9999	.999	σ_c ,ξ_w ,σ_l ,ρ_w, g_y
σ_c	elast.inter.subst.	.997	.980	β ,α ,λ ,ι_w ,σ_l ,$r_{\triangle y}$,ρ_a
λ	habit	.993	.968	δ ,σ_c , ,ρ_b ,ρ_w ,σ_b
Φ	fixed cost	.991	.976	φ ,λ ,ι_p ,ξ_p ,σ_l ,ρ_w ,
ξ_w	Calvo wages	.999	.994	λ_w ,σ_l ,ρ_w
ξ_p	Calvo prices	.997	.980	λ_w ,Φ ,ξ_w ,ι_p ,σ_l ,ρ_p
σ_l	elast. hours	.993	.962	λ_w ,σ_c ,ξ_w ,r_y
r_π	policy inflation	.999	.999	g_y ,$r_{\triangle y}$,r_y ,ρ
r_y	policy output	.996	.971	λ_w ,g_y ,β ,ι_w ,ξ_w ,ξ_p ,σ_l ,r_π ,ρ_w
ρ	policy smoothing	.999	.991	δ ,φ ,ι_p ,r_π ,$r_{\triangle y}$

Note: Parameters that have multiple collinearity coefficients larger than .95 everywhere in Θ. Column 3 shows the values of the multiple collinearity at the posterior mean when all other parameters in θ are used. Column 4 gives the multiple collinearity when only a subset of the most important parameters (shown in column 5) are used.

Table 4.6 presents the deep parameters that tend to be very poorly identified everywhere in the parameter space. They are the ones with a value of the multiple collinearity exceeding .95 at all points in Θ I checked. Column 3 gives the multiple collinearity (computed at posterior mean from Smets and Wouters (2007)) when all other parameters in θ are used, i.e. when column H_i is projected onto the space spanned by all other columns of matrix H. However, not all other parameters are very important in explaining the role a given deep parameter plays in the model. For each of the worst identified parameters I determined the parameters that are most important using techniques from the model selection literature[23]. These parameters are listed in the last column, and column 4 shows the values of the multiple correlation when only the subsets of the most important parameters are used. For instance, the Calvo parameter for wages (ξ_w) has a multiple collinearity of .999 when all other deep parameters are used. If the multiple collinearity is computed only with respect to σ_l, λ_w, and ρ_w, the value is .994. These parameters are: elasticity of labor sup-

[23]Specifically, I used the elastic net algorithm proposed by Zou and Hastie (2005). See the appendix for more details.

ply, steady state wage markup, and autocorrelation coefficient of the shock to wage markup. I know (see Table 4.4) that the presence of λ_w alone is sufficient to have a collinearity of .99. However, even if λ_w is removed (i.e. assumed to be known), the degree of multiple collinearity would exceed .95 as long as σ_l, and ρ_w remain among the included parameters. On the other hand even if σ_l, and ρ_w are removed in addition to λ_w, the degree of multiple collinearity of ξ_w would still exceed .9. The only way to reduce it to below that level is to remove *all* of the following 8 parameters: $\lambda_w,\ \sigma_c,\ \sigma_l\ , \rho_w,\ \rho_p,\ \alpha,\ \xi_p,\ \lambda$

Remark. It is worth mentioning that parameter interdependence problems in the latest version of the "Smets and Wouters" model are somewhat less severe than those in earlier versions. This is due to the simpler structure of the current model. For instance, the autocorrelation coefficient of the preference shocks, present in the previous versions of the model, was very difficult to distinguish from the habit persistence and elasticity of intertemporal substitution parameters λ and σ_c. In the current version a similar role is played by the risk premium shock, but the interdependence among ρ_b, λ and σ_c is not as strong as before. Another change that improves the identifiability of the model is the simplified monetary policy rule. In the earlier versions of the model the central bank responded to both past inflation and output gap, which was making it difficult to separately identify the response coefficients for current and past inflation and output gap.

Strong interdependence among the parameters makes it difficult to identify them separately. As a consequence not only the point estimates will be imprecise, but also the standard measures of estimation uncertainty, based on the *marginal* distributions of the estimates, will be misleading. For instance, constructing confidence intervals using the estimated standard errors would underestimate the true sampling uncertainty. The same holds for the highest (marginal) posterior density intervals, typically reported when Bayesian techniques are used for estimation. The problem

with these measures of estimation uncertainty is that they allow for variations in only one parameter estimate at a time. If two parameter estimates are correlated, allowing for simultaneous variation in both will cover a much wider range of values. I return to this point in Section 4 where I explain how I construct confidence intervals for the maximum likelihood estimates.

The above analysis suggests an extension of the identification analysis procedure outlined in Section 4, which would help gain a better understanding of the causes for weak identification in the model. Repeating steps 1 to 4 from Section 4 many times provides information on whether the necessary and sufficient condition for identification are satisfied, and keeping track of the condition numbers of $\mathcal{I}_\theta$ and H, tells us about the strength of identification. The model-related causes for weak identification result in a large condition number of H. To find out what features of the model are responsible one should:

1. Compute the sensitivity $s_i(\theta)$ using (4.47) for each column i of H. Small values of $s_i(\theta)$ imply that the parameter θ_i has only a marginal effect in the model.
2. Compute the pairwise and multiple collinearity measures (4.48). Values close to one imply that the role of parameter θ_i in the model is very well approximated by a combination of other deep parameters.

In Appendix B I provide a summary of all steps involved in the identification and weak identification analysis presented here and in section 4

To summarize, the objective in this section was to study the identifiability of model described in section 4. I started by drawing randomly a large number of points from the parameter space, and evaluating the conditioning of the Information matrix at those points. I found that the matrix is generally very poorly conditioned, which suggest that the identification of the model parameters as a whole is week. I also found that matrix $H(\theta)$, which depends only on the parameter values and the structure of the model, is poorly condition too, indicating that problems originate in the model. Studying the columns of that matrix and relationships among them

allowed us to determine the causes for poor identification in the model, as well as to get a better sense of severity of these problems. I found that a large number of deep parameters are strongly interdependent - the effect of each one of them can be very well replicated by a combination of other deep parameters. Moreover, parameters for which interdependence is not a serious problem, such as most of the stochastic shock parameters, are also those with respect to which the reduced-form model, and therefore the likelihood in not very sensitive. I found that these problems occur pretty much everywhere in the parameter space, suggesting that the problem is a global one. Moreover, since the poor identifiability is largely due to the model, it is unlikely that having more observed variables or longer time series would be of much help.

Here I studied the identification of the theoretical model as it is, without reference to a particular data set used for estimation. Thus the problems I found may arise whenever this or similar DSGE models are estimated. To find out how strong identification is at a particular parameter estimate, that is, conditional on a specific data set, one should examine the conditioning of the information matrix evaluated at that particular point. Furthermore, if Bayesian techniques are used for estimation, in addition to the posterior mode, one could also evaluate the conditioning of the information matrix for all points from the posterior distribution. I return to that in the next section, after the estimation results are presented.

4.4 Case Study: Estimation

The results from the previous section suggest that the likelihood, and therefore the data, is not very informative about the parameters of the model. One consequence of this is that estimating the model using Bayesian techniques, as in Smets and Wouters (2007), one places relatively large weight on the priors compared to the likelihood. To explore this further, in this section I estimate the model by maximizing the the likeli-

hood only, and then compare the results with the posterior mode estimates reported in Smets and Wouters (2007)

I start by describing the data to which the model is applied. Then I turn to estimation of the model.

4.4.1 Data

The model is estimated using quarterly US data over the period 1966:1-2004:4. The observed variables are: real consumption (c), real investment (i), real output (y), real wages (w), hours (h), inflation (π), and the nominal interest rate (r).

Consumption is personal consumption expenditures. Investment is fixed private investment. Wages are hourly compensation for nonfarm business. Real consumption, investment and wages are obtained by deflating the nominal variables with the GDP implicit price deflator. Real output is real GDP. Hours are average hours for nonfarm business. Inflation is the first difference of the log GDP implicit price deflator. Consumption, investment, and output are expressed in per capita terms by dividing with civilian population of 16 and older. The nominal interest rate is the quarterly average of the Federal Funds rate.

More details on the definitions and data sources used are provided in the data Appendix to Smets and Wouters (2007).

4.4.2 Estimation

Maximizing the likelihood function and the posterior density

Both MLE and Bayesian estimation require the evaluation of the likelihood function. To do that I first solve the linearized structural model (4.3) to find the state equation (4.4); then the Kalman filter is used to evaluate the log-likelihood $l(Z;\theta) = \ln L(Z;\theta)$ of the reduced-form model (4.4)-(4.5). In order to keep the estimate of θ within the-

oretically meaningful bounds, I optimize the likelihood with respect to unbounded variables that are one-to-one transformations of the restricted variables in the θ. The bounds on the parameters in θ are shown in Table B.13, and are the same as those used by Smets and Wouters (2007). In addition, when computing the likelihood I impose the restriction that the model has a unique solution. This is achieved by setting the value of the likelihood to a very small number for values of θ that result in multiple or no solutions.

Using the Bayes rule, the posterior density can be expressed as

$$p(\theta; Z) = \frac{L(Z; \theta)p(\theta)}{p(Z)} \propto L(Z; \theta)p(\theta) \tag{4.49}$$

where $p(\theta)$ denotes the prior distribution of θ. Thus, to maximize the posterior density, I evaluate the likelihood, as before, and the prior $p(\theta)$, which alternatively may be though of as a penalty function.

A well-known practical problem with non-linear optimization/estimation is that one cannot be certain that a global maximum is found, and not just a local one. A common strategy for dealing with this is to try many different starting values. My approach was to combine simulation techniques, gradient and non-gradient based optimization methods. I started with picking ten of the points drawn for the purpose of identification analysis (see section 4), which yielded the highest values of the likelihood or the posterior density. Then, taking these points as starting values, I run ten Markov chains generated by the random walk implementation of the Metropolis-Hastings algorithm (I follow Schorfheide (2000)). The modes of the distributions generated by each chain were then used as starting values for several optimization routines, and the final maximizer was determined by direct comparison of the resulting values.

Results

I estimate two different parameterizations of the model. In the first one three of the identified parameters - depreciation rate δ, wage markup λ_w, and government spending share in output g_y, are assumed known instead of estimated. This is the parametrization estimated in Smets and Wouters (2007). The values at which these parameters are fixed - .025, 1.5 and .18, respectively, are also taken from that paper. In the second parametrization these parameters are estimated.

I follow Smets and Wouters (2007) and estimate the model using data for the full sample period (1966:1-2004:4), and for two subperiods (1966:1-1979:2 and 1984:1-2004:4). This is done in order to investigate the sources of the differences in the economic environment during these two periods.

The estimation results for the first parametrization are presented in Tables 4.7 (deep parameters), and 4.8 (shock parameters). In addition to the maximum likelihood estimates, and the posterior mode values from Smets and Wouters (2007), I report the 90% confidence intervals. Also, the values of the log likelihood as well as the condition number of the information matrix evaluated at the respective point estimates are shown.

Before I turn to the discussion of the results, I should explain how the confidence intervals I report were obtained. For the Bayesian estimates I show the 5-th and 90-th percentile of the marginal posterior distribution of the parameters. The numbers are taken from Smets and Wouters (2007), and are obtained from the output of the Metropolis-Hastings algorithm used for sampling from the posterior distribution. Regarding the ML estimates, theoretically one could compute confidence intervals using the fact that the Information matrix is the inverse of the asymptotic covariance matrix. The diagonal elements of the inverse are, therefore, the estimated standard errors, and can be used to construct asymptotic confidence intervals. There are two problems with this approach. First, as I explained in Section 4, even small errors

Table 4.7: Estimation Results: Deep Parameters 1966:1-2004:4

Param.	Prior			Bayesian			MLE		
	Distr.	Mean	St.dev.	5%	mode	95%	5%	mode	95%
φ	$\mathcal{N}$	4.00	1.50	3.97	5.49	7.42	1.84	8.00	21.31
σ_c	$\mathcal{N}$	1.50	0.38	1.16	1.40	1.59	1.22	1.78	2.88
λ	$\mathcal{B}$	0.70	0.10	0.64	0.71	0.78	0.41	0.70	0.86
ξ_w	$\mathcal{B}$	0.50	0.10	0.60	0.74	0.81	0.62	0.88	1.03
σ_l	$\mathcal{N}$	2.00	0.75	0.91	1.92	2.78	-0.04	2.94	8.69
ξ_p	$\mathcal{B}$	0.50	0.10	0.56	0.66	0.74	0.41	0.67	0.87
ι_w	$\mathcal{B}$	0.50	0.15	0.38	0.59	0.78	0.05	0.73	1.61
ι_p	$\mathcal{B}$	0.50	0.15	0.10	0.23	0.38	0.01	0.01	0.01
ψ	$\mathcal{B}$	0.50	0.15	0.36	0.55	0.72	0.25	0.76	1.56
Φ	$\mathcal{N}$	1.25	0.12	1.48	1.61	1.73	1.34	1.86	2.56
r_π	$\mathcal{N}$	1.50	0.25	1.74	2.03	2.33	1.78	2.62	10.66
ρ	$\mathcal{B}$	0.75	0.10	0.77	0.82	0.85	0.82	0.87	0.98
r_y	$\mathcal{N}$	0.12	0.05	0.05	0.08	0.12	0.07	0.13	0.79
$r_{\triangle y}$	$\mathcal{N}$	0.12	0.05	0.10	0.22	0.38	0.16	0.25	0.52
$\bar{\pi}$	$\mathcal{G}$	0.62	0.10	0.61	0.82	0.96	0.61	0.98	1.66
$100(\beta^{-1}-1)$	$\mathcal{G}$	0.25	0.10	0.07	0.16	0.26	0.01	0.01	0.01
$\bar{l}$	$\mathcal{N}$	0.00	2.00	0.07	-0.10	0.26	-2.62	-0.30	2.03
γ	$\mathcal{N}$	0.40	0.10	0.40	0.43	0.45	0.33	0.42	0.47
α	$\mathcal{N}$	0.30	0.05	0.16	0.19	0.21	0.10	0.18	0.25
Log Likelihood:					-840.11			-820.36	
$cond(\mathfrak{I}_\theta)$:					2.7e7			4.4e7	

Note: $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed. $\bar{\pi}$, and $\bar{l}$ are quarterly steady state inflation rate, and steady state hours worked.

in a poorly conditioned Information matrix lead to large errors in its inverse. The simulation evidence discussed in Section 4 show that standard errors and confidence intervals obtained in this fashion are likely to be very misleading. Second, as I discussed in section 4, even when the standard errors ($S.E.$) are precisely estimated, the

usual confidence intervals of the form $\hat{\theta}_j \pm S.E.(\hat{\theta}_j) \times crit.value$ may be very misleading if a strong degree of interdependence exists among the parameter estimates. As I explained in section 4, the reason is that standard confidence intervals are based the marginal distribution of the estimates, and when dependence among parameter estimates is strong, the product of the marginal distributions is quite different from the joint distribution.

Because of these two reasons I use an alternative approach for constructing confidence intervals, namely inverting the likelihood ratio test statistic. This is a standard approach for obtaining approximate confidence regions, and uses the result that, for θ in the neighborhood of the MLE $\hat{\theta}$, the likelihood ratio statistic $2\big(l(\hat{\theta}) - l(\theta)\big)$ is asymptotically distributed as $\chi^2(k)$, where k is the dimension of θ. The $100(1-a)\%$ asymptotic confidence region contains all θ in the neighborhood of the $\hat{\theta}$ for which the likelihood ratio statistic does not exceed the upper a percentile of the chi-squared distribution with k-degrees of freedom.

Starting with the deep parameters estimated over the whole sample, the results show significant differences between the MLE and Bayesian estimates for most of them. Particularly large is the effect on φ, σ_l, ι_p, ι_w , ξ_w , ψ, and r_π. Smaller, but still substantial are the differences for σ_c, Φ, r_y, $\bar{\pi}$, and $\bar{l}$. For the remaining parameters the estimates are very close.

The maximum likelihood estimates of both Calvo parameters, ξ_p and ξ_w, are higher than their Bayesian estimates. This implies longer average duration of the wage (6.3 vs. 3.9 quarters) and price (3.1 vs. 2.9 quarters) contracts. The estimates of ι_p and ι_w suggest much larger degree of indexation of wages, and much weaker degree of price indexation than those implied by the Bayesian estimates.[24]

The elasticity of the investment adjustment cost function (φ) is also larger according to the ML estimates, as are fixed cost parameter (Φ), and the elasticity of

[24]This findings are consistent with the remarks in Smets and Wouters (2007) on the effect of relaxing their priors. See their footnote 9.

the capacity utilization adjustment cost function (ψ).

Overall, for all frictions in the model, except the habit persistence parameter (h), the ML estimates are substantially different and larger than the Bayesian ones. The latter are in turn larger than the respective means of the prior distribution, which is therefore the most likely explanation of the observed discrepancies.

The ML estimate of the monetary policy rule parameters suggest a much stronger interest rate response to inflation and output gap, and slightly stronger response to the change in output gap; the degree of interest rate smoothing is also higher, according to the ML estimate. Again, these differences between the Bayesian and the maximum likelihood estimates can be attributed to the use of the particular prior values.

Turning to the estimates of the exogenous shock parameters, presented in Table 4.8, we see that the MLE and Bayesian estimates are quite close. One exception is the autocorrelation parameter of the policy shock (ρ_r), which is estimated to be substantially larger when a prior (with mean of .5) is used. This confirms the observation made in Smets and Wouters (2007) that "the data appear to be very informative on the stochastic processes of for the exogenous disturbances" (p.9). One implication of this is that we should expect that the forecast error variance decompositions of the model variables will be quite similar across the two sets of estimates.

The results from the estimation of the model using data for the two subsamples are shown in Tables B.14 and B.15. There we observe much larger discrepancies between the maximum likelihood and Bayesian estimates of the deep parameters. For some parameters, for instance r_π for the first subperiod, and φ - for the second, the ML estimates were pushed towards the bounds for those parameters. Similar experience, resulting from relaxation of the prior precision, was reported in Onatski and Williams (2004). One possible explanation of these discrepancies is that much less data is used for estimation, which makes the likelihood relative less informative, and

Table 4.8: Estimation Results: Shock Processes 1966:1-2004:4

Param.	Prior			Bayesian			MLE		
	Distr.	Mean	St.dev.	5%	mode	95%	5%	mode	95%
ρ_a	$\mathcal{B}$	0.50	0.20	0.94	0.96	0.97	0.93	0.97	0.99
ρ_b	$\mathcal{B}$	0.50	0.20	0.07	0.18	0.36	-0.27	0.15	0.69
ρ_g	$\mathcal{B}$	0.50	0.20	0.96	0.98	0.99	0.95	0.98	1.00
ρ_I	$\mathcal{B}$	0.50	0.20	0.61	0.71	0.80	0.36	0.70	0.95
ρ_r	$\mathcal{B}$	0.50	0.20	0.04	0.13	0.24	0.01	0.01	0.01
ρ_p	$\mathcal{B}$	0.50	0.20	0.80	0.90	0.96	0.70	0.93	1.00
ρ_w	$\mathcal{B}$	0.50	0.20	0.94	0.97	0.99	0.88	0.98	1.00
ρ_{ga}	$\mathcal{B}$	0.50	0.25	0.37	0.53	0.66	0.03	0.45	0.82
μ_w	$\mathcal{B}$	0.50	0.20	0.75	0.89	0.93	0.85	0.96	0.99
μ_p	$\mathcal{B}$	0.50	0.20	0.54	0.74	0.85	0.12	0.73	0.93
σ_a	$\mathcal{IG}$	0.10	2.00	0.41	0.45	0.50	0.36	0.44	0.54
σ_b	$\mathcal{IG}$	0.10	2.00	0.19	0.24	0.27	0.13	0.24	0.37
σ_g	$\mathcal{IG}$	0.10	2.00	0.48	0.52	0.58	0.45	0.54	0.69
σ_I	$\mathcal{IG}$	0.10	2.00	0.37	0.45	0.53	0.33	0.45	0.73
σ_r	$\mathcal{IG}$	0.10	2.00	0.22	0.24	0.27	0.19	0.23	0.30
σ_p	$\mathcal{IG}$	0.10	2.00	0.11	0.14	0.16	0.05	0.12	0.21
σ_w	$\mathcal{IG}$	0.10	2.00	0.20	0.24	0.28	0.20	0.26	0.38
Log Likelihood:					-840.11			-820.36	
$cond(\mathfrak{I}_\theta)$:					2.7e7			4.4e7	

Note: $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed. $\bar{\pi}$, and $\bar{l}$ are quarterly steady state inflation rate, and steady state hours worked.

the priors - relative more influential with respect to the posterior distribution. This is indicated by the high value of the condition numbers of the information matrix. These values are quite high even when all data is used, but particularly so for two subsample estimates.

Remark. Unlike in Section 4, where I computed condition numbers of the information matrix at the true values of θ, here the parameters are estimated, and therefore

subject to sampling uncertainty. Accounting for this uncertainty is straightforward for the estimates obtained with Bayesian methods. I can simply find the posterior distribution of cond($\mathfrak{I}_\theta$). The 5-th and 95-th percentiles of the distribution are $2.4e7$ and $2.9e7$, respectively. It is not obvious how to put similar confidence bounds on the condition number of the information matrix evaluated at the ML estimates.

The results reported in Tables 4.7 and 4.8 were obtained under the assumption that δ, λ_w, and g_y are known and fixed at the values assumed in Smets and Wouters (2007). As I discussed in section 4 the reason given for not estimating these parameters was their poor identification. However, I found evidence supporting that claim only with respect to λ_w. In Table B.16 I report the maximum likelihood estimates of the model parameters obtained when δ, λ_w, and g_y are assumed unknown and also estimated. The values I estimated for these parameters are $\hat{\delta} = .021$, $\hat{\lambda}_w = 1.77$, and $\hat{g}_y = .3$. Turning to the other parameters, the effect is most noticeable for the policy rule parameters, the estimates of all of which increase substantially. The higher condition numbers (6.7e10 vs. 2.7e10) suggest that the identification of this parametrization is indeed weaker. However, the difference is not particularly large and is, at least partly, due to the large number of parameter estimated in the second case.

Overall, I find that the use of priors have significant effects on the parameter estimates for the model I consider. This by itself does not imply that the model behavior is also affected substantially. To assess the implications of different estimates on the internal dynamics and the propagation mechanism of the model, I next compare the impulse responses to the structural shocks, and the variance decompositions for the observed variables.

Impulse responses and variance decompositions

Impulse responses and variance decompositions are standard tools for gauging the behavior of macroeconomic models, and assessing their credibility. Impulse response analysis allows us to trace the dynamic interactions among economic variables, while the variance decompositions measure the contribution of each structural shock to the total variation of each variable. Here I compare the implications along these two dimensions of three different parameter estimates for the whole sample period (1966:1-2004:4) - the Bayesian and ML estimates for the first parameterizations (columns 2 and 3 of tables 4.7 and 4.8), and the ML estimate for the second parametrization (columns 2 and 6 of Table B.14 in the Appendix). For ease of notation, henceforth I refer to the first two estimates as SW and MLE1, and the the last one - as MLE2.

Figures B.2 - B.8 plot the impulse responses (percent deviations from steady state level) of the seven observed variables (output, consumption, investment, hours, inflation, wages, and interest rate) to a one standard deviation in each of the seven structural shocks (productivity, risk premium, government spending, investment, monetary policy, price and wage markup shocks). Overall, the responses seem reasonable, and are, in most cases, qualitatively similar in the sense of having the same sign on impact and similar dynamics. In particular, most impulse responses implied by the two ML estimates are very close. The most common difference between MLE1 and MLE2 on one hand, and SW - on the other, are in the magnitude and persistence of the responses. For instance, the responses of output and consumption to productivity, investment or price markup shocks, take longer to reach their peaks, and last longer under the MLE, compared to SW estimates.

The opposite is true for the response of most variables, and particularly investment and wages, to a wage markup shock. In some cases there is also a substantial difference in the impact effect of the shocks. For instance, wages and inflation respond much more strongly to monetary policy, productivity, risk premium, or government

spending shocks, under the SW estimates compared to the MLE ones. In the case of response of wages to exogenous spending shock, the impact effects are also in different directions (see Figure B.4). Under SW the response is positive and remains so for up to 10 quarters, while the two ML estimates imply a smaller and negative response.

Tables B.17 - B.19 report, for the three parameter estimates - SW, MLE1 and MLE2 respectively, the contributions of each structural shock to the forecast error variances of the observed variables at different horizons. As with the impulse responses, the results are broadly similar, with some differences emerging in the medium to long-run horizon. With respect to the determinants of output, for instance, the Bayesian parameter estimates overemphasize, relative to the ML ones, the importance of wage markup, exogenous spending, and risk premium shocks, and underestimate that of sector-neutral productivity, and price markup shocks. Similar differences may be observed regarding inflation. Relative to the ML estimates, the Bayesian estimates overestimate the importance of risk premium, exogenous spending, investment and monetary policy shocks, and underestimate the importance of price markup shocks. These differences are again more significant at medium and long-run horizons.

Similar differences in the importance assigned to different structural shocks can be observed with respect to the other variables in the model. One property that all estimates have in common is that "demand" shocks, such as government spending, risk premium, or investment-specific shocks, are the main driving forces behind the fluctuations in output in a short run. According to both the Bayesian and ML estimates, these shocks 50% to 70% of the forecast error variance of output at horizons of 1 to 4 quarters. On the other hand, at medium to long-run, "supply" shocks - productivity, price and particularly wage markup shocks, are the main driving forces behind the fluctuations in output, explaining between 60% and 80% of the forecast error variance of output at horizons of 10 years and beyond. These observations were made in Smets and Wouters (2007), and as the results show, are robust to the method

used for estimation.

4.5 Conclusion

One of the main promises of the rapidly expanding literature on empirical evaluation of DSGE models, is that we can now estimate rich micro-founded structural models that until recently had to be calibrated. However, the extent to which this is of practical use depends crucially on whether the parameters we want to estimate are well identified. In this paper I developed a new methodology that can be used to address the following questions - are the parameters identified, how strong is identification, are the identification problems inherent in the structure of the model, or due to data deficiencies - for any linearized DSGE model. I then applied this methodology to study parameters identification of a state-of-the-art monetary DSGE model, that is widely regarded as one of the success stories of the empirical DSGE literature. I found that many of the parameters of the model are very poorly identified virtually everywhere in the parameter space. In addition, my results suggest that the problem to a large extent originates in the structure of the model. Thus, it is likely that other models in the empirical DSGE literature, that share features of the model I considered, also suffer from weak parameter identifiability. I showed how parameter interdependence problems can be detected and possibly alleviated by reparametrization. For the model I considered this improved, but unfortunately did not fully solve the identification problem. Estimating the model by maximum likelihood, I found substantial differences in the parameter estimates compared to those obtained with Bayesian methods. I attribute those differences to the the use of priors in the latter.

Are these differences important? The answer of this question depends on the purpose of estimating the model in the first place. For instance, using estimated DSGE models solely for forecasting purposes does not require knowledge of the values of

behavioral or technology parameters. Similarly, if the estimated model is used to conduct impulse response and variance decomposition analysis, then the strength of parameter identification is not very important. We saw evidence to that effect in the last section, where quite different parameter values often implied very similar, and even identical impulse response functions, or variance decomposition results. This should not be surprising, as by definition weak local identification means that different deep parameters imply very similar reduced-form dynamics. However, when estimated DSGE models are used for policy analysis, such as designing optimal monetary policy, the values of the deep parameters may be of crucial importance. This is because for the purpose of such analysis one needs to work with non-linear versions of the model, for which the implications of different parameter values are likely to be stronger than in the linearized version of the model.

These results may cause one to seriously doubt the validity of parameter estimates reported in some of the empirical DSGE literature. For instance, in their empirical comparison of the US and Euro area business cycles, Smets and Wouters (2005) conclude that the structures of the two economies are very similar, and have not changed much over time. Since the model they estimate is similar to the one in this paper, these findings may be explained with the fact that they use the same prior distributions for both economic areas, and the different sample periods. Of course, if the priors are chosen so that they truly reflect the researcher's a priori beliefs for the parameters of interest, weak identification is not an issue, as long as care is taken to sample from the true posterior distribution. I believe, however, that even when this is the case, conducting and reporting the results of identification analysis as described here, would help in communicating one's findings to a broader audience, who may not hold the same subjective beliefs as the author. Providing such information would help the reader assess the relative importance of the data and the priors, and let her judge for herself the credibility of the reported estimates. Also, as we saw in Section

4, the current practice of reporting percentiles of the marginal posterior distributions, or showing plots of these distributions along with the distributions of the priors, may sometimes be misleading. This would be the case when there is strong dependence among some parameters. What these parameters are could be determined with the help of matrix $H(\theta)$ defined in Section 4. Instead of the marginal distributions, one should report results based on the joint posterior distributions for parameters parameters that are found to be strongly dependent.

Given the increasing popularity of empirical DSGE analysis, one may wonder whether the problems I have discussed in this paper are specific to the model I consider, or endemic, as the analysis in Beyer and Farmer (2004) may lead one to believe. To partially answer this question, I carried out the identification analysis described in section 4 for three different DSGE models - a prototypical three-equations New Keynesian model, a standard one-sector stochastic growth model, and a two-country monetary New Open Economy model. The first two are stripped-down versions of the main model, focusing on features that are important in the New Keynesian and the RBC economics, respectively. The third one is an example of a model which is comparable, in terms of size and number of parameters, to the Smets and Wouters (2007) model, but simpler in terms of structural features. More information on the models, and the results from the identification analysis is provided in the Appendix. I find that parameter identification in these models, is much stronger than in the large scale New Keynesian model adopted in this paper. Thus the problem with identification is not necessarily generic, and should be addressed for each DSGE model separately.

One way to deal with the identification problems, when such are detected, is to re-parameterize the structural model and estimate parameters that are well identified. This would be an useful approach in situations where the values of the individual deep parameters are not of primary interest, and estimating functions of such parameters is also acceptable. As I suggested above, if the DSGE model is used for forecasting,

or to study the dynamic responses of economic variables to structural shocks, this can be accomplished without estimating deep parameters. Moreover, in such situations many of the cross-equation restrictions imposed when the deep parameters are estimated, can be relaxed, thus making the results robust to larger classes of models.

Another possible solution is to work with higher order approximations instead of linearized models. McManus (1992) proves that identification failures are much rearer in non-linear than in linear models, and argues that using linear approximations is a major cause for poor parameter identifiability in econometrics. Although the estimation of non-linear DSGE models is computationally much more demanding, recent work by Fernandez-Villaverde and Rubio-Ramirez (2005), An (2005), and **?** have shown how it could be accomplished. However, the procedures for studying identification proposed here cannot be applied to non-linear models. The development of appropriate methods is left for future work. Another question suggested by the findings in this paper, is whether the difficulties with identification of some of the preference parameters is specific to the Smets and Wouters (2007) model, or would arise in any model with the same specification of the consumer preferences. This is also left for future investigation.

Appendices

Appendix A

Appendix for Chapter 1

Derivation of (2.15).

Consider equation (2.13). Vectorizing both sides yields

$$(A' \otimes I_m)\mathbf{vec}(\Gamma_0) - (A'^2 \otimes I_m)\mathbf{vec}(\Gamma_1) - \mathbf{vec}(\Gamma_2) = 0_m \tag{A.1}$$

where I_m and 0_m are $m \times m$ identity matrix and a zero matrix, respectively. Each of the three Γ matrices contains zeros, other known constants, and elements of γ. Let $\tilde{\gamma} = [1,\ \gamma']'$. Then, for $i = 0, 1, 2$, we can write

$$\mathbf{vec}(\Gamma_i) = G_i\tilde{\gamma} \tag{A.2}$$

where G_i is a $m^2 \times (l+1)$ matrix containing only zeros and ones.

Thus equation (A.1) becomes

$$\big((A' \otimes I_m)G_0 - (A'^2 \otimes I_m)G_1 - G_2\big)\tilde{\gamma} = 0_m \tag{A.3}$$

where $\tilde{\gamma} = \tilde{\gamma}(\theta)$.

Denoting the first column of $\big((A' \otimes I_m)G_0 - (A'^2 \otimes I_m)G_1 - G_2\big)$ with $-\Xi_1$, and the remaining l columns with Ξ_1, (A.3) becomes

$$-\Xi_1 + \Xi_2\gamma = 0$$

or

$$\Xi_2\gamma = \Xi_1$$

Proof of Lemma 2.4.1

1 The consistency of $\hat{\theta}_{MD}$ follows from the Slutsky's theorem, and the assumption that $W_T \longrightarrow W$ and $\hat{\tau} \xrightarrow{p} \tau_0$. By continuity of $f(.)$ we have

$$f(\hat{\tau}, \theta) \xrightarrow{p} f(\tau_0, \theta)$$

and

$$f(\hat{\tau},\theta)'W_T f(\hat{\tau},\theta) \xrightarrow{p} f(\tau_0,\theta)'W f(\tau_0,\theta)$$

Finally, since

$$\hat{\theta}_{MD} = \operatorname{argmin}\left(f(\hat{\tau},\theta)'W_T f(\hat{\tau},\theta)\right)$$

and θ_0 is identified, it follows that

$$\hat{\theta}_{MD} \xrightarrow{p} \theta_0$$

2. By the Mean Value theorem we have

$$f(\hat{\tau},\hat{\theta}_{MD}) = f(\hat{\tau},\theta_0) + f_\theta(\hat{\tau},\theta')(\hat{\theta}_{MD}-\theta_0) \tag{A.4}$$

$$f(\hat{\tau},\theta_0) = f(\tau_0,\theta_0) + f_\tau(\tau',\theta')(\hat{\tau}-\tau_0) \tag{A.5}$$

where $\|\theta'-\theta_0\| \leq \|\hat{\theta}_{MD}-\theta_0\|$ and $\|\tau'-\tau_0\| \leq \|\hat{\tau}-\tau_0\|$. Substituting the expression for $f(\hat{\tau},\theta_0)$ from (A.5) in (A.4) yields

$$f(\hat{\tau},\hat{\theta}_{MD}) = f_\tau(\tau',\theta')(\hat{\tau}-\tau_0) + f_\theta(\hat{\tau},\theta')(\hat{\theta}_{MD}-\theta_0) \tag{A.6}$$

where I have used that $f(\tau_0,\theta_0)=0$.
The consistency of $\hat{\tau}$ and $\hat{\theta}_{MD}$ allows us to write (A.6) as

$$f(\hat{\tau},\hat{\theta}_{MD}) = f_\tau(\tau_0,\theta_0)(\hat{\tau}-\tau_0) + f_\theta(\tau_0,\theta_0)(\hat{\theta}_{MD}-\theta_0) + o_p(1) \tag{A.7}$$

Pre-multiplying both sides of (A.7) by $f_\theta(\hat{\tau},\hat{\theta}_{MD})'W_T$ and using that

$$f_\theta(\hat{\tau},\hat{\theta}_{MD})'W_T f(\hat{\tau},\hat{\theta}_{MD}) = 0$$

as a first order condition of the minimum distance estimator of θ, gives

$$f_\theta(\hat{\tau},\hat{\theta}_{MD})'W_T f_\tau(\tau_0,\theta_0)(\hat{\tau}-\tau_0) + f_\theta(\hat{\tau},\hat{\theta}_{MD})'W_T f_\theta(\tau_0,\theta_0)(\hat{\theta}_{MD}-\theta_0) + o_p(1) \tag{A.8}$$

and therefore

$$\sqrt{T}(\hat{\theta}_{MD}-\theta_0) = \left(f_\theta(\hat{\tau},\hat{\theta}_{MD})'W_T f_\theta(\tau_0,\theta_0)\right)^{-1} f_\theta(\hat{\tau},\hat{\theta}_{MD})'W_T f_\tau(\tau_0,\theta_0)(\hat{\tau}-\tau_0)$$

From the continuous mapping theorem and the asymptotic normality of $\hat{\tau}$, $\hat{\tau}_{MD}$ is asymptotically normally distributed with covariance matrix given in (2.48).

3 Here I have to show that

$$(f_\theta' W_T f_\theta)^{-1} f_\theta' W_T f_\tau V(\hat{\tau}) f_\tau' W_T f_\theta (f_\theta' W_T f_\theta)^{-1} - \left(f_\theta'(f_\tau V_\tau f_\tau')^{-1} f_\theta'\right)^{-1}$$

is a positive semidefinite matrix, or, equivalently, that

$$f_\theta'(f_\tau V_\tau f_\tau')^{-1} f_\theta' - f_\theta' W_T f_\theta \left(f_\theta' W_T f_\tau V(\hat{\tau}) f_\tau' W_T f_\theta\right)^{-1} f_\theta' W_T f_\theta$$

To show that, we first observe that this matrix can be written as

$$f_\theta' C' \left(I - (C')^{-1} W_T f_\theta \left(f_\theta' W_T C C' W_T f_\theta \right)^{-1} f_\theta' W_T C \right) C f_\theta \tag{A.9}$$

where C is nonsingular matrix such that $C'C = (f_\tau V_\tau f_\tau')^{-1}$. The matrix inside the parentheses is a projection matrix, hence it is symmetric and idempotent. That the matrix in (A.9) is positive semidefinite then follows from the Lemma below.

Lemma Let B be square symmetric and idempotent matrix. Then ABA' is positive definite.

Proof: To prove the lemma I have to show that for any non-zero vector a, $a'A'BAa \geq 0$. We have

$$a'A'BAa = a'A'B'BAa = (BAa)'(BAa) = \sum_i \tilde{a}_i^2 \geq 0$$

where $\tilde{a} \equiv BAa$.

Appendix B

Appendix for Chapter 3

B.1 Case Study: Identification

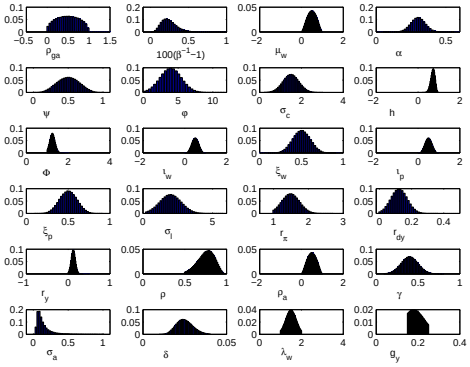

Figure B.1: Distributions of the draws of parameters used in the identification analysis.

Table B.1: Prior Distribution of θ

	Prior		
Parameter	Distr.	Mean	Stdd.
α	$\mathcal{N}$	0.300	0.050
ψ	$\mathcal{B}$	0.500	0.150
φ	$\mathcal{N}$	4.000	1.500
σ_c	$\mathcal{N}$	1.500	0.375
h	$\mathcal{B}$	0.700	0.100
$100(\beta^{-1}-1)$	$\mathcal{G}$	0.250	0.100
Φ	$\mathcal{N}$	1.250	0.125
ι_w	$\mathcal{B}$	0.500	0.150
ξ_w	$\mathcal{B}$	0.500	0.100
ι_p	$\mathcal{B}$	0.500	0.150
ξ_p	$\mathcal{B}$	0.500	0.100
σ_l	$\mathcal{N}$	2.000	0.750
r_π	$\mathcal{N}$	1.500	0.250
$r_{\triangle y}$	$\mathcal{N}$	0.125	0.050
r_y	$\mathcal{N}$	0.125	0.050
ρ	$\mathcal{B}$	0.750	0.100
γ	$\mathcal{N}$	0.400	0.100
δ	$\mathcal{B}$	0.025	0.005
λ_w	$\mathcal{N}$	1.500	0.250
g_y	$\mathcal{N}$	0.180	0.050
ρ_{ga}	$\mathcal{B}$	0.500	0.250
ρ_a	$\mathcal{B}$	0.500	0.200
ρ_b	$\mathcal{B}$	0.500	0.200
ρ_g	$\mathcal{B}$	0.500	0.200
ρ_I	$\mathcal{B}$	0.500	0.200
ρ_r	$\mathcal{B}$	0.500	0.200
ρ_p	$\mathcal{B}$	0.500	0.200
ρ_w	$\mathcal{B}$	0.500	0.200
μ_w	$\mathcal{B}$	0.500	0.200
μ_p	$\mathcal{B}$	0.500	0.200
σ_a	$\mathcal{IG}$	0.100	2.000
σ_b	$\mathcal{IG}$	0.100	2.000
σ_g	$\mathcal{IG}$	0.100	2.000
σ_I	$\mathcal{IG}$	0.100	2.000
σ_r	$\mathcal{IG}$	0.100	2.000
σ_p	$\mathcal{IG}$	0.100	2.000
σ_w	$\mathcal{IG}$	0.100	2.000

Note:$\mathcal{N}$ is Normal distribution, $\mathcal{B}$ is Beta-distribution, $\mathcal{G}$ is Gamma distribution, $\mathcal{IG}$ is Inverse Gamma distribution. The inverse Gamma priors are in the form $p(\sigma;\nu,s) \propto \sigma^{-\nu-1}\exp^{-\nu s^2/2\sigma^2}$; s and ν are given in the Mean column and Stdd. column respectively.

Table B.2: Admissability of draws

Param.	Non-existence	Indeterminacy	Admissible
1	0.30%	3.20%	96.50%
2	0.10%	2.00%	97.90%
3	0.30%	3.10%	96.60%
4	0.20%	3.40%	96.40%
5	0.10%	4.10%	95.80%
6	0.20%	2.40%	97.40%

Note: The total number of draws is $1,000,000$.

Table B.3: Conditioning of H for different parameterizations.

Param.	Decile of cond(H)										
	min	1	2	3	4	5	6	7	8	9	max
1	6.4e1	2.2e2	2.9e2	3.7e2	4.7e2	6.0e2	7.9e2	1.1e3	1.6e3	3.2e3	3.1e11
2	4.8e1	2.0e2	2.8e2	3.6e2	4.5e2	5.8e2	7.6e2	1.0e3	1.6e3	3.1e3	2.9e11
3	4.3e1	1.5e2	1.9e2	2.3e2	2.8e2	3.4e2	4.2e2	5.4e2	7.3e2	1.2e3	2.8e8
4	6.4e1	2.1e2	2.8e2	3.6e2	4.6e2	5.9e2	7.7e2	1.1e3	1.6e3	3.1e3	3.0e11
5	7.0e1	2.8e2	3.9e2	4.9e2	6.1e2	7.4e2	9.1e2	1.1e3	1.5e3	2.1e3	2.8e8
6	3.4e1	1.3e2	1.7e2	2.1e2	2.5e2	3.1e2	3.8e2	4.9e2	6.6e2	1.1e3	2.8e8

Note: $H = \frac{\partial \tau}{\partial \theta'}$ is the gradient of the reduced-form parameters w.r.t. θ. $\text{rank}(H) = \dim(\theta)$ is a necessary condition for identification of θ. Large values of $\text{cond}(H)$ imply near failure of this condition, thus indicating weak identification. The statistics were computed on the basis of $1,000,000$ random draws of θ.

Table B.4: Conditioning of $H'H$ for different parameterizations.

Param.	Decile of cond($H'H$)										
	min	1	2	3	4	5	6	7	8	9	max
1	4.1e3	4.8e4	8.6e4	1.4e5	2.2e5	3.7e5	6.2e5	1.2e6	2.6e6	1.0e7	9.5e22
2	2.3e3	4.1e4	7.6e4	1.3e5	2.1e5	3.4e5	5.8e5	1.1e6	2.4e6	9.4e6	8.5e22
3	1.8e3	2.2e4	3.6e4	5.5e4	8.0e4	1.2e5	1.8e5	2.9e5	5.3e5	1.4e6	7.6e15
4	4.1e3	4.5e4	8.1e4	1.3e5	2.1e5	3.5e5	6.0e5	1.1e6	2.5e6	9.6e6	9.0e22
5	4.9e3	8.0e4	1.5e5	2.4e5	3.7e5	5.5e5	8.3e5	1.3e6	2.2e6	4.5e6	7.6e15
6	1.2e3	1.6e4	2.7e4	4.2e4	6.3e4	9.4e4	1.4e5	2.4e5	4.4e5	1.1e6	7.6e15

Note: $\text{cond}(\mathfrak{I}_\theta) = \text{cond}(H'H)$ if $\mathfrak{I}_\tau$ is perfectly well conditioned. Thus $\text{cond}(H'H)$ can be thought of as the unattainable lower bound for $\text{cond}(\mathfrak{I}_\theta)$.

Table B.5: Conditioning of $\mathfrak{I}_\theta$ for different parameterizations

Param.	Decile of cond($\mathfrak{I}_\theta$)										
	min	1	2	3	4	5	6	7	8	9	max
1	4.2e5	1.6e6	2.1e6	4.9e6	8.1e6	1.8e7	5.0e7	6.4e8	2.3e9	2.2e10	4.4e24
2	2.7e5	4.7e5	1.3e6	2.9e6	3.3e6	3.7e6	4.9e7	3.5e8	2.2e9	2.1e10	4.1e25
3	1.8e5	1.6e6	1.9e6	2.6e6	4.5e6	1.2e7	1.6e7	4.4e8	1.1e9	2.2e10	1.8e14
4	4.1e5	1.4e6	2.1e6	4.6e6	7.1e6	1.8e7	4.9e7	6.1e8	2.3e9	2.2e10	2.8e24
5	4.3e5	7.6e5	1.3e6	1.5e6	1.8e6	2.0e6	1.5e7	2.8e8	1.0e9	2.1e10	1.6e14
6	1.0e5	4.2e5	1.1e6	1.5e6	1.8e6	2.0e6	1.4e7	1.9e8	1.0e9	2.1e10	1.6e14

Note: $\mathfrak{I}_\theta = H'\mathfrak{I}_\tau H$ is the information matrix for θ. $\text{rank}(\mathfrak{I}_\theta) = \dim(\theta)$ is a necessary and sufficient condition for identification of θ. Large values of $\text{cond}(\mathfrak{I}_\theta)$ imply near failure of this condition, thus indicating weak identification. These statistics were computed on the basis of $1,000$ random draws of θ.

Table B.6: Cross-correlations

	λ_w	β	μ_p	ψ	σ_c	h	Φ	ξ_w	σ_l	r_π	ρ	ρ_b	ρ_I
α	.77	**.98**	-.54	-.88	.82	-.84	-.75	-.82	**-.92**	**.94**	.87	-.74	.89
ψ	**-.97**	**-.94**	.85	1	**-.98**	**.99**	**.97**	**.98**	**.93**	**-.97**	**-.97**	.59	**-.94**
σ_c	**.99**	.89	-.87	**-.98**	1	-.99	**-.95**	**-.99**	**-.93**	**.96**	**.98**	-.56	**.94**
h	**-.99**	**-.91**	.84	.99	**-.99**	1	**.94**	**.99**	**.95**	**-.97**	**-.97**	.59	**-.93**
ξ_w	**-.99**	-.89	.86	**.98**	**-.99**	**.99**	**.95**	1	**.93**	**-.96**	**-.97**	.58	**-.93**
ξ_p	**.96**	.87	-.85	**-.97**	**.95**	**-.95**	**-.99**	**-.95**	-.83	**.90**	**.92**	-.55	.87
r_π	**.93**	**.97**	-.75	**-.97**	**.96**	**-.97**	-.89	**-.96**	**-.97**	1	**.98**	-.67	**.97**
ρ	**.97**	**.92**	-.84	**-.97**	**.98**	**-.97**	**-.92**	**-.97**	**-.93**	**.98**	1	-.58	**.97**
ρ_I	**.92**	**.92**	-.77	**-.94**	**.94**	**-.93**	-.86	**-.93**	**-.92**	**.97**	**.97**	-.57	1
σ_b	.50	.69	-.24	-.57	.54	-.57	-.48	-.56	-.66	.64	.56	**-.99**	.55
σ_I	**-.91**	**-.94**	.73	.94	**-.94**	**.94**	.84	**.94**	**.97**	**-.99**	**-.97**	.63	**-.99**
σ_p	**-.95**	-.71	**.98**	**.90**	**-.93**	**.90**	**.92**	**.93**	.77	-.82	-.89	.35	-.82
σ_w	**-.99**	-.86	**.90**	**.97**	**-.99**	**.98**	**.94**	**.99**	**.92**	**-.94**	**-.97**	.53	**-.92**

Note: Pairwise correlation coefficients $corr(\hat{\theta}_i, \hat{\theta}_j)$ exceeding .95 in absolute value. The values are obtained by inverting and normalizing the information matrix evaluated at θ for which the condition number of the matrix is equal to the 7-th percentile from Table B.3. High correlation between the estimates of two deep parameters indicates that they are difficult to identify.

Table B.7: Percent error in diag($V(\hat{\theta})$) for 1% error in diag($\mathfrak{I}_\theta(\hat{\theta})$)

Param.	Decile of $\mathfrak{I}_\theta$									
	min	1	2	3	4	5	6	7	8	9
δ	93.6	202.2	-136.8	-294.9	355.7	-96.8	266.6	2245.6	49.3	-236.7
λ_w	102.6	218.2	97.3	211.3	-132.5	-173.0	-265.0	-177.6	62.4	-56.2
g_y	0.4	9.1	-495.3	-56.6	49.6	-587.8	159.5	-23.8	1308.6	1530.8
ρ_{ga}	-19.7	-123.8	-57.8	-34.1	-17.1	22.6	9.2	-16.9	-32.8	-12.3
$\bar{\beta}$	65.5	54.9	-138.4	-736.3	68.6	-225.4	90.2	569.2	-58.5	-89.8
μ_w	-11.3	28.5	2.8	6.7	23.7	-7.5	-25.1	-23.2	-34.7	-324.4
μ_p	18.8	89.3	114.5	-539.5	7.1	-76.6	-68.8	-14.0	128.0	-119.6
α	-39.2	-369.0	69.1	-135.4	-31.2	-116.3	81.2	-27.5	-141.0	-234.6
ψ	52.8	54.0	-64.4	-75.5	-46.8	-13.5	203.6	-30.3	-1471.1	521.7
φ	66.7	65.9	56.0	-98.7	-35.9	-633.6	-164.0	-992.9	291.9	-1407.6
σ_c	-40.2	-47.7	107.7	171.4	-1720.9	-176.7	127.8	203.4	199.3	156.9
λ	163.7	-60.5	-42.8	83.9	36.6	1920.0	179.5	61.9	107.0	-136.9
Φ	160.5	-63.2	388.8	509.4	-119.7	1251.6	2346.8	185.1	-113.1	144.4
ι_w	-9.3	6.8	-382.7	-231.4	-34.5	654.0	-123.2	-361.7	109.0	-112.5
ξ_w	319.4	153.1	-1231.1	187.7	59.1	159.9	104.6	327.6	310.5	-150.1
ι_p	99.3	309.5	178.8	549.4	612.1	57.7	-180.6	24.1	-88.5	-69.0
ξ_p	-67.0	59.4	-122.4	172.9	89.2	78.1	-68.2	-78.9	-81.0	-76.3
σ_l	-144.1	134.3	-128.9	1450.5	73.7	30.9	-241.8	41.5	-114.9	123.9
r_π	77.6	-139.4	337.8	102.2	-61.3	-872.6	138.7	-256.9	-9506.0	-4013.2
$r_{\triangle y}$	-118.2	32.0	-48.8	24.9	-86.9	-24.1	-171.3	-4639.8	72.8	-38.5
r_y	71.7	-50.7	143.9	41.8	-216.4	-98.2	84.0	113.2	-198.2	-167.6
ρ	-70.4	108.5	396.0	625.5	-121.6	289.8	-2027.8	97.0	-149.8	-408.1
ρ_a	-36.7	-22.4	772.4	461.4	0.4	-32.5	8.9	-12.5	-29.7	-19.4
ρ_b	-0.8	90.4	-69.7	-233.8	-204.0	-37.4	-90.3	957.3	118.8	179.8
ρ_g	0.8	1.4	0.8	14.5	-4.3	-4.7	38.5	-4.3	38.4	-20.2
ρ_I	-0.6	-87.2	-38.1	2.8	1.4	-179.7	214.4	43.1	12.3	-35.1
ρ_r	6.4	-182.2	5.5	-189.0	1306.1	-78.6	-183.6	-86.1	68.0	11.1
ρ_p	-5.1	1.8	-4.2	748.3	-8.4	-2.9	-227.3	-7.2	-7.7	-21.0
ρ_w	-1.2	-0.4	58.0	-9.5	3.4	2.8	-2.8	63.1	-2.3	136.8
σ_a	-88.9	-70.7	79.8	101.0	39.8	119.0	67.2	17.2	999.0	-348.2
σ_b	-5.9	-70.8	149.8	-148.7	76.2	60.1	268.5	-34.9	173.3	107.3
σ_g	-63.9	14.3	7.5	13.0	34.5	-24.8	-1221.2	14.2	-27.0	-28.7
σ_I	1.0	-15.1	-58.5	-11.4	-27.8	34.8	-234.5	35.2	-14.5	10550.7
σ_r	-167.0	-41.1	-3.4	-191.1	-57.3	1022.8	154.5	-162.7	-646.0	-656.9
σ_p	19.7	50.0	167.8	-227.5	-14.6	117.5	173.0	12.5	42.2	25.4
σ_w	-0.2	-8.6	7.6	-168.3	-12.0	-24.0	-256.9	128.7	25.6	-44.8
cond($\mathfrak{I}_\theta$)	4.2e5	1.6e6	2.1e6	4.9e6	8.1e6	1.8e7	5.0e7	6.4e8	2.3e9	2.2e10

Table B.8: Parameter Importance

Parameter		$s_i(\theta)$										
		Decile of cond$(H'H)$										
		min	1	2	3	4	5	6	7	8	9	10
ρ_w	autocorr. wage shock	.104	.086	.082	.071	.082	.081	.134	.074	.087	2.207	.188
ρ	policy smoothing	.141	.140	.831	1.217	.540	.727	.324	2.305	.437	1.919	2.200
ρ_p	autocorr. price shock	.136	.355	.109	.125	.127	.151	.123	.100	.155	.087	.219
ξ_w	Calvo wages	.065	.120	.227	.180	.252	.361	.189	.240	.316	.980	.317
ξ_p	Calvo prices	.132	.319	.224	.186	.178	.290	.166	.198	.246	.300	.308
r_π	policy inflation	.156	.192	.310	.202	.143	.594	.265	.614	.263	.605	.533
λ_w	wage markup	.038	.074	.162	.146	.181	.288	.140	.184	.264	.508	.218
λ	habit	.098	.434	.673	.362	1.397	1.376	1.639	.715	3.432	.461	1.808
Φ	fixed cost	.145	.338	.434	.298	.187	.459	.300	.217	.378	.551	.357
ρ_g	autocorr. gov. spending	.062	.124	.110	.083	.164	.159	.193	.107	.113	.146	.090
ρ_a	autocorr. TFP	.118	.171	.140	.192	.225	.170	.197	.186	.234	.138	.166
σ_c	elast.inter.subst.	.201	.227	.380	.151	.414	.452	.406	.274	.529	.359	.522
$r_{\triangle y}$	policy output growth	.026	.032	.068	.069	.047	.076	.036	.066	.046	.175	.162
ρ_I	autocorr. investment	.105	.104	.179	.126	.147	.081	.190	.117	.105	.124	.077
μ_w	MA wage shock	.066	.067	.058	.057	.058	.061	.079	.057	.062	.170	.089
μ_p	MA price shock	.079	.140	.062	.073	.071	.077	.072	.067	.075	.055	.113
ψ	cap. utilization cost	.077	.122	.183	.136	.127	.217	.126	.128	.170	.121	.222
α	capital share	.035	.055	.066	.077	.063	.079	.083	.079	.098	.083	.162
φ	invest. adj. cost	.067	.085	.113	.336	.079	.226	.155	.299	.339	.074	2.410
σ_l	elast. hours	.027	.028	.062	.045	.047	.064	.046	.054	.054	.080	.053
ι_w	indexation wages	.021	.030	.038	.026	.022	.060	.028	.044	.032	.073	.043
δ	depreciation rate	.006	.012	.014	.020	.015	.019	.019	.016	.030	.034	.026
ι_p	indexation prices	.016	.022	.028	.024	.018	.040	.020	.033	.021	.040	.046
r_y	policy output	.010	.017	.024	.022	.014	.039	.018	.030	.018	.103	.064
ρ_b	autocorr. risk premium	.032	.120	.361	.112	.241	.560	.279	.190	.477	.116	.356
σ_b	std. dev. risk premium	.025	.082	.120	.110	.162	.198	.200	.197	.335	.117	.220
σ_I	std. dev. investment	.050	.053	.068	.052	.064	.046	.071	.053	.055	.061	.040
σ_a	std. dev. TFP	.046	.071	.050	.081	.101	.053	.095	.105	.093	.050	.069
g_y	G/Y	.018	.027	.047	.024	.030	.048	.031	.027	.037	.030	.021
σ_g	std. dev. gov. spending	.036	.079	.068	.068	.099	.097	.120	.078	.101	.094	.053
σ_r	std. dev. policy	.029	.019	.047	.074	.042	.044	.039	.115	.030	.056	.122
σ_w	std. dev. wages	.015	.016	.015	.015	.015	.018	.016	.015	.018	.017	.022
σ_p	std. dev. prices	.011	.016	.009	.009	.011	.013	.012	.012	.010	.010	.010
ρ_r	autocorr. policy	.028	.013	.045	.085	.036	.052	.045	.203	.021	.056	.217
β	discount factor	.002	.002	.004	.004	.003	.003	.004	.004	.005	.004	.011

Note: The table shows the values of the sensitivity statistic $s_i(\theta) = \sqrt{\frac{1}{J}\sum_j^J \left(\theta_i \frac{\partial \tau_j}{\partial \theta_i}\right)^2}$ evaluated at values of θ corresponding to the minimum and the deciles of the distribution of cond$(H'H)$ computed on the basis of 1 million draws from Θ.

Table B.9: Maximum pairwise correlations

	Decile of cond($H'H$)										
	min	1	2	3	4	5	6	7	8	9	10
λ_w	.767	.943	.938	.959	.945	.941	.973	.908	.932	.984	.998
ξ_w	.767	.943	.938	.959	.945	.941	.973	.908	.932	.984	.998
ξ_p	.862	.910	.635	.860	.825	.878	.629	.650	.870	.815	.932
ι_w	.361	.661	.661	.775	.833	.841	.667	.586	.748	.625	-.540
ι_p	.656	.833	.635	.860	.825	.878	.629	.616	.870	.815	.932
μ_w	.502	.213	.255	.218	.225	.071	.236	.284	.114	.179	.611
μ_p	.570	.643	.353	.324	.182	.150	.341	.190	.406	.341	.122
α	-.585	-.698	-.854	.580	-.853	-.769	-.755	-.802	-.833	-.976	-.650
ψ	.117	.157	.065	.165	.100	.087	-.072	-.179	-.085	-.143	-.367
φ	-.291	.349	-.475	.477	-.556	-.404	-.385	.395	-.559	.718	.572
σ_c	.514	.940	.869	.956	.828	.982	.953	.943	.966	.905	.806
λ	.534	.940	.869	.956	.768	.982	.953	.943	.966	.955	.674
Φ	.862	.910	.554	.642	.513	.549	.485	.650	.561	-.607	.848
σ_l	.514	.589	-.716	-.707	.828	-.783	-.610	-.622	-.695	-.867	-.888
r_π	-.614	-.735	-.970	.923	-.989	.968	-.866	-.719	-.781	-.982	-.951
$r_{\triangle y}$	.358	.646	.588	.661	-.502	.725	.454	.440	.462	.881	.959
r_y	-.614	-.703	-.915	.923	-.959	.968	.767	-.698	-.794	.881	.959
ρ	.530	-.735	-.970	-.861	-.989	-.874	-.866	-.719	-.766	-.982	-.951
δ	.256	-.655	-.708	-.708	-.653	.408	-.680	-.618	-.722	.809	-.728
g_y	-.664	-.470	.337	-.570	.530	.431	-.415	.396	.448	-.464	-.481
β	-.664	-.698	-.854	-.708	-.853	-.769	-.755	-.802	-.833	-.976	-.650
ρ_a	.184	.399	.322	.220	.411	.418	.464	.402	.440	.161	.219
ρ_b	.534	.790	.561	.909	.666	-.957	.821	.900	.951	.955	.650
ρ_g	.221	.293	-.165	.168	.259	.173	.227	.270	-.238	.140	.244
ρ_I	.151	.238	.302	.303	.129	.123	.304	.395	.137	.086	-.222
ρ_r	.530	.115	-.781	.424	-.941	.441	.413	.674	.154	-.934	.298
ρ_p	.503	.893	.420	.323	.182	.311	.391	.270	.501	.418	.151
ρ_w	.524	.213	.255	.218	.225	.071	.236	.284	.114	.301	.974
ρ_{ga}	-.186	-.741	-.648	-.142	-.854	.418	-.856	-.882	-.788	-.565	-.356
σ_a	-.186	-.741	-.648	-.142	-.854	-.279	-.856	-.882	-.788	-.565	-.356
σ_b	.085	.486	.664	.144	.508	.083	.692	.414	.168	.253	.384
σ_g	.044	.062	-.123	.044	-.067	.105	.174	.163	-.159	.019	.049
σ_I	.049	.191	.263	.208	.031	.123	.087	.157	.043	.006	-.136
σ_r	.183	.056	-.125	.196	-.276	.073	.071	.090	.163	-.256	.098
σ_p	-.042	-.055	-.227	-.034	-.067	-.093	-.055	-.075	-.094	-.302	-.083
σ_w	-.021	-.149	-.017	-.024	-.036	-.031	-.020	-.172	-.045	-.050	-.022
cond	170	1346	1458	2807	8234	8413	9067	10188	20638	48567	305217

Note: The table shows for each deep parameter θ_i the value of $\max_{j \neq i}\left(\frac{H_i'H_j}{(H_i'H_i)^{1/2}(H_j'H_j)^{1/2}}\right)$, where $H_l = \frac{\partial \tau}{\partial \theta_l}$ gives the effect on the reduced-form model of changes in θ_l. Values close to 1 or -1 indicate that H_i and H_k are nearly collinear.

Table B.10: Multiple correlations

	Decile of cond($H'H$)										
	min	1	2	3	4	5	6	7	8	9	10
λ_w	.975	.992	.996	.997	.999	.995	.999	.998	.999	.999	.999
ξ_w	.968	.991	.996	.997	.999	.997	.999	.998	.999	.999	.999
ξ_p	.959	.991	.936	.968	.955	.978	.945	.954	.957	.962	.992
ι_w	.638	.764	.832	.939	.954	.971	.844	.875	.896	.857	.903
ι_p	.715	.894	.824	.934	.927	.966	.802	.818	.916	.928	.973
μ_w	.669	.411	.489	.433	.473	.156	.566	.648	.241	.345	.837
μ_p	.778	.894	.682	.653	.406	.398	.585	.426	.657	.767	.310
α	.770	.860	.943	.928	.941	.934	.903	.920	.926	.989	.916
ψ	.253	.328	.190	.262	.246	.201	.181	.249	.188	.230	.393
φ	.656	.729	.849	.868	.892	.928	.767	.842	.832	.957	.799
σ_c	.969	.997	.996	.998	.999	.999	.999	.999	.999	.999	.999
λ	.947	.995	.991	.997	.995	.999	.999	.999	.999	.999	.990
Φ	.939	.982	.922	.947	.921	.944	.936	.960	.932	.899	.994
σ_l	.901	.952	.971	.938	.992	.975	.987	.986	.989	.981	.997
r_π	.967	.962	.990	.994	.998	.995	.979	.910	.977	.999	.999
$r_{\triangle y}$	.422	.717	.625	.728	.582	.840	.504	.521	.563	.913	.969
r_y	.959	.945	.972	.975	.972	.988	.908	.910	.957	.990	.998
ρ	.873	.839	.992	.984	.997	.977	.966	.879	.933	.998	.997
δ	.508	.862	.922	.805	.904	.881	.865	.862	.967	.892	.968
g_y	.844	.902	.789	.927	.819	.874	.915	.911	.797	.746	.941
β	.827	.899	.956	.936	.958	.936	.923	.942	.971	.989	.912
ρ_a	.621	.770	.748	.672	.788	.718	.851	.829	.829	.655	.651
ρ_b	.737	.957	.831	.995	.922	.997	.982	.984	.996	.997	.856
ρ_g	.500	.697	.529	.538	.662	.526	.646	.712	.630	.421	.774
ρ_I	.340	.344	.559	.587	.362	.174	.483	.649	.335	.394	.398
ρ_r	.846	.169	.898	.663	.984	.769	.567	.751	.219	.994	.767
ρ_p	.697	.972	.656	.647	.354	.784	.660	.589	.782	.794	.487
ρ_w	.572	.290	.278	.374	.296	.264	.456	.642	.358	.740	.998
ρ_{ga}	.310	.750	.653	.283	.858	.606	.865	.888	.792	.574	.562
σ_a	.350	.746	.664	.214	.861	.338	.863	.884	.795	.598	.377
σ_b	.123	.804	.876	.422	.870	.356	.957	.867	.767	.874	.677
σ_g	.145	.225	.190	.168	.202	.354	.552	.543	.282	.068	.275
σ_I	.069	.262	.497	.323	.049	.208	.124	.272	.101	.031	.262
σ_r	.364	.079	.189	.307	.529	.115	.095	.098	.238	.675	.209
σ_p	.102	.157	.414	.075	.145	.215	.125	.175	.188	.484	.177
σ_w	.047	.245	.030	.046	.067	.064	.042	.277	.092	.094	.038
cond	170	1346	1458	2807	8234	8413	9067	10188	20638	48567	305217

Note: The table shows for each deep parameter θ_i the value of $\left(\frac{H_i'H_{-j}}{(H_i'H_i)^{1/2}(H_{-j}'H_{-j})^{1/2}}\right)$, where $H_i = \frac{\partial \tau}{\partial \theta_i}$ gives the effect on the reduced-form model of changes in θ_i, and H_{-i} is the projection of H_i onto the space spanned by the other columns of H. Values close to 1 or -1 indicate that H_i and H_{-i} are nearly collinear.

B.2 Identification and weak identification analysis procedure

Here I outline the steps involved in the identification analysis performed in Chapter 4.

Steps:

1 Define a discrete approximation $\hat{\Theta}$ of the the parameter space Θ. The parameter space is (usually a continuous) set of values that are possible, from a theoretical point of view, for the deep parameters to take. Typically, for each deep parameter there is an open or closed interval such that $-\infty \leq \theta_i^{min} \leq \theta_i \leq \theta_i^{max} \leq \infty$. In $\hat{\Theta}$ these intervals are approximated by a grid with a finite number of points. In addition, $\hat{\Theta}$ is constrained to includes only points θ_i for which the linearized model has a unique solution. Lastly, the grid should be finer for regions in the parameter space that are considered as a priori more likely.

2 Evaluate rank(H), cond(H) and the singular value decomposition (SVD) $H = LQR$ of $H(\theta)$ at each $\theta_i \in \hat{\Theta}$. Since $H(\theta)$ depends only on the structure of the linearized model and the value of θ

(a) if rank$(H) <$ dim(θ), some deep parameter or parameters cannot be identified because they have no effect in the model, or their effect cannot be distinguished from that of other deep parameters. In the first case one or more columns of H are zeros, which implies that the corresponding elements of θ are unidentifiable. In the second case there is one or more sets of columns of H that are exactly linearly dependent. The number of such sets is equal to the number of singular values of H that are equal to zero. The columns of H that belong to each such set can be established by identifying the non-zero elements of L_i - the i-th column of matrix L from the SVD of H that correspond to a singular value $Q_i = 0$.

(b) if rank$(H) =$ dim(θ), but $cond(H) >> 1$, then some deep parameter or parameters are not well identified either because they have a very small effect in the model, or their effect cannot be easily distinguished from that of other deep parameters. To find out which parameters are involved

- compute the sensitivity $s_i(\theta)$ using (4.47) for each column i of H. Small values of $s_i(\theta)$ imply that the parameter θ_i has only a marginal effect in the model.

- compute the pairwise and multiple collinearity measures (4.48). Values close to one imply that the role of parameter θ_i in the model is very well approximated by a combination of the other deep parameters.

3 Evaluate the rank and the condition number of $\mathcal{I}_\theta$. In general DSGE models, some of the state variables are unobserved, and because of that some reduced-form parameters τ may be unidentifiable. By verifying that $\mathcal{I}_\theta$ has full rank, we make sure that θ can be identified from the identifiable parameters or combinations of parameters in τ. This is usually the case since typically $\dim(\theta) << \dim(\tau)$, and $\mathcal{I}_\theta = H(\theta)\mathcal{I}_\tau H(\theta)'$ has full rank even though $\mathcal{I}_\tau$ does not.

The multiple collinearity coefficient (4.48), computed in step 2(b) measures the severity of the parameter interdependence problem for each deep parameter θ_i. It is the absolute value of the cosine of the angle between $H_i(\theta) = \frac{\partial \tau}{\partial \theta_i}$ and the linear space spanned by the other columns of $H(\theta)$. A large value implies that locally the effect of θ_i on τ can be very well approximated by the effect of a combination of all other deep parameters. However, not all other deep parameters are equally important in this approximation. In fact 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.

The problem of selecting the important deep parameters is similar to that of selecting a parsimonious set of regressors, in the linear regression framework. The motivation, however, is different. In linear regression the goal is to improve the precision of the parameter estimates. Our goal is to find out what feature of the model are closely related and therefore difficult to distinguish. Specifically, we want to select a small set J_i of parameters $\theta_{j \neq i}$ such that adding one or more of the other parameters $\theta_l \notin J_i$ leads to only a small increase in the multiple collinearity coefficient, and replacing some $\theta_j \in J_i$ by $\theta_l \notin J_i$ would lead to a significantly lower value of the multiple collinearity coefficient. To select candidates for J_i we can use the naive elastic net algorithm proposed by Zou and Hastie (2005). The algorithm finds a vector $\mathbf{a}_i$ that minimizes the following function

$$\|H_i - H_{-i}\mathbf{a}_i\|_2 + \lambda_1\|\mathbf{a}_i\|_1 + \lambda_2\|\mathbf{a}_i\|_2$$

where H_i is the i-th column of H, and H_{-i} is H with the i-th column deleted. Our interest is in finding the non-zero entries in the solution $\mathbf{a}_i$. Their number increases with the value of λ_1. A positive value of λ_2, on the other hand, instructs the algorithm to keep in J_i all columns of H_{-i} that are important in approximating H_i, even is they exhibit strong pairwise collinearity, and therefore have only a marginal contribution in the approximation.[1]

[1]More precisely, we need $\frac{\lambda_2}{\lambda_1+\lambda_2} > 0$ for the regression coefficients of highly collinear regressors to be similar, see Zou and Hastie (2005, Theorem 2)

B.3 Identification: Three Alternative Models

The three models we consider are: a simple New Keynesian model (see An and Schorfheide (2005) for details), a simple real business cycle (RBC) model (see Chang, Doh, and Schorfheide (2007) for details), and a two-country Open Economy model (see Lubik and Schorfheide (2005) for details). The New Keynesian model has nominal rigidities only in prices, and no capital accumulation. It has 11 deep parameters, and 3 structural shocks - productivity, government consumption, and monetary policy. The RBC is a standard stochastic growth model with 10 deep parameters and 2 stochastic shocks - productivity and labor supply. The New Open Economy model is a two-country version of the New Keynesian model with nominal rigidities in domestic and import prices. It has 32 deep parameters, and 8 structural shocks - country-specific productivity, government consumption, and monetary policy for both countries, a world-wide technology shock, and a shock capturing deviations from the purchasing power parity.

Table B.11: Conditioning of H for 3 different DSGE models

model		Decile										
	$dim(\theta)$	min	1	2	3	4	5	6	7	8	9	10
New Keynesian	11	8.2	22.8	28.1	33.2	38.7	45.1	52.7	62.7	77.1	103.4	8.7e2
RBC	10	4.9	17.5	23.7	30.1	36.7	44.1	53.5	68.1	95.7	177.7	4.0e8
NOE	32	12.3	41.3	52.2	62.9	75.0	89.6	108.9	137.0	185.4	303.5	1.1e10

Note:$H = \frac{\partial \tau}{\partial \theta}$ is the gradient of the reduced-form parameters w.r.t. θ. $\text{rank}(H) = \dim(\theta)$ is a necessary condition for identification of θ. Large values of $\text{cond}(H)$ imply near failure of this condition, thus indicating weak identification. The statistics were computed on the basis of $1,000,000$ random draws of θ.

B.4 Monte Carlo Study: Small Example

Structural Model

$$\Gamma_0 y_t = \Gamma_1 E_t y_{t+1} + \Gamma_1 y_{t-1} + \Gamma_3 u_t, \tag{B.1}$$

where y is univariate and

$$\Gamma_0 = (1+\delta), \quad \Gamma_1 = (1+\gamma+\gamma^2/2), \quad \Gamma_2 = (\delta-\gamma-\gamma^2/2), \quad \Gamma_3 = e^\gamma$$

Parameters: δ, γ.

The **reduced form** solution is:

$$y_t = Ay_{t-1} + Be_t \tag{B.2}$$

where A and B can be calculated by hand:

$$A = \frac{2\delta - 2\gamma + \gamma^2}{2+2\gamma+\gamma^2}, \quad B = \frac{2e^\gamma}{2+2\gamma+\gamma^2}$$

Identification problems δ and γ are difficult to identify separately when $\gamma \approx 0$. One way to see that is by computing H given by

$$H = \begin{bmatrix} \frac{\partial A}{\partial \delta} & \frac{\partial A}{\partial \gamma} \\ \frac{\partial B}{\partial \delta} & \frac{\partial B}{\partial \gamma} \end{bmatrix} = \begin{bmatrix} 2/(2+2\gamma+\gamma^2) & -4(1+\gamma)(1+\delta)/(2+2\gamma+\gamma^2)^2 \\ 0 & 2e^\gamma\gamma^2/(2+2\gamma+\gamma^2)^2 \end{bmatrix}$$

When $\gamma \approx 0$ the columns of H are almost collinear, which implies that, locally, the effect on A and B of perturbing δ is very similar to that of perturbing γ. Since the likelihood function depends on the parameters only through A and B, this implies that they are poorly identified for $\gamma \approx 0$. For instance, if $\delta = .25$ and $\gamma = .01$, the *condition number* of H is 51247. If $\delta = 3.6$ and $\gamma = 1.4$, on the other hand, the *condition number* of H is 11.

We can also see why the problem arises directly, by realizing that δ and γ only enter the likelihood function as either $f = \frac{1+\gamma+\gamma^2/2}{1+\delta}$ or $g = \frac{e^\gamma}{1+\delta}$ (we can write $A = \frac{1-f}{f}$, and $B = \frac{g}{f}$). When $\gamma \approx 0$, f and g are very similar, which make it difficult to separate δ from γ.

Table B.12: Condition number and finite sample properties of MLE: Example

	Relative Bias					Relative MSE				
Parameter	1	2	3	4	5	1	2	3	4	5
δ	-0.3	0.6	1.0	1.0	1.1	1.0	2.7	3.2	3.3	3.5
γ	-0.5	-0.6	1.4	12.5	68.7	0.9	3.9	37.8	376.2	766.8
cond(H)	2.6e1	5.1e2	5.1e4	5.1e6	2.0e7	2.6e1	5.1e2	5.1e4	5.1e6	2.0e7

Note: Results from Monte Carlo study with 1000 repetitions.

B.5 Estimation

Table B.13: Parameter Bounds

Parameter	lower bounds	upper bounds
φ	2.000	15.000
σ_c	0.250	3.000
λ	0.001	0.990
ξ_w	0.300	0.950
σ_l	0.250	10.000
ξ_p	0.500	0.950
ι_w	0.010	0.990
ι_p	0.010	0.990
ψ	0.010	1.000
Φ	1.000	3.000
r_π	1.000	6.000
ρ	0.500	0.975
r_y	0.001	0.500
$r_{\triangle y}$	0.001	0.500
$100(\beta^{-1}-1)$	0.010	2.000
α	0.010	1.000
δ	0.010	0.400
λ_w	1.000	2.000
ρ_a	0.010	1.000
ρ_b	0.010	1.000
ρ_g	0.010	1.000
ρ_I	0.010	1.000
ρ_r	0.010	1.000
ρ_p	0.010	1.000
ρ_w	0.001	1.000
ρ_{ga}	0.010	2.000
μ_w	0.010	1.000
μ_p	0.010	1.000
σ_a	0.010	3.000
σ_b	0.025	5.000
σ_g	0.010	3.000
σ_I	0.010	3.000
σ_r	0.010	3.000
σ_p	0.010	3.000
σ_w	0.010	3.000

Note:Taken from Smets and Wouters (2007)

B.5.1 Estimation: Restricted Model ($\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$)

Table B.14: Estimation Results:1966:1-1979:2

Param.	Prior			Bayesian			MLE		
	Distr.	Mean	St.dev.	5%	mode	95%	5%	mode	95%
φ	$\mathcal{N}$	4.00	1.50	1.93	3.62	5.31	1.02	2.12	5.83
σ_c	$\mathcal{N}$	1.50	0.38	1.03	1.39	1.75	0.73	1.21	2.29
λ	$\mathcal{B}$	0.70	0.10	0.52	0.63	0.75	0.28	0.53	0.75
ξ_w	$\mathcal{B}$	0.50	0.10	0.54	0.66	0.77	0.50	0.73	0.94
σ_l	$\mathcal{N}$	2.00	0.75	0.45	1.52	2.59	-0.21	1.55	3.85
ξ_p	$\mathcal{B}$	0.50	0.10	0.42	0.56	0.69	0.43	0.63	0.87
ι_w	$\mathcal{B}$	0.50	0.15	0.37	0.59	0.80	0.35	0.86	1.41
ι_p	$\mathcal{B}$	0.50	0.15	0.16	0.46	0.75	-0.16	0.25	0.80
ψ	$\mathcal{B}$	0.50	0.15	0.13	0.35	0.56	0.00	0.16	0.71
Φ	$\mathcal{N}$	1.25	0.12	1.29	1.43	1.58	1.05	1.38	1.71
r_π	$\mathcal{N}$	1.50	0.25	1.35	1.66	1.97	3.00	3.00	3.00
ρ	$\mathcal{B}$	0.75	0.10	0.76	0.81	0.86	0.81	0.91	0.96
r_y	$\mathcal{N}$	0.12	0.05	0.13	0.18	0.22	0.25	0.40	0.60
$r_{\triangle y}$	$\mathcal{N}$	0.12	0.05	0.16	0.21	0.26	0.18	0.27	0.40
$100(\frac{1}{\beta} - 1)$	$\mathcal{G}$	0.25	0.10	0.05	0.15	0.24	0.01	0.01	0.01
α	$\mathcal{N}$	0.30	0.05	0.16	0.20	0.23	0.07	0.15	0.24
ρ_a	$\mathcal{B}$	0.50	0.20	0.96	0.97	0.99	0.94	0.99	1.00
ρ_b	$\mathcal{B}$	0.50	0.20	0.12	0.40	0.68	0.22	0.60	0.92
ρ_g	$\mathcal{B}$	0.50	0.20	0.86	0.91	0.96	0.74	0.91	1.00
ρ_I	$\mathcal{B}$	0.50	0.20	0.44	0.61	0.77	0.17	0.47	0.97
ρ_r	$\mathcal{B}$	0.50	0.20	0.17	0.22	0.27	-0.34	0.07	0.48
ρ_p	$\mathcal{B}$	0.50	0.20	0.12	0.51	0.90	0.26	0.82	1.00
ρ_w	$\mathcal{B}$	0.50	0.20	0.93	0.97	1.00	0.96	1.00	1.00
μ_w	$\mathcal{B}$	0.50	0.20	0.73	0.85	0.96	0.87	0.97	1.00
μ_p	$\mathcal{B}$	0.50	0.20	0.13	0.46	0.79	0.34	0.98	1.17
σ_a	$\mathcal{IG}$	0.10	2.00	0.50	0.58	0.66	0.43	0.61	0.91
σ_b	$\mathcal{IG}$	0.10	2.00	0.16	0.23	0.29	0.11	0.20	0.39
σ_g	$\mathcal{IG}$	0.10	2.00	0.46	0.54	0.63	0.40	0.52	0.87
σ_I	$\mathcal{IG}$	0.10	2.00	0.37	0.52	0.67	0.24	0.56	1.03
σ_r	$\mathcal{IG}$	0.10	2.00	0.17	0.20	0.24	0.16	0.21	0.31
σ_p	$\mathcal{IG}$	0.10	2.00	0.18	0.22	0.27	0.12	0.26	0.42
σ_w	$\mathcal{IG}$	0.10	2.00	0.17	0.20	0.24	0.17	0.25	0.36
Log Likelihood:					-320.24			-303.56	
$cond(\mathfrak{I}_\theta)$:					4.0e7			1.0e9	

Note: $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed. $\bar{\pi}$, and $\bar{l}$ are quarterly steady state inflation rate, and steady state hours worked.

Table B.15: Estimation Results: 1984:1-2004:4

Param.	Prior			Bayesian			MLE		
	Distr.	Mean	St.dev.	5%	mode	95%	5%	mode	95%
φ	$\mathcal{N}$	4.00	1.50	4.39	6.23	8.07	14.84	14.97	15.04
σ_c	$\mathcal{N}$	1.50	0.38	1.26	1.48	1.69	1.33	1.71	2.38
λ	$\mathcal{B}$	0.70	0.10	0.62	0.69	0.75	0.59	0.72	0.84
ξ_w	$\mathcal{B}$	0.50	0.10	0.53	0.75	0.96	0.94	0.95	0.95
σ_l	$\mathcal{N}$	2.00	0.75	1.20	2.30	3.40	1.19	2.42	5.14
ξ_p	$\mathcal{B}$	0.50	0.10	0.67	0.74	0.80	0.61	0.83	0.92
ι_w	$\mathcal{B}$	0.50	0.15	0.20	0.47	0.73	-0.14	0.27	0.81
ι_p	$\mathcal{B}$	0.50	0.15	0.07	0.21	0.36	0.01	0.01	0.01
ψ	$\mathcal{B}$	0.50	0.15	0.52	0.70	0.88	0.99	1.00	1.00
Φ	$\mathcal{N}$	1.25	0.12	1.39	1.54	1.69	1.40	1.59	2.11
r_π	$\mathcal{N}$	1.50	0.25	1.29	1.77	2.25	1.56	2.39	3.17
ρ	$\mathcal{B}$	0.75	0.10	0.81	0.84	0.88	0.83	0.89	0.95
r_y	$\mathcal{N}$	0.12	0.05	0.00	0.09	0.17	0.01	0.10	0.21
$r_{\triangle y}$	$\mathcal{N}$	0.12	0.05	0.13	0.16	0.19	0.11	0.20	0.30
$100(\frac{1}{\beta}-1)$	$\mathcal{G}$	0.25	0.10	0.05	0.13	0.21	0.00	0.00	0.34
α	$\mathcal{N}$	0.30	0.05	0.18	0.22	0.25	0.12	0.17	0.24
ρ_a	$\mathcal{B}$	0.50	0.20	0.91	0.94	0.97	0.92	0.98	1.00
ρ_b	$\mathcal{B}$	0.50	0.20	0.01	0.14	0.28	-0.38	0.09	0.37
ρ_g	$\mathcal{B}$	0.50	0.20	0.95	0.97	0.98	0.92	0.97	1.00
ρ_I	$\mathcal{B}$	0.50	0.20	0.53	0.65	0.76	0.49	0.75	0.92
ρ_r	$\mathcal{B}$	0.50	0.20	0.13	0.30	0.46	-0.18	0.12	0.47
ρ_p	$\mathcal{B}$	0.50	0.20	0.53	0.75	0.96	0.67	0.89	1.00
ρ_w	$\mathcal{B}$	0.50	0.20	0.58	0.83	1.07	0.18	0.65	0.90
μ_w	$\mathcal{B}$	0.50	0.20	0.34	0.62	0.90	-0.09	0.52	0.88
μ_p	$\mathcal{B}$	0.50	0.20	0.30	0.60	0.89	0.41	0.81	0.97
σ_a	$\mathcal{IG}$	0.10	2.00	0.32	0.35	0.39	0.28	0.37	0.45
σ_b	$\mathcal{IG}$	0.10	2.00	0.16	0.19	0.22	0.15	0.21	0.30
σ_g	$\mathcal{IG}$	0.10	2.00	0.37	0.42	0.46	0.34	0.47	0.55
σ_I	$\mathcal{IG}$	0.10	2.00	0.32	0.40	0.48	0.20	0.31	0.51
σ_r	$\mathcal{IG}$	0.10	2.00	0.11	0.12	0.14	0.09	0.12	0.15
σ_p	$\mathcal{IG}$	0.10	2.00	0.10	0.12	0.13	0.05	0.13	0.18
σ_w	$\mathcal{IG}$	0.10	2.00	0.17	0.22	0.27	0.15	0.22	0.34
Log Likelihood:					-337.76			-304.35	
$cond(\mathfrak{I}_\theta)$:					6.6e7			3.1e8	

Note: $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed. $\bar{\pi}$, and $\bar{l}$ are quarterly steady state inflation rate, and steady state hours worked.

B.5.2 Estimation: Unrestricted Model

Table B.16: Estimation Results: MLE

Parameter	1966:1-2004:4			1966:1-1979:2			1984:1-2004:4		
	5%	mode	95%	5%	mode	95%	5%	mode	95%
φ	4.13	7.92	13.63	2.00	2.00	2.00	14.96	15.00	15.10
σ_c	1.27	1.68	2.37	0.83	1.19	2.00	1.35	1.62	2.38
λ	0.59	0.72	0.86	0.28	0.53	0.68	0.55	0.70	0.83
ξ_w	0.61	0.85	0.98	0.47	0.74	0.96	0.95	0.95	0.96
σ_l	1.08	2.88	5.18	0.08	1.61	3.63	0.82	2.47	5.17
ξ_p	0.51	0.67	0.85	0.45	0.64	0.84	0.66	0.79	0.93
ι_w	0.32	0.79	1.18	0.39	0.84	1.26	-0.14	0.44	0.91
ι_p	0.01	0.01	0.01	-0.13	0.22	0.69	0.01	0.01	0.01
ψ	0.37	0.75	1.17	0.00	0.18	0.76	1.00	1.00	1.01
Φ	1.52	1.81	2.15	1.08	1.35	1.61	1.38	1.63	2.01
r_π	3.00	3.00	3.00	3.00	3.00	3.00	1.76	2.60	3.33
ρ	0.83	0.89	0.93	0.81	0.90	0.95	0.83	0.88	0.96
r_y	0.07	0.19	0.26	0.24	0.41	0.59	0.03	0.09	0.22
$r_{\triangle y}$	0.19	0.27	0.40	0.19	0.28	0.41	0.12	0.21	0.30
$100(\frac{1}{\beta}-1)$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.28
α	0.14	0.20	0.28	0.10	0.17	0.25	0.12	0.19	0.27
δ	0.00	0.02	0.05	0.01	0.02	0.03	0.01	0.02	0.05
λ_w	1.25	1.77	4.15	0.93	1.54	3.10	1.01	1.53	2.55
g_y	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
ρ_a	0.93	0.97	0.99	0.97	0.99	1.00	0.96	0.98	1.00
ρ_b	-0.11	0.12	0.47	0.19	0.61	0.91	-0.32	0.09	0.40
ρ_g	0.94	0.98	1.00	0.76	0.93	0.99	0.91	0.96	0.99
ρ_I	0.48	0.69	0.85	0.07	0.47	0.87	0.44	0.68	0.88
ρ_r	0.01	0.01	0.01	-0.23	0.07	0.39	-0.11	0.19	0.42
ρ_p	0.72	0.94	1.00	0.38	0.83	1.00	0.69	0.93	1.00
ρ_w	0.92	0.98	1.00	0.96	0.99	1.00	0.19	0.66	0.90
μ_w	0.89	0.97	1.00	0.86	0.97	1.02	-0.05	0.55	0.87
μ_p	0.29	0.76	0.93	0.85	0.99	1.22	0.44	0.83	0.96
σ_a	0.33	0.43	0.54	0.42	0.61	0.85	0.26	0.37	0.47
σ_b	0.19	0.26	0.35	0.11	0.20	0.31	0.13	0.20	0.30
σ_g	0.43	0.51	0.63	0.38	0.50	0.73	0.30	0.38	0.50
σ_I	0.33	0.46	0.63	0.32	0.57	0.92	0.23	0.35	0.51
σ_r	0.20	0.23	0.28	0.16	0.21	0.31	0.09	0.12	0.15
σ_p	0.06	0.12	0.20	0.20	0.27	0.41	0.06	0.12	0.17
σ_w	0.21	0.27	0.33	0.17	0.25	0.32	0.16	0.24	0.35
Log Likelihood:		-814.06			-301.31			-299.65	
$cond(\mathfrak{I}_\theta)$:		6.0e7			1.8e9			4.2e8	

Note: δ, λ_w, and g_y are estimated

B.6 Impulse responses

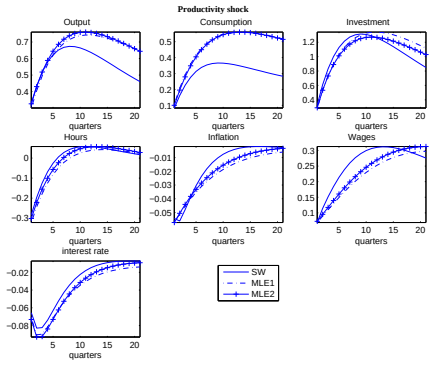

Figure B.2: Impulse Responses to a productivity shock

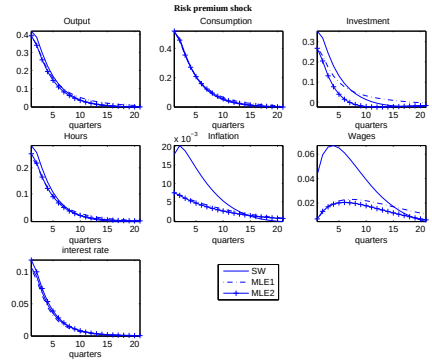

Figure B.3: Impulse Responses to risk premium shock

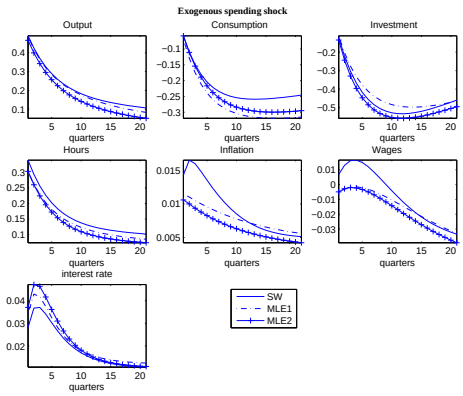

Figure B.4: Impulse Responses to exogenous spending shock

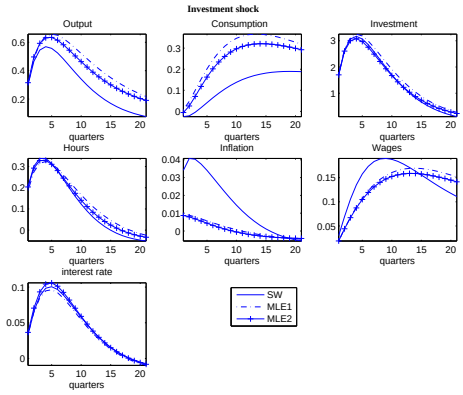

Figure B.5: Impulse Responses to investment shock

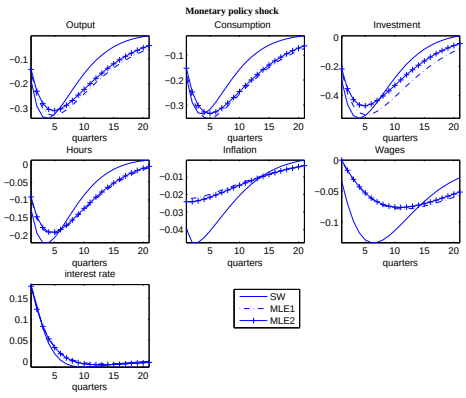

Figure B.6: Impulse Responses to monetary policy shock

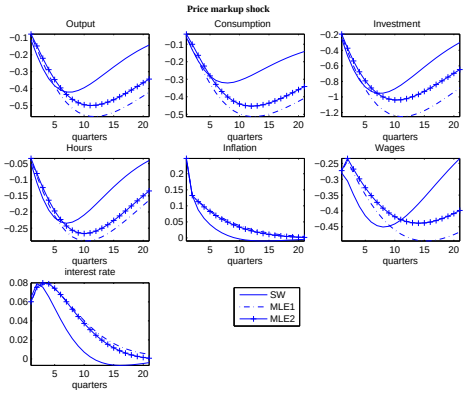

Figure B.7: Impulse Responses to price markup shock

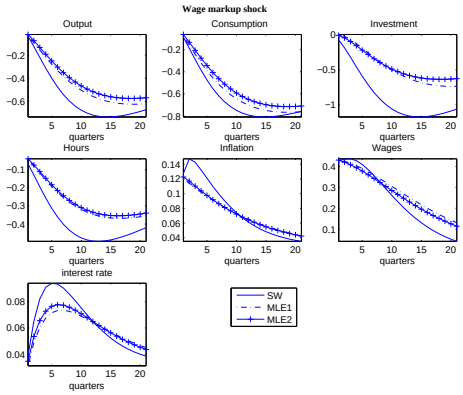

Figure B.8: Impulse Responses to wage markup shock

B.7 Variance Decompositions

Table B.17: Variance Decomposition: Bayesian 1966:1-2004:4

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.164	0.264	0.363	0.136	0.053	0.019	0.002
	Consumption	0.022	0.817	0.011	0.002	0.112	0.012	0.024
	Investment	0.033	0.037	0.004	0.879	0.026	0.019	0.002
	Hours	0.234	0.239	0.337	0.123	0.047	0.007	0.014
	Inflation	0.036	0.004	0.002	0.014	0.019	0.725	0.199
	Wages	0.018	0.007	0.000	0.003	0.004	0.292	0.676
	interest rate	0.077	0.204	0.014	0.020	0.583	0.075	0.027
2	Output	0.193	0.210	0.274	0.200	0.077	0.035	0.011
	Consumption	0.047	0.671	0.022	0.001	0.169	0.030	0.060
	Investment	0.043	0.020	0.006	0.877	0.027	0.025	0.004
	Hours	0.172	0.215	0.292	0.193	0.077	0.021	0.031
	Inflation	0.047	0.006	0.004	0.022	0.030	0.597	0.295
	Wages	0.031	0.009	0.000	0.010	0.011	0.297	0.642
	interest rate	0.106	0.190	0.020	0.052	0.477	0.100	0.055
4	Output	0.242	0.124	0.179	0.250	0.093	0.067	0.044
	Consumption	0.097	0.413	0.043	0.001	0.207	0.072	0.167
	Investment	0.063	0.009	0.008	0.846	0.026	0.038	0.010
	Hours	0.101	0.151	0.231	0.262	0.108	0.056	0.091
	Inflation	0.052	0.007	0.005	0.031	0.042	0.464	0.399
	Wages	0.056	0.011	0.001	0.026	0.022	0.330	0.555
	interest rate	0.132	0.152	0.026	0.129	0.327	0.115	0.118
40	Output	0.308	0.018	0.045	0.089	0.026	0.071	0.443
	Consumption	0.116	0.030	0.078	0.034	0.029	0.052	0.661
	Investment	0.196	0.002	0.045	0.472	0.014	0.069	0.201
	Hours	0.021	0.029	0.096	0.092	0.037	0.067	0.659
	Inflation	0.041	0.006	0.008	0.035	0.048	0.298	0.564
	Wages	0.291	0.004	0.004	0.072	0.022	0.398	0.208
	interest rate	0.105	0.084	0.033	0.206	0.168	0.077	0.327
100	Output	0.295	0.016	0.042	0.079	0.023	0.063	0.482
	Consumption	0.105	0.023	0.090	0.032	0.022	0.042	0.686
	Investment	0.192	0.002	0.051	0.456	0.013	0.066	0.219
	Hours	0.020	0.026	0.105	0.086	0.034	0.062	0.668
	Inflation	0.040	0.006	0.010	0.034	0.046	0.285	0.579
	Wages	0.314	0.004	0.010	0.072	0.021	0.377	0.202
	interest rate	0.104	0.077	0.039	0.194	0.154	0.071	0.361

Note: Based on the posterior mode of θ reported in Smets and Wouters (2007). $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed.

Table B.18: Variance Decomposition: MLE1 1966:1-2004:4

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.150	0.252	0.393	0.151	0.039	0.012	0.002
	Consumption	0.037	0.828	0.015	0.000	0.090	0.010	0.020
	Investment	0.025	0.022	0.004	0.918	0.017	0.014	0.000
	Hours	0.344	0.195	0.304	0.116	0.030	0.005	0.006
	Inflation	0.041	0.001	0.002	0.001	0.006	0.769	0.180
	Wages	0.018	0.000	0.000	0.001	0.000	0.283	0.698
	interest rate	0.091	0.193	0.021	0.021	0.593	0.064	0.017
2	Output	0.183	0.194	0.296	0.235	0.061	0.025	0.007
	Consumption	0.079	0.669	0.031	0.003	0.145	0.027	0.046
	Investment	0.033	0.011	0.005	0.913	0.018	0.019	0.000
	Hours	0.277	0.176	0.273	0.188	0.055	0.016	0.014
	Inflation	0.052	0.001	0.002	0.001	0.009	0.694	0.241
	Wages	0.025	0.000	0.000	0.005	0.001	0.260	0.709
	interest rate	0.130	0.181	0.030	0.052	0.484	0.089	0.033
4	Output	0.234	0.109	0.186	0.312	0.080	0.056	0.024
	Consumption	0.157	0.387	0.056	0.021	0.192	0.074	0.112
	Investment	0.049	0.005	0.007	0.884	0.020	0.033	0.001
	Hours	0.186	0.128	0.224	0.278	0.091	0.052	0.041
	Inflation	0.058	0.001	0.003	0.002	0.012	0.622	0.303
	Wages	0.039	0.001	0.000	0.014	0.003	0.291	0.652
	interest rate	0.162	0.142	0.036	0.125	0.345	0.124	0.067
10	Output	0.312	0.037	0.081	0.276	0.067	0.138	0.090
	Consumption	0.225	0.101	0.073	0.070	0.120	0.161	0.250
	Investment	0.110	0.002	0.015	0.752	0.022	0.090	0.009
	Hours	0.083	0.062	0.145	0.257	0.101	0.177	0.176
	Inflation	0.059	0.001	0.004	0.001	0.015	0.528	0.391
	Wages	0.083	0.001	0.000	0.041	0.010	0.416	0.449
	interest rate	0.156	0.093	0.034	0.207	0.214	0.146	0.149
40	Output	0.370	0.012	0.031	0.130	0.027	0.152	0.278
	Consumption	0.222	0.020	0.085	0.077	0.031	0.133	0.431
	Investment	0.234	0.001	0.040	0.476	0.016	0.156	0.078
	Hours	0.036	0.025	0.089	0.115	0.047	0.178	0.509
	Inflation	0.052	0.001	0.006	0.003	0.015	0.432	0.491
	Wages	0.219	0.001	0.004	0.057	0.009	0.540	0.171
	interest rate	0.142	0.071	0.039	0.186	0.166	0.124	0.271
100	Output	0.371	0.010	0.026	0.110	0.022	0.130	0.330
	Consumption	0.215	0.015	0.104	0.064	0.023	0.105	0.473
	Investment	0.240	0.001	0.049	0.453	0.015	0.149	0.094
	Hours	0.039	0.022	0.097	0.105	0.041	0.158	0.539
	Inflation	0.052	0.001	0.008	0.004	0.014	0.406	0.516
	Wages	0.275	0.001	0.011	0.055	0.008	0.504	0.147
	interest rate	0.145	0.063	0.048	0.172	0.146	0.116	0.311

Note: $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed.

Table B.19: Variance Decomposition: MLE2 1966:1-2004:4

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.176	0.255	0.361	0.165	0.033	0.010	0.001
	Consumption	0.032	0.863	0.011	0.000	0.075	0.006	0.013
	Investment	0.027	0.023	0.006	0.918	0.015	0.012	0.000
	Hours	0.303	0.215	0.306	0.138	0.028	0.004	0.005
	Inflation	0.041	0.001	0.001	0.001	0.007	0.762	0.187
	Wages	0.020	0.000	0.000	0.001	0.000	0.280	0.698
	interest rate	0.091	0.239	0.023	0.023	0.542	0.061	0.020
2	Output	0.209	0.194	0.271	0.249	0.052	0.020	0.004
	Consumption	0.073	0.723	0.024	0.001	0.128	0.018	0.033
	Investment	0.035	0.011	0.007	0.914	0.016	0.017	0.000
	Hours	0.240	0.192	0.273	0.217	0.051	0.013	0.012
	Inflation	0.051	0.001	0.002	0.001	0.010	0.685	0.250
	Wages	0.028	0.000	0.000	0.005	0.001	0.255	0.712
	interest rate	0.129	0.222	0.033	0.058	0.435	0.086	0.037
4	Output	0.265	0.109	0.170	0.320	0.071	0.046	0.018
	Consumption	0.162	0.441	0.049	0.013	0.185	0.058	0.092
	Investment	0.053	0.004	0.011	0.884	0.018	0.029	0.001
	Hours	0.158	0.139	0.224	0.310	0.088	0.045	0.037
	Inflation	0.056	0.001	0.002	0.001	0.014	0.613	0.313
	Wages	0.046	0.001	0.000	0.014	0.003	0.279	0.657
	interest rate	0.157	0.172	0.040	0.138	0.303	0.117	0.073
10	Output	0.359	0.038	0.073	0.268	0.062	0.119	0.081
	Consumption	0.254	0.117	0.071	0.057	0.123	0.143	0.235
	Investment	0.120	0.002	0.023	0.749	0.019	0.077	0.011
	Hours	0.074	0.069	0.146	0.271	0.103	0.163	0.175
	Inflation	0.055	0.001	0.003	0.001	0.018	0.518	0.404
	Wages	0.103	0.001	0.000	0.041	0.010	0.389	0.456
	interest rate	0.147	0.114	0.038	0.228	0.187	0.134	0.153
40	Output	0.426	0.012	0.027	0.126	0.025	0.123	0.260
	Consumption	0.252	0.023	0.084	0.073	0.031	0.110	0.426
	Investment	0.252	0.001	0.057	0.488	0.014	0.119	0.070
	Hours	0.036	0.028	0.087	0.123	0.049	0.158	0.520
	Inflation	0.047	0.001	0.004	0.002	0.017	0.423	0.505
	Wages	0.279	0.000	0.006	0.060	0.009	0.468	0.178
	interest rate	0.130	0.091	0.040	0.207	0.151	0.114	0.267
100	Output	0.424	0.010	0.022	0.106	0.021	0.103	0.314
	Consumption	0.248	0.017	0.097	0.062	0.023	0.083	0.472
	Investment	0.262	0.001	0.066	0.459	0.013	0.113	0.086
	Hours	0.039	0.024	0.089	0.109	0.042	0.138	0.559
	Inflation	0.045	0.001	0.005	0.003	0.016	0.394	0.537
	Wages	0.350	0.000	0.015	0.058	0.008	0.418	0.151
	interest rate	0.131	0.081	0.045	0.191	0.135	0.105	0.310

Note: δ, λ_w, and g_y are estimated.

Table B.20: Variance Decomposition: Bayesian 1966:1-1979:2

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.279	0.265	0.300	0.091	0.049	0.013	0.002
	Consumption	0.086	0.792	0.000	0.001	0.110	0.011	0.000
	Investment	0.061	0.083	0.020	0.794	0.034	0.008	0.000
	Hours	0.141	0.315	0.370	0.111	0.057	0.005	0.001
	Inflation	0.072	0.011	0.004	0.008	0.022	0.644	0.240
	Wages	0.106	0.023	0.002	0.003	0.009	0.359	0.499
	interest rate	0.114	0.342	0.038	0.018	0.345	0.068	0.075
2	Output	0.336	0.240	0.218	0.119	0.068	0.019	0.001
	Consumption	0.155	0.676	0.001	0.001	0.143	0.021	0.003
	Investment	0.088	0.057	0.028	0.779	0.036	0.010	0.002
	Hours	0.091	0.328	0.312	0.161	0.091	0.013	0.004
	Inflation	0.097	0.017	0.007	0.014	0.038	0.427	0.401
	Wages	0.204	0.030	0.003	0.006	0.016	0.283	0.458
	interest rate	0.129	0.338	0.046	0.040	0.247	0.063	0.139
4	Output	0.457	0.161	0.140	0.129	0.076	0.024	0.012
	Consumption	0.312	0.457	0.005	0.001	0.158	0.034	0.033
	Investment	0.154	0.031	0.044	0.713	0.035	0.011	0.012
	Hours	0.055	0.279	0.259	0.212	0.128	0.027	0.039
	Inflation	0.092	0.020	0.008	0.019	0.053	0.257	0.551
	Wages	0.382	0.029	0.003	0.012	0.025	0.185	0.363
	interest rate	0.132	0.280	0.051	0.079	0.144	0.041	0.274
10	Output	0.622	0.068	0.064	0.078	0.045	0.015	0.109
	Consumption	0.503	0.173	0.013	0.000	0.081	0.020	0.209
	Investment	0.329	0.013	0.079	0.468	0.022	0.007	0.081
	Hours	0.054	0.165	0.170	0.163	0.101	0.024	0.322
	Inflation	0.066	0.015	0.008	0.018	0.055	0.181	0.657
	Wages	0.682	0.014	0.001	0.015	0.020	0.080	0.187
	interest rate	0.096	0.180	0.044	0.098	0.086	0.026	0.471
40	Output	0.659	0.032	0.031	0.037	0.021	0.007	0.212
	Consumption	0.553	0.059	0.014	0.008	0.028	0.007	0.330
	Investment	0.528	0.006	0.066	0.223	0.010	0.003	0.163
	Hours	0.043	0.092	0.100	0.096	0.057	0.014	0.599
	Inflation	0.056	0.012	0.007	0.015	0.044	0.146	0.720
	Wages	0.874	0.005	0.002	0.010	0.008	0.030	0.071
	interest rate	0.080	0.132	0.034	0.076	0.064	0.019	0.594
100	Output	0.670	0.029	0.028	0.034	0.019	0.006	0.212
	Consumption	0.600	0.044	0.013	0.009	0.021	0.005	0.308
	Investment	0.563	0.006	0.059	0.199	0.009	0.003	0.160
	Hours	0.078	0.086	0.095	0.091	0.053	0.013	0.584
	Inflation	0.060	0.012	0.007	0.015	0.042	0.140	0.725
	Wages	0.895	0.004	0.002	0.009	0.006	0.024	0.060
	interest rate	0.098	0.122	0.033	0.071	0.059	0.018	0.600

Note: Based on the posterior mode of θ reported in Smets and Wouters (2007). $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed.

Table B.21: Variance Decomposition: MLE1 1966:1-1979:2

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.403	0.252	0.225	0.037	0.068	0.014	0.001
	Consumption	0.267	0.562	0.011	0.005	0.120	0.032	0.004
	Investment	0.039	0.166	0.041	0.660	0.084	0.001	0.008
	Hours	0.102	0.380	0.345	0.057	0.102	0.013	0.001
	Inflation	0.160	0.006	0.001	0.000	0.033	0.421	0.379
	Wages	0.188	0.007	0.000	0.000	0.001	0.121	0.683
	interest rate	0.089	0.527	0.030	0.004	0.299	0.000	0.050
2	Output	0.473	0.239	0.152	0.040	0.085	0.009	0.001
	Consumption	0.371	0.449	0.015	0.005	0.123	0.026	0.012
	Investment	0.065	0.150	0.058	0.616	0.102	0.001	0.007
	Hours	0.057	0.424	0.277	0.072	0.151	0.015	0.004
	Inflation	0.151	0.006	0.001	0.001	0.039	0.392	0.411
	Wages	0.306	0.013	0.000	0.000	0.005	0.108	0.567
	interest rate	0.089	0.589	0.032	0.009	0.194	0.008	0.078
4	Output	0.603	0.167	0.093	0.034	0.091	0.005	0.006
	Consumption	0.543	0.274	0.018	0.005	0.107	0.013	0.039
	Investment	0.134	0.116	0.096	0.516	0.123	0.010	0.004
	Hours	0.041	0.396	0.228	0.081	0.213	0.011	0.029
	Inflation	0.118	0.006	0.001	0.001	0.043	0.419	0.413
	Wages	0.465	0.016	0.001	0.001	0.013	0.129	0.376
	interest rate	0.084	0.570	0.032	0.016	0.119	0.038	0.140
10	Output	0.745	0.068	0.040	0.016	0.054	0.027	0.050
	Consumption	0.698	0.099	0.015	0.003	0.050	0.015	0.121
	Investment	0.336	0.052	0.153	0.274	0.100	0.072	0.014
	Hours	0.058	0.245	0.153	0.057	0.186	0.076	0.225
	Inflation	0.091	0.005	0.001	0.000	0.052	0.342	0.509
	Wages	0.724	0.008	0.000	0.001	0.017	0.101	0.150
	interest rate	0.068	0.453	0.028	0.020	0.085	0.053	0.293
40	Output	0.787	0.025	0.015	0.006	0.020	0.015	0.130
	Consumption	0.746	0.027	0.007	0.001	0.014	0.006	0.198
	Investment	0.636	0.019	0.100	0.099	0.037	0.046	0.064
	Hours	0.063	0.109	0.072	0.026	0.083	0.050	0.596
	Inflation	0.064	0.003	0.000	0.000	0.039	0.242	0.652
	Wages	0.928	0.002	0.000	0.000	0.005	0.032	0.033
	interest rate	0.045	0.297	0.019	0.013	0.057	0.037	0.531
100	Output	0.710	0.012	0.007	0.003	0.010	0.007	0.250
	Consumption	0.669	0.009	0.003	0.000	0.004	0.002	0.313
	Investment	0.749	0.008	0.045	0.044	0.017	0.021	0.116
	Hours	0.123	0.038	0.026	0.009	0.029	0.018	0.757
	Inflation	0.030	0.002	0.000	0.000	0.018	0.112	0.839
	Wages	0.971	0.001	0.000	0.000	0.002	0.013	0.013
	interest rate	0.024	0.125	0.008	0.006	0.024	0.016	0.797

Note: $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed.

Table B.22: Variance Decomposition: MLE2 1966:1-1979:2

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.434	0.250	0.190	0.047	0.062	0.015	0.001
	Consumption	0.275	0.558	0.014	0.007	0.109	0.034	0.004
	Investment	0.032	0.180	0.038	0.666	0.078	0.002	0.005
	Hours	0.069	0.413	0.319	0.079	0.103	0.016	0.001
	Inflation	0.176	0.006	0.001	0.001	0.028	0.410	0.378
	Wages	0.191	0.006	0.000	0.000	0.001	0.118	0.685
	interest rate	0.095	0.547	0.028	0.007	0.270	0.000	0.052
2	Output	0.498	0.238	0.127	0.050	0.078	0.009	0.001
	Consumption	0.380	0.445	0.018	0.008	0.111	0.027	0.012
	Investment	0.054	0.166	0.055	0.625	0.094	0.001	0.004
	Hours	0.035	0.452	0.248	0.095	0.148	0.016	0.005
	Inflation	0.163	0.006	0.001	0.001	0.032	0.400	0.398
	Wages	0.307	0.011	0.000	0.001	0.004	0.106	0.571
	interest rate	0.092	0.606	0.029	0.013	0.172	0.009	0.078
4	Output	0.621	0.167	0.076	0.041	0.082	0.006	0.006
	Consumption	0.552	0.272	0.022	0.007	0.095	0.014	0.038
	Investment	0.115	0.133	0.093	0.532	0.117	0.008	0.002
	Hours	0.035	0.418	0.199	0.102	0.203	0.012	0.031
	Inflation	0.129	0.006	0.001	0.001	0.035	0.429	0.400
	Wages	0.467	0.013	0.000	0.001	0.010	0.127	0.381
	interest rate	0.084	0.588	0.027	0.022	0.104	0.041	0.134
10	Output	0.758	0.068	0.033	0.019	0.049	0.027	0.046
	Consumption	0.710	0.099	0.019	0.003	0.044	0.015	0.110
	Investment	0.300	0.064	0.156	0.291	0.097	0.071	0.021
	Hours	0.074	0.256	0.133	0.069	0.174	0.076	0.219
	Inflation	0.102	0.005	0.001	0.000	0.043	0.358	0.492
	Wages	0.725	0.007	0.000	0.001	0.013	0.102	0.151
	interest rate	0.069	0.480	0.024	0.025	0.076	0.056	0.271
40	Output	0.807	0.026	0.013	0.007	0.019	0.015	0.113
	Consumption	0.773	0.027	0.010	0.002	0.012	0.006	0.170
	Investment	0.582	0.024	0.112	0.110	0.038	0.048	0.085
	Hours	0.068	0.123	0.070	0.034	0.084	0.054	0.567
	Inflation	0.073	0.003	0.000	0.000	0.033	0.259	0.631
	Wages	0.929	0.001	0.000	0.000	0.004	0.032	0.032
	interest rate	0.049	0.335	0.017	0.018	0.055	0.041	0.486
100	Output	0.794	0.013	0.007	0.004	0.010	0.008	0.165
	Consumption	0.772	0.009	0.004	0.001	0.004	0.002	0.208
	Investment	0.738	0.011	0.051	0.049	0.017	0.021	0.113
	Hours	0.165	0.057	0.033	0.016	0.039	0.025	0.665
	Inflation	0.044	0.002	0.000	0.000	0.019	0.154	0.780
	Wages	0.972	0.001	0.000	0.000	0.001	0.013	0.013
	interest rate	0.034	0.189	0.010	0.010	0.031	0.023	0.702

Note: δ, λ_w, and g_y are estimated.

Table B.23: Variance Decomposition: Bayesian 1984:1-2004:4

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.082	0.237	0.390	0.222	0.057	0.011	0.000
	Consumption	0.005	0.844	0.006	0.001	0.129	0.010	0.005
	Investment	0.019	0.024	0.004	0.918	0.026	0.008	0.001
	Hours	0.271	0.187	0.314	0.174	0.044	0.005	0.005
	Inflation	0.027	0.002	0.005	0.014	0.024	0.756	0.172
	Wages	0.002	0.004	0.001	0.009	0.006	0.127	0.851
	interest rate	0.119	0.192	0.036	0.052	0.476	0.088	0.035
2	Output	0.095	0.182	0.294	0.317	0.090	0.021	0.001
	Consumption	0.013	0.703	0.014	0.008	0.218	0.025	0.018
	Investment	0.025	0.013	0.006	0.913	0.029	0.011	0.002
	Hours	0.210	0.162	0.269	0.258	0.077	0.012	0.012
	Inflation	0.040	0.003	0.008	0.022	0.039	0.617	0.270
	Wages	0.003	0.005	0.001	0.017	0.012	0.108	0.854
	interest rate	0.142	0.153	0.045	0.103	0.403	0.092	0.062
4	Output	0.124	0.109	0.196	0.396	0.124	0.038	0.013
	Consumption	0.038	0.451	0.032	0.033	0.316	0.060	0.070
	Investment	0.041	0.007	0.010	0.885	0.034	0.016	0.007
	Hours	0.139	0.114	0.215	0.341	0.122	0.029	0.039
	Inflation	0.049	0.005	0.012	0.032	0.060	0.470	0.371
	Wages	0.006	0.006	0.002	0.035	0.023	0.100	0.830
	interest rate	0.160	0.110	0.054	0.201	0.279	0.083	0.113
10	Output	0.208	0.050	0.110	0.364	0.121	0.057	0.089
	Consumption	0.092	0.166	0.072	0.091	0.245	0.083	0.251
	Investment	0.098	0.003	0.027	0.769	0.039	0.026	0.038
	Hours	0.077	0.067	0.162	0.319	0.142	0.056	0.178
	Inflation	0.052	0.005	0.018	0.038	0.085	0.380	0.422
	Wages	0.025	0.006	0.002	0.079	0.045	0.112	0.733
	interest rate	0.153	0.075	0.061	0.299	0.176	0.057	0.179
40	Output	0.324	0.032	0.075	0.270	0.085	0.044	0.170
	Consumption	0.150	0.070	0.170	0.150	0.117	0.045	0.299
	Investment	0.167	0.003	0.077	0.620	0.033	0.023	0.077
	Hours	0.062	0.052	0.160	0.265	0.118	0.050	0.293
	Inflation	0.053	0.005	0.026	0.047	0.087	0.371	0.410
	Wages	0.109	0.005	0.006	0.117	0.052	0.113	0.598
	interest rate	0.152	0.067	0.085	0.306	0.161	0.053	0.175
100	Output	0.331	0.031	0.075	0.268	0.084	0.044	0.168
	Consumption	0.153	0.063	0.209	0.154	0.105	0.041	0.274
	Investment	0.167	0.003	0.082	0.617	0.033	0.023	0.076
	Hours	0.063	0.051	0.170	0.264	0.116	0.049	0.288
	Inflation	0.054	0.005	0.030	0.049	0.087	0.367	0.407
	Wages	0.117	0.005	0.012	0.118	0.051	0.111	0.587
	interest rate	0.154	0.065	0.099	0.305	0.155	0.052	0.171

Note: Based on the posterior mode of θ reported in Smets and Wouters (2007). $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed.

Table B.24: Variance Decomposition: MLE1 1984:1-2004:4

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.103	0.222	0.394	0.234	0.032	0.011	0.005
	Consumption	0.027	0.832	0.004	0.017	0.088	0.016	0.017
	Investment	0.009	0.006	0.002	0.968	0.007	0.009	0.000
	Hours	0.340	0.163	0.290	0.172	0.023	0.008	0.003
	Inflation	0.031	0.000	0.001	0.001	0.001	0.819	0.147
	Wages	0.004	0.000	0.000	0.001	0.000	0.110	0.885
	interest rate	0.085	0.190	0.027	0.031	0.537	0.116	0.015
2	Output	0.115	0.156	0.290	0.356	0.050	0.021	0.011
	Consumption	0.058	0.651	0.008	0.049	0.154	0.038	0.042
	Investment	0.012	0.003	0.002	0.963	0.008	0.012	0.000
	Hours	0.274	0.137	0.256	0.261	0.044	0.018	0.009
	Inflation	0.045	0.000	0.002	0.002	0.002	0.729	0.221
	Wages	0.006	0.000	0.000	0.001	0.000	0.093	0.900
	interest rate	0.118	0.150	0.037	0.066	0.481	0.120	0.028
4	Output	0.135	0.083	0.177	0.468	0.069	0.042	0.025
	Consumption	0.113	0.352	0.015	0.131	0.213	0.080	0.096
	Investment	0.019	0.001	0.003	0.945	0.010	0.020	0.001
	Hours	0.182	0.094	0.204	0.369	0.077	0.046	0.029
	Inflation	0.058	0.000	0.003	0.003	0.002	0.633	0.300
	Wages	0.010	0.000	0.000	0.003	0.000	0.099	0.888
	interest rate	0.151	0.106	0.046	0.138	0.369	0.129	0.059
10	Output	0.181	0.029	0.078	0.470	0.072	0.090	0.080
	Consumption	0.163	0.086	0.020	0.250	0.153	0.130	0.197
	Investment	0.051	0.001	0.007	0.863	0.016	0.052	0.010
	Hours	0.078	0.046	0.130	0.388	0.107	0.127	0.124
	Inflation	0.071	0.000	0.005	0.007	0.003	0.526	0.388
	Wages	0.029	0.000	0.000	0.011	0.000	0.155	0.804
	interest rate	0.166	0.067	0.051	0.215	0.239	0.127	0.134
40	Output	0.296	0.010	0.031	0.302	0.045	0.101	0.214
	Consumption	0.230	0.019	0.028	0.245	0.061	0.102	0.315
	Investment	0.170	0.001	0.022	0.624	0.018	0.096	0.069
	Hours	0.042	0.022	0.088	0.233	0.080	0.154	0.381
	Inflation	0.080	0.000	0.009	0.013	0.003	0.474	0.421
	Wages	0.157	0.000	0.002	0.058	0.002	0.281	0.501
	interest rate	0.176	0.053	0.064	0.220	0.190	0.106	0.191
100	Output	0.325	0.009	0.029	0.281	0.041	0.096	0.219
	Consumption	0.260	0.017	0.035	0.225	0.055	0.092	0.315
	Investment	0.180	0.001	0.024	0.611	0.018	0.096	0.071
	Hours	0.059	0.021	0.090	0.234	0.074	0.157	0.365
	Inflation	0.081	0.000	0.010	0.013	0.003	0.473	0.420
	Wages	0.238	0.000	0.005	0.064	0.003	0.259	0.430
	interest rate	0.191	0.050	0.070	0.226	0.178	0.099	0.187

Note: $\delta = .025$, $\lambda_w = 1.5$ and $g_y = .18$ are fixed.

Table B.25: Variance Decomposition: MLE2 1984:1-2004:4

qrt		product-ivity	risk premium	exog. spend.	invest-ment	monetary policy	price markup	wage markup
1	Output	0.118	0.237	0.371	0.228	0.032	0.010	0.004
	Consumption	0.041	0.834	0.003	0.010	0.084	0.013	0.014
	Investment	0.008	0.005	0.002	0.970	0.006	0.008	0.000
	Hours	0.345	0.176	0.276	0.169	0.024	0.007	0.003
	Inflation	0.031	0.000	0.001	0.001	0.001	0.815	0.150
	Wages	0.004	0.000	0.000	0.001	0.000	0.111	0.884
	interest rate	0.090	0.225	0.030	0.032	0.497	0.111	0.015
2	Output	0.135	0.168	0.271	0.348	0.050	0.019	0.009
	Consumption	0.086	0.661	0.007	0.033	0.148	0.031	0.035
	Investment	0.011	0.003	0.002	0.965	0.007	0.011	0.000
	Hours	0.277	0.151	0.244	0.259	0.045	0.017	0.008
	Inflation	0.045	0.000	0.002	0.002	0.001	0.725	0.225
	Wages	0.006	0.000	0.000	0.001	0.000	0.091	0.901
	interest rate	0.121	0.180	0.041	0.069	0.444	0.116	0.029
4	Output	0.621	0.167	0.076	0.041	0.082	0.006	0.006
	Consumption	0.552	0.272	0.022	0.007	0.095	0.014	0.038
	Investment	0.115	0.133	0.093	0.532	0.117	0.008	0.002
	Hours	0.035	0.418	0.199	0.102	0.203	0.012	0.031
	Inflation	0.129	0.006	0.001	0.001	0.035	0.429	0.400
	Wages	0.467	0.013	0.000	0.001	0.010	0.127	0.381
	interest rate	0.084	0.588	0.027	0.022	0.104	0.041	0.134
10	Output	0.215	0.030	0.070	0.460	0.070	0.083	0.072
	Consumption	0.227	0.091	0.017	0.216	0.150	0.118	0.181
	Investment	0.049	0.001	0.008	0.873	0.014	0.048	0.007
	Hours	0.080	0.051	0.125	0.392	0.110	0.124	0.119
	Inflation	0.071	0.000	0.003	0.008	0.002	0.523	0.392
	Wages	0.032	0.000	0.000	0.010	0.000	0.154	0.804
	interest rate	0.157	0.080	0.053	0.230	0.220	0.124	0.136
40	Output	0.352	0.010	0.026	0.291	0.042	0.091	0.188
	Consumption	0.307	0.020	0.022	0.226	0.058	0.089	0.278
	Investment	0.179	0.001	0.022	0.635	0.017	0.089	0.057
	Hours	0.045	0.025	0.082	0.239	0.084	0.153	0.372
	Inflation	0.080	0.000	0.006	0.014	0.003	0.470	0.427
	Wages	0.177	0.000	0.002	0.053	0.002	0.275	0.491
	interest rate	0.166	0.064	0.061	0.231	0.176	0.104	0.197
100	Output	0.395	0.009	0.024	0.264	0.038	0.084	0.186
	Consumption	0.358	0.017	0.026	0.203	0.050	0.078	0.268
	Investment	0.197	0.001	0.023	0.617	0.016	0.088	0.058
	Hours	0.069	0.023	0.081	0.239	0.077	0.156	0.354
	Inflation	0.081	0.000	0.006	0.014	0.003	0.469	0.427
	Wages	0.284	0.000	0.004	0.058	0.002	0.246	0.406
	interest rate	0.186	0.060	0.064	0.234	0.165	0.098	0.192

Note: δ, λ_w, and g_y are estimated.

Bibliography

Altig, D., L. Christiano, M. Eichenbaum, and J. Linde (2005): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," NBER Working Papers 11034, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberwo/11034.html.

An, A. (2005): "Bayesian Estimation of DSGE Models: Lessons from Second Order Approximations," .

An, S., and F. Schorfheide (2005): "Bayesian Analysis of DSGE Models," CEPR Discussion Papers 5207, C.E.P.R. Discussion Papers, available at http://ideas.repec.org/p/cpr/ceprdp/5207.html.

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.

Andre Klein, G. M., and T. Zahaf (2000): "Construction of the exact Fisher information matrix of Gaussian time series models by means of matrix differential rules," *Linear Algebra and its Applications*, 321(1-3), 209–232.

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.

Belsley, D., E. Kuh, and R. Welsch (1980): *Regression Diagnostics - Identifying Influential Data and Sources of Variation*. New York: Wiley.

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.

Boivin, J., and M. Giannoni (2006): "DSGE Models in a Data-Rich Environment," NBER Working Papers 12772, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberwo/12772.html.

Boswijk, H. P., and J. Doornik (2003): "Identifying, Estimating and Testing Restricted Cointegrated Systems: An Overview," Economics Papers 2003-W10, Economics Group, Nuffield College, University of Oxford, available at http://ideas.repec.org/p/nuf/econwp/0310.html.

Canova, F., and L. Sala (2006): "Back to square one: identification issues in DSGE models," Working Paper Series 583, European Central Bank, available at http://ideas.repec.org/p/ecb/ecbwps/20060583.html.

Chang, Y., T. Doh, and F. Schorfheide (2007): "Non-stationary Hours in a DSGE Model," *Journal of Money, Credit and Banking*, 39(6), 1357–1373, available at http://ideas.repec.org/a/mcb/jmoncb/v39y2007i6p1357-1373.html.

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).

DeJong, D. N., B. F. Ingram, and C. H. Whiteman (2000): "Keynesian impulses versus Solow residuals: identifying sources of business cycle fluctuations," *Journal of Applied Econometrics*, 15(3), 311–329, available at http://ideas.repec.org/a/jae/japmet/v15y2000i3p311-329.html.

DelNegro, M., F. Schorfheide, F. Smets, and R. Wouters (2005): "On the fit and forecasting performance of New-Keynesian models," Working Paper Series 491, European Central Bank, available at http://ideas.repec.org/p/ecb/ecbwps/20050491.html.

Demmel, J. (1987): "The condition number and the distance to the nearest ill-posed problem," *Numerische Mathematik*, 51(251289).

Dufour, J.-M. (2003): "Identification, weak instruments, and statistical inference in econometrics," *Canadian Journal of Economics*, 36(4), 767–808, available at http://ideas.repec.org/a/cje/issued/v36y2003i4p767-808.html.

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

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez (2005): "Estimating dynamic equilibrium economies: linear versus nonlinear likelihood," *Journal of Applied Econometrics*, 20(7), 891–910, available at http://ideas.repec.org/a/jae/japmet/v20y2005i7p891-910.html.

Fisher, F. (1966): *The identification problem in econometrics*. McGraw-Hill.

Forchini, G., and G. Hillier (2005): "Ill-conditioned problems, Fisher information and weak instruments," CeMMAP working papers CWP04/05, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, available at http://ideas.repec.org/p/ifs/cemmap/04-05.html.

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

Hamilton, J. (1994): *Time Series Analysis*. Princeton University Press.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14(3), 262–80, available at http://ideas.repec.org/a/bes/jnlbes/v14y1996i3p262-80.html.

Hansen, L. P., E. R. McGrattan, and T. J. Sargent (1994): "Mechanics of forming and estimating dynamic linear economies," Staff Report 182, Federal Reserve Bank of Minneapolis, available at http://ideas.repec.org/p/fip/fedmsr/182.html.

Horn, R., and C. Johnson (1985): *Matrix Analysis*. Cambridge University Press.

Iskrev, N. (2007a): "Identification and Estimation of DSGE model - An Integrated Approach," .

——— (2007b): "Identification and Estimation of DSGE model - An Integrated Approach," .

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee (1982): *Introduction to the Theory and Practice of Econometrics*. Wiley, New York.

Juillard, M., P. Karam, D. Laxton, and P. Pesenti (2006): "Welfare-based monetary policy rules in an estimated DSGE model of the US economy," Working Paper Series 613, European Central Bank, available at http://ideas.repec.org/p/ecb/ecbwps/20060613.html.

Justiniano, A., and G. E. Primiceri (2006): "The Time Varying Volatility of Macroeconomic Fluctuations," NBER Working Papers 12022, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberwo/12022.html.

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

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

Koopmans, T., and Q. Reiersol (1950): "The identification of Structural Characersistics," *Annals of Mathematical Statistics*, 21, 165–181.

Levin, A. T., A. Onatski, J. C. Williams, and N. Williams (2005): "Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models," NBER Working Papers 11523, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberwo/11523.html.

Li, H. (2004): "The Econometrics of the Lucas Critique: Estimation and Testing of Euler Equation Models with Time-varying Reduced-form Coefficients," Princeton University.

Lubik, T., and F. Schorfheide (2005): "A Bayesian Look at New Open Economy Macroeconomics," Economics Working Paper Archive 521, The Johns Hopkins University,Department of Economics, available at http://ideas.repec.org/p/jhu/papers/521.html.

Lubik, T. A., and F. Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94(1), 190–217, available at http://ideas.repec.org/a/aea/aecrev/v94y2004i1p190-217.html.

McManus, D. A. (1992): "How common is identification in parametric models?," *Journal of Econometrics*, 53(1-3), 5–23, available at http://ideas.repec.org/a/eee/econom/v53y1992i1-3p5-23.html.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048, available at http://ideas.repec.org/a/ecm/emetrp/v71y2003i4p1027-1048.html.

Neely, C. J., A. Roy, and C. H. Whiteman (2001): "Risk Aversion versus Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model," *Journal of Business & Economic Statistics*, 19(4), 395–403, available at http://ideas.repec.org/a/bes/jnlbes/v19y2001i4p395-403.html.

Onatski, A., and N. Williams (2004): "Empirical and policy performance of a forward-looking monetary model," *Proceedings*, (Mar), available at http://ideas.repec.org/a/fip/fedfpr/y2004imarx6.html.

Pesaran, M. H. (1981): "Identification of rational expectations models," *Journal of Econometrics*, 16(3), 375–398, available at http://ideas.repec.org/a/eee/econom/v16y1981i3p375-398.html.

——— (1989): *The Limits to Rational Expectations*. Basil Blackwell, Oxford.

Rotemberg, J. J., and M. Woodford (1998): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy: Expanded Version," NBER Technical Working Papers 0233, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberte/0233.html.

Rothenberg, T. J. (1966): "Efficient Estimation with a priori Information: A Classical Approach," Cowles Foundation Discussion Papers 205, Cowles Foundation, Yale University, available at http://ideas.repec.org/p/cwl/cwldpp/205.html.

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.

Sargen, D. J. (1988): *Lectures on Advanced Econometric Theory*. Basil Blackwell, Oxford.

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.

Sargent, T. J. (1979): "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure," *Journal of Monetary Economics*, 5, 133–143.

Sbordone, A. M. (2002): "Prices and unit labor costs: a new test of price stickiness," *Journal of Monetary Economics*, 49(2), 265–292, available at http://ideas.repec.org/a/eee/moneco/v49y2002i2p265-292.html.

——— (2005a): "Do expected future marginal costs drive inflation dynamics?," Staff Reports 204, Federal Reserve Bank of New York, available at http://ideas.repec.org/p/fip/fednsr/204.html.

——— (2005b): "A Limited Information Approach to the Simultaneous Estimation of Wage and Price Dynamics," Computing in Economics and Finance 2005 321, Society for Computational Economics, available at http://ideas.repec.org/p/sce/scecf5/321.html.

Schmitt-Grohe, S., and M. Uribe (2004): "Optimal Operational Monetary Policy in the Christiano-Eichenbaum-Evans Model of the U.S. Business Cycle," NBER Working Papers 10724, National Bureau of Economic Research, Inc, available at http://ideas.repec.org/p/nbr/nberwo/10724.html.

Schorfheide, F. (2000): "Loss function-based evaluation of DSGE models," *Journal of Applied Econometrics*, 15(6), 645–670, available at http://ideas.repec.org/a/jae/japmet/v15y2000i6p645-670.html.

Segal, M., and E. Weinstein (1989): "A new method for evaluating the log-likelihood gradient, the Hessian, and the Fisher information matrix for linear dynamic systems.," *IEEE Transactions on Information Theory*, 35(3), 682–.

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 (2005): "Comparing shocks and frictions in US and euro area business cycles: a Bayesian DSGE Approach," *Journal of Applied Econometrics*, 20(2), 161–183, available at http://ideas.repec.org/a/jae/japmet/v20y2005i2p161-183.html.

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

Stock, J. H., and J. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096, available at http://ideas.repec.org/a/ecm/emetrp/v68y2000i5p1055-1096.html.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518–29, available at http://ideas.repec.org/a/bes/jnlbes/v20y2002i4p518-29.html.

Svensson, L. E., and N. Williams (2005): "Monetary policy with model uncertainty: distribution forecast targeting," Discussion paper.

Terceiro, J. (1990): *Estimation of Dynamic Econometric Models with Errors in Variables*. Berlin: Springer-Verlag.

Wallis, K. F. (1980): "Econometric Implications of the Rational Expectations Hypothesis," *Econometrica*, 48(1), 49–73, available at http://ideas.repec.org/a/ecm/emetrp/v48y1980i1p49-73.html.

Wegge, L. L., and M. Feldman (1983): "Identifiability criteria for Muth-rational expectations models," *Journal of Econometrics*, 21(2), 245–254, available at http://ideas.repec.org/a/eee/econom/v21y1983i2p245-254.html.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Zadrozny, P. A. (1989): "Analytic derivatives for estimation of linear dynamic models," *Computers and Mathematics with Applications*, 18(6-7), 539–553.

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.