

Back to square one: identification issues in DSGE models*

Fabio Canova
ICREA, UPF, CREI, AMeN and CEPR

Luca Sala
Istituto di Economia Politica and IGIER - Università Bocconi

First Draft: May 2005
This draft: September 2006

Abstract

We investigate identifiability issues in DSGE models and their consequences for parameter estimation and model evaluation when the objective function measures the distance between estimated and model impulse responses. Observational equivalence, partial and weak identification problems are widespread and they lead to biased estimates, unreliable t-statistics and may induce investigators to select false models. We examine whether different objective functions affect identification and study how small samples interact with parameters and shock identification. We provide diagnostics and tests to detect identification failures and apply them to a state-of-the-art model.

JEL classification numbers: C10, C52, E32, E50

Key words: Identification, Impulse Responses, DSGE models, small samples.

*Among the many who provided us with useful comments and suggestions we would like to thank A. Beyer, T. Cogley, M. Del Negro, J. Faust, P. Feve, J. Lindè, A. Levin, S. Johansen, J. Nason, F. Schorfheide, J. Stock, M. Villani, T. Zha and the participants of seminars at Bocconi, CREI-UPF, Warwick, UPenn, the Atlanta Fed, the Board, the Riksbank, Johns Hopkins, the ECB and at conferences (ESSIM, Cyprus; New methods for empirical macroeconomics, Oslo; Practical issues in DSGE modelling at Central Banks, Helsinki, the fourth Hydra Macro conference, Hydra; the SCE conference, Cyprus, International Macroeconomic Conference, Valencia).

1 Introduction

The 1990's have seen a remarkable development in the specification of DSGE models. The literature has added considerable realism to the constructions popular in the 1980's and a number of shocks and frictions have been introduced into first generation RBC models driven by technological disturbances. Steps forward have also been made in comparing the models' approximation to the data: while 10 years ago it was standard to calibrate the parameters of a model and informally evaluate the quality of its fit, now maximum likelihood or Bayesian estimation of the structural parameters is common both in academic and policy circles (see e.g. Smets and Wouters (2003), Ireland (2004), Canova (2004), Rubio and Rabanal (2005), Gali and Rabanal (2005)) and new techniques have been introduced for evaluation purposes (see Del Negro et. al. (2005)).

Given the complexities involved in estimating state-of-the-art DSGE models and the difficulties in designing criteria which are informative about their discrepancy to the data, a strand of the literature has considered less demanding limited information methods and focused on whether the model matches the data only along certain dimensions. Following Rotemberg and Woodford (1997) and others, it is now common to estimate structural parameters by quantitatively matching conditional dynamics in response to certain structural shocks (Canova (2002) proposes an alternative limited information approach where only a qualitative matching of responses is sought). One crucial but often neglected condition needed for any methodology to deliver sensible estimates and meaningful inference is the one of identifiability: the objective function must have a unique minimum and should display "enough" curvature in all relevant dimensions. Since dynamic responses depend nonlinearly on the structural parameters, it is unknown if these identifiability conditions are met and far from straightforward to check for them in practice.

This paper investigates identifiability issues in DSGE models and explores their consequences for parameter estimation and model evaluation when the objective function measures the distance between impulse responses obtained from a structural VAR and from a model. While the approach we consider falls into the class of moment estimators, and results about the interaction between identification and estimation exist in the literature (see e.g. Choi and Phillips (1992), Stock and Wright (2000)), to the best of our knowledge we are the first to address these issues in the context of DSGE models. Our special interest in impulse response matching is motivated by the popularity

of the technique among applied researchers and the fact that several peculiarities of the procedure make standard theoretical conclusions inapplicable. Our work investigates identification in frameworks commonly used in modern macroeconomics; examines its consequences for structural parameter estimation; and provides simple diagnostics to detect problems in practice.

We start in Section 2 discussing the generics of identification and providing definitions for several practically relevant situations. Section 3 provides examples of simple structures generating four commonly encountered problems: observational equivalence; under-identification; partial and weak identification; and limited information identification problems. In the context of these examples we study: (a) the consequences of altering the weights responses receive in the objective function and the number of variables considered in the analysis; (b) whether and in what way different objective functions provide different identification of the parameters; (c) whether higher order solution techniques necessarily improve parameter identifiability, (d) what features of the economic environment are potentially responsible for the problems. It turns out that all identification problems lead to objective functions with large flat surfaces in the economically reasonable portion of the parameter space; that identification depends on the objective function used, and that Bayesian methods, if properly employed, can help to detect identification problems but, if improperly used, may cover them up. We also show that the common practice of fixing some of the troublesome parameters at arbitrary values, may create distortions in the distribution of parameter estimates and that identification failures are not necessarily reduced when higher order solution techniques are employed contrary, for example, to the likelihood based conclusions of An and Schorfheide (2005). Finally, we find that these problems emerge when the law of motion of states of the model is relatively insensitive to variations in certain parameters. Hence parameter identification, in practice, depends on the structure of the model, the solution technique, the objective function and the type of information used.

Section 4 investigates the interaction between parameters' identifiability, shock identification and small samples. We argue in the context of a three equation New-Keynesian model that many structural parameters are only partially or weakly identifiable from impulse responses and that limited information identification problems are present. Our results suggest that flat objective functions lead to serious biases in large sample estimates and uninformative standard errors, and that small samples and incorrect shock identification pile up to induce major distortions in parameter estimates.

Section 5 examines what happens when the model is unknown and an investigator uses the dynamic implications of a small number of shocks to find estimates of the parameters. We are interested in examining the case in which, because of near-observational equivalence of alternative economic structures, an investigator may end up estimating as significant features which do not appear in the data generating process. In the context of a state-of-the-art model with real and nominal frictions, we demonstrate that many of the features introduced to generate endogenous persistence are only very weakly identified. Hence, investigators using responses to monetary and/or technology shocks could mistakenly select the wrong model with high degree of confidence.

Section 6 presents simple diagnostics to detect identification problems and uses them to highlight why problems in the model used in section 5 emerge. Section 7 summarizes the results and provides suggestions for empirical practice.

Chari, et. al. (2005), Christiano, et. al. (2006) and Fernandez-Villaverde et. al (2005) have recently studied invertibility problems in DSGE models and the ability of structural VARs to recover deep parameters and the dynamics in response to shocks. One interpretation of their evidence is that when invertibility problems are present, the empirical strategy of matching impulse responses is potentially flawed. Our work indicates that even when invertibility problems are absent, matching impulse responses may be problematic and inference erratic because of widespread identification problems. In this sense, the issues we address complement those brought to light in this literature but, given their generality, appear to be more relevant in practice.

2 A few definitions

Identification problems have been extensively studied in theory; the literature on this issue goes back at least to Koopmans (1950), and more recent contributions include Rothenberg (1971), Pesaran (1981), and Hsiao (1983). While the theoretical concepts are relatively straightforward, it is uncommon to see these issues explicitly considered in empirical analyses.

To set ideas, identification has to do with the ability to draw inference about the parameters of a theoretical model from an observed sample. There are several reasons that may prevent researchers to perform such an exercise. First, the mapping between structural parameters and reduced form statistics may not be unique. Hence, different structural models having potentially different economic interpretations may be indis-

tinguishable from the point of view of the chosen objective function. We call this issue *observational equivalence* problem. Second, the population objective function may be independent of certain structural parameters - a structural parameter may disappear from a log-linearized solution. We call this issue *under-identification* problem. A special case of this phenomenon emerges when two structural parameters enter the objective function only proportionally, making them separately unrecoverable. This phenomenon, well known in traditional systems of simultaneous linear equations, is called here *partial identification* problem. Third, even though all parameters enter the objective function independently and the population objective function has a unique minimum, identification problems may emerge because only a subset of the model's implications are used. We name this situation *limited information identification* problem. Fourth, even though all parameters enter the objective function independently and the population objective function has a unique minimum, its curvature may be "insufficient". We call this phenomenon *weak identification* problem. One interesting special case arises when the objective function is asymmetric in the neighborhood of the minimum and its curvature deficient only in a portion of the parameter space.

We formalize the above concepts as follows. Let m_0 be the true model, θ_0 be a $h_0 \times 1$ vector of true parameters and θ_i be a $h_i \times 1$ vector parameters of model m_i , $i = 1, 2, \dots$. Let y^T be a sample of data of length T and let the objective function be $g(y^T; m_i, \theta_i, W) = (irf(y^T) - irf^m(m_i, \theta_i))'W(irf(y^T) - irf^m(m_i, \theta_i))$, where $irf(y^T)$ is a $k \times 1$ vector of data-based impulse responses, $irf^m(m_i, \theta_i)$ is a $k \times 1$ vector of impulse responses obtained with model m_i and W is a weighting matrix. A (minimum distance) estimator for θ_i is defined as $\hat{\theta}_i(W) \equiv \arg \min_{\theta_i(W)} g(y^T; m_i, \theta_i, W)$. Furthermore, $g(y^T; m_i, \theta_i(W), W) \geq 0$ with equality holding if and only if $m_i = m_0$.

The identification problems we are interested in can be formulated as follow:

- *Observational equivalence between two models.* Two models m_1 and m_2 with parameters θ_1 and θ_2 are observationally equivalent if $g(y^T; m_1, \hat{\theta}_1(W), W) = g(y^T; m_2, \hat{\theta}_2(W), W)$, all y^T .

The next set of definitions refer to the local properties of the objective function. For global ones, simply let $\Theta_1 = \Theta$.

- *Observational equivalence of two parameter vectors, given a model.* Two parameter vectors $\hat{\theta}_1(W) \in \Theta_1$ and $\hat{\theta}_2(W) \in \Theta_1$ are observationally equivalent, given m_1 , if $g(y^T; m_1, \hat{\theta}_1(W), W) = g(y^T; m_1, \hat{\theta}_2(W), W)$ and for any other $\theta \in \Theta_1$, $g(y^T; m_j, \hat{\theta}_j, W) < g(y^T; m_j, \theta, W)$, all $y^T, j = 1, 2$.

- *Under-identification of the elements of a parameter vector, given a model* : If for some $\theta = [\theta_1, \theta_2] \in \Theta_1^1 \times \Theta_1^2 = \Theta_1$, $g(y^T; m_1, [\theta^1 \ \theta^2], W) = g(y^T; m_1, [\cdot, \theta^2], W)$ for all $\theta^1 \in \Theta_1^1$, and all y^T , then θ^1 is under-identified.

- *Partial identification of the elements of a parameter vector, given a model*: If for some $\theta = [\theta_1, \theta_2] \in \Theta_1^1 \times \Theta_1^2 = \Theta_1$, $g(y^T; m_1, [\theta_1, \theta_2], W) = g(y^T; m_1, f(\theta_1, \theta_2), W)$ for all y^T and for all $\theta_1 \in \Theta_1^1$ and $\theta_2 \in \Theta_1^2$, where f is some continuous function, then θ_1 and θ_2 are partially identified.

- *Limited information identification*: If we can write $W = \mathcal{S}W$, where \mathcal{S} is a selection matrix with ones on some elements of the main diagonal and zero everywhere else in any of the above definitions, then observational equivalence, under and partial identification are produced by limited informations approaches.

- *Weak identification*: Weak identification of some of the components of θ_1 occurs if there exists a $\hat{\theta}_1(W)$ such that $g(y^T; m_1, \hat{\theta}_1(W), W) < g(y^T; m_1, \theta, W)$ for all y^T and all $\theta \neq \hat{\theta}_1(W) \in \Theta_1$. However, $|g(y^T; m_1, \hat{\theta}_{1i}(W), W) - g(y^T; m_1, \theta_{1i}, W)| < \epsilon$ for some $\theta_{1i} \neq \hat{\theta}_{1i}(W) \in \Theta_1$, $i = 1, 2, \dots, h_1$.

- *Asymmetric Weak identification*: Asymmetric weak identification is present if there exists a $\hat{\theta}_1(W)$ such that $g(y^T; m_1, \hat{\theta}_1(W), W) < g(y^T; m_1, \theta, W)$ for all y^T and all $\theta \neq \hat{\theta}_1(W) \in \Theta_1$. However $|g(y^T; m_1, \hat{\theta}_{1i}(W), W) - g(y^T; m_1, \theta_{1i}, W)| < \epsilon$ for some $\theta_{1i} > \hat{\theta}_{1i}(W) \in \Theta_1$, or for some $\theta_{1i} < \hat{\theta}_{1i}(W) \in \Theta_1$ $i = 1, 2, \dots, h_1$.

3 Identification problems in DSGE models

This section provides a few examples intended to show (a) the pervasiveness of identification problems in DSGE models, (b) the consequences of using limited information approaches for parameter identification, (c) the advantages/disadvantages of employing different objective functions, (d) the relative informational gain obtained using higher order solution methods.

3.1 Observational equivalence: Structural models have the same impulse responses.

Consider the following three models:

$$y_t = \frac{1}{\lambda_2 + \lambda_1} E_t y_{t+1} + \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} y_{t-1} + v_t \quad (1)$$

$$y_t = \lambda_1 y_{t-1} + w_t \quad (2)$$

$$y_t = \frac{1}{\lambda_1} E_t y_{t+1} \text{ where } y_{t+1} = E_t y_{t+1} + e_t \quad (3)$$

where $\lambda_2 \geq 1 \geq \lambda_1 \geq 0$ and v_t , w_t and e_t are iid processes with zero mean and variance σ_v^2 , σ_w^2 , σ_e^2 respectively. It is well known that the unique stable rational expectations solution of (1) is $y_t = \lambda_1 y_{t-1} + \frac{\lambda_2 + \lambda_1}{\lambda_2} v_t$ and that the stable solution of (3) is $y_t = \lambda_1 y_{t-1} + e_t$. Therefore, if $\sigma_w = \sigma_e = \frac{\lambda_2 + \lambda_1}{\lambda_2} \sigma_v$, a unitary impulse in the three innovations will produce the same responses of y_{t+j} , $j = 0, 1, \dots$

What makes the three processes equivalent in terms of impulse responses? Clearly, the unstable root λ_2 in (1) enters the solution only contemporaneously. Since the variance of the shocks is not estimable from normalized impulse responses (any value simply implies a proportional increase in all the elements of the impulse response function), we can arbitrarily select so as to capture the effects of the unstable root.

While the models in (1)-(3) are stylized, it should be kept in mind that many refinements of currently used DSGE models add some backward looking component to a standard forward looking one and that the current debate about the great inflation moderation in the US relies on the existence of determinate vs. indeterminate solutions to explain the evidence. What this example suggests is that these features may be indistinguishable when one looks at impulse responses. Therefore, information external to the models needs to be brought in to disentangle various structural representations (see Lubik and Schorfheide (2004), An and Schorfheide (2005) and Nason and Smith (2005) for similar examples). Note that the equivalence results presented here are the basis for Beyer and Farmer's (2004) claim that the data cannot distinguish whether a Phillips curve is backward looking or forward looking and are the cornerstone of Pesaran's (1981) critique of tests of rational vs. adaptive expectations models.

Examples of more complicated setups which produce observationally equivalent structures appear in Ma (2002), Kim (2003), Altig et. al. (2004) and Ellison (2005).

3.2 Under-identification: Structural parameters disappearing from impulse responses.

Consider the following three equations model:

$$y_t = a_1 E_t y_{t+1} + a_2 (i_t - E_t \pi_{t+1}) + c_1 + v_{1t} \quad (4)$$

$$\pi_t = a_3 E_t \pi_{t+1} + a_4 y_t + c_2 + v_{2t} \quad (5)$$

$$i_t = a_5 E_t \pi_{t+1} + c_3 + v_{3t} \quad (6)$$

where y_t is the output gap, π_t the inflation rate, i_t the nominal interest rate and c_1, c_2, c_3 are constants. The first equation is a forward looking IS curve, the second a forward looking Phillips curve and the third characterizes monetary policy. Since there are no states, the solution is a linear in v_{jt} , $j = 1, 2, 3$ and given by:

$$\begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} \frac{a_2 c_3 + c_1 - a_3 a_2 c_3 - a_3 c_1 + c_2 a_5 a_2 - c_2 a_2}{1 - a_3 - a_1 + a_1 a_5 - a_4 a_5 a_2 + a_4 a_2} \\ \frac{-c_2 + a_1 c_2 - a_4 a_2 c_3 - a_4 c_1}{1 - a_3 - a_1 + a_1 a_5 - a_4 a_5 a_2 + a_4 a_2} \\ \frac{a_5 c_2 + c_2 a_5 a_1 - a_5 a_4 c_1 - c_3 + c_3 a_3 + c_3 a_1 - c_3 a_1 a_3 - a_2 c_3 a_4}{1 - a_3 - a_1 + a_1 a_5 - a_4 a_5 a_2 + a_4 a_2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & a_2 \\ a_4 & 1 & a_2 a_4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix}$$

A few useful points can be made. First, the parameters a_1, a_3, a_5 disappear from the dynamics in response to shocks. Interestingly, they are those characterizing the forward looking dynamics of the model. Second, different shocks carry different information for the parameters: for example, responses to v_{1t} allow us to recover only a_4 ; while responses to v_{2t} have no information for either a_4 or a_2 . Similarly, responses of different variables to each shock carry different information about the structural parameters. Third, different objective functions may have different information about the parameters. For example, if in addition to the dynamics in response to shocks one also considers steady state information (the constant in the solution), one could have some hope of identifying some of the missing parameters. Nevertheless, it should be clear that since six unknown parameter enter the 3×1 vector of constants, not all the parameters will be identifiable even in this latter case.

What do we learn from this example? First, the dynamics of the model may not contain information about certain parameters of interest. Second, while appropriately choosing the objective function may reduce identification problems, there is no guarantee that it will solve them. Third, matching responses to a limited number of shocks may exacerbate identification problems.

3.3 Partial and weak identification

The situations considered in the two previous examples are, in a way, pathological. In practice, there are less extreme but equally interesting settings where the population objective function (locally) has a unique minimum and no parameter disappears, but identification problems may still emerge.

To show this we use a standard RBC structure. We work with the simplest version of the model since we can study whether and how structural parameters affect the impulse responses and therefore highlight both the problems and the reasons for their occurrence. The social planner maximizes $E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\phi}}{1-\phi}$ and the resource constraint is $c_t + k_t = k_{t-1}^\eta z_t + (1-\delta)k_{t-1}$, where c_t is consumption and ϕ is the risk aversion coefficient, z_t is a first order autoregressive process of with persistence ρ , steady state value z^{ss} and variance σ_e^2 , k_{t-1} is the current capital stock, η is the share of capital in production and δ the depreciation rate of capital. The parameters of interest are $\theta = [\beta, \phi, \delta, \eta, \rho, z^{ss}]$. Using the method of undetermined coefficients and letting output be $y_t \equiv k_{t-1}^\eta z_t$, the solution for $w_t = [z_t, k_t, c_t, y_t, r_t]$, in log-deviations from the steady state, is of the form $Aw_t = Bw_{t-1} + Ce_t$ where:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -v_{kz} & 0 & 0 & 0 & 0 \\ -v_{cz} & 0 & 0 & 0 & 0 \\ -v_{yz} & 0 & 0 & 0 & 0 \\ -v_{rz} & 0 & 0 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} \rho & 0 & 0 & 0 & 0 \\ 0 & v_{kk} & 0 & 0 & 0 \\ 0 & v_{ck} & 0 & 0 & 0 \\ 0 & v_{yk} & 0 & 0 & 0 \\ 0 & v_{rk} & 0 & 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$v_{kk} = \frac{1}{2}\gamma - \sqrt{(\frac{1}{2}\gamma)^2 - \beta^{-1}}$; $v_{kz} = \frac{(1-\beta(1-\delta))\rho - \phi(1-\rho)\frac{y^{ss}}{c^{ss}}}{(1-\beta(1-\delta))(1-\eta) + \phi v_{ck} + \phi(1-\rho)\frac{k^{ss}}{c^{ss}}}$; $v_{ck} = (\beta^{-1} - v_{kk}) \frac{k^{ss}}{c^{ss}}$; $v_{cz} = \frac{y^{ss}}{c^{ss}} - \frac{k^{ss}}{c^{ss}} v_{vk}$, $\gamma = \frac{(1-\beta(1-\delta))(1-\eta)(1-\beta+\beta\delta(1-\eta))}{\phi\eta\beta+\beta^{-1}+1}$ and the superscript ss indicates steady states values. We choose $\beta = 0.985$, $\phi = 2.0$, $\rho = 0.95$, $\eta = 0.36$, $\delta = 0.025$, $z^{ss} = 1$ and use the model solution to construct "true responses". To show the features of the objective function, we compute the distance between the "true responses" and the responses obtained varying one or two parameters at a time in an economically reasonable neighborhood of the selected values. Twenty equally weighted responses of $x_t = [k_t, c_t, y_t, r_t]$ are used in computing the objective function.

The first row of Figure 1 presents (the negative of) two of these three-dimensional surfaces and the corresponding contour plots. Each point on the graph gives the distance between the responses of x_t given the true parameters and those obtained with the true parameters except for (ϕ, ρ) or (δ, β) , which take the values on the horizontal

axes. While there is a unique minimum in correspondence of the true parameter vector, the objective function has approximately similar height when the depreciation rate δ and the discount factor β , run from $(\delta = 0.005, \beta = 0.975)$ up to $(\delta = 0.03, \beta = 0.99)$, indicating that the two parameters could only be partially identifiable. Interestingly the 0.01 contour includes the whole range of economically interesting values of these two parameters. In addition, the objective function is quite flat in some dimensions. For example, the persistence parameter ρ is weakly identified in the interval $[0.8, 1.0]$. These are not isolated cases: the share of capital η is also only weakly identifiable in the range $[0.3, 0.6]$ and another ridge appears plotting the objective function against z^{ss} and the depreciation rate δ .

Given our solution, we can check which element of A and B is responsible for this state of affairs. It turns out that the objective function is flat in ρ because the dynamics of the capital stocks are only weakly influenced by this parameter. Since the law of motion of the capital stock determines the dynamics of c_t, y_t, r_t , their responses carry little additional information about this parameter. The local derivatives of v_{kk} and v_{kz} with respect to β and δ have similar magnitude but opposite sign. Hence, the dynamics of the capital stock are also roughly insensitive to proportional changes in these two parameters.

The distance surface plotted in the first row of figure 1 uses the full vector of responses and equally weight responses at all horizons. Would its shape change if, say, only consumption and output responses were used, or responses were weighted by $1/h^2$, where $h = 1, \dots, 20$? In the first setup one would expect some loss of information relative to the baseline case; the question is how large the loss is. In the second setup, the outcome is unclear: identifiability could improve if information in long horizons is noisy or worsen because cross horizon restrictions are partially neglected. The second and third rows of Figure 1 shows that both choices lead to a uniform loss of curvature in the objective function but to minor shape changes. Therefore, cross equation and cross horizons restrictions do help with the identification of these parameters.

One may wonder if matching the coefficients of the D matrix in the VAR(1) representation : $w_t = Dw_{t-1} + v_t$, where $D = A^{-1}B$ and $v_t = A^{-1}Ce_t$, as suggested by Smith (1993), would help in the identification purposes. Intuitively, this choice could be beneficial because shocks' identification is entirely sidestepped, but could also be detrimental since information present in v_t is neglected. The fourth row of Figure 1 indicates that the latter dominates.

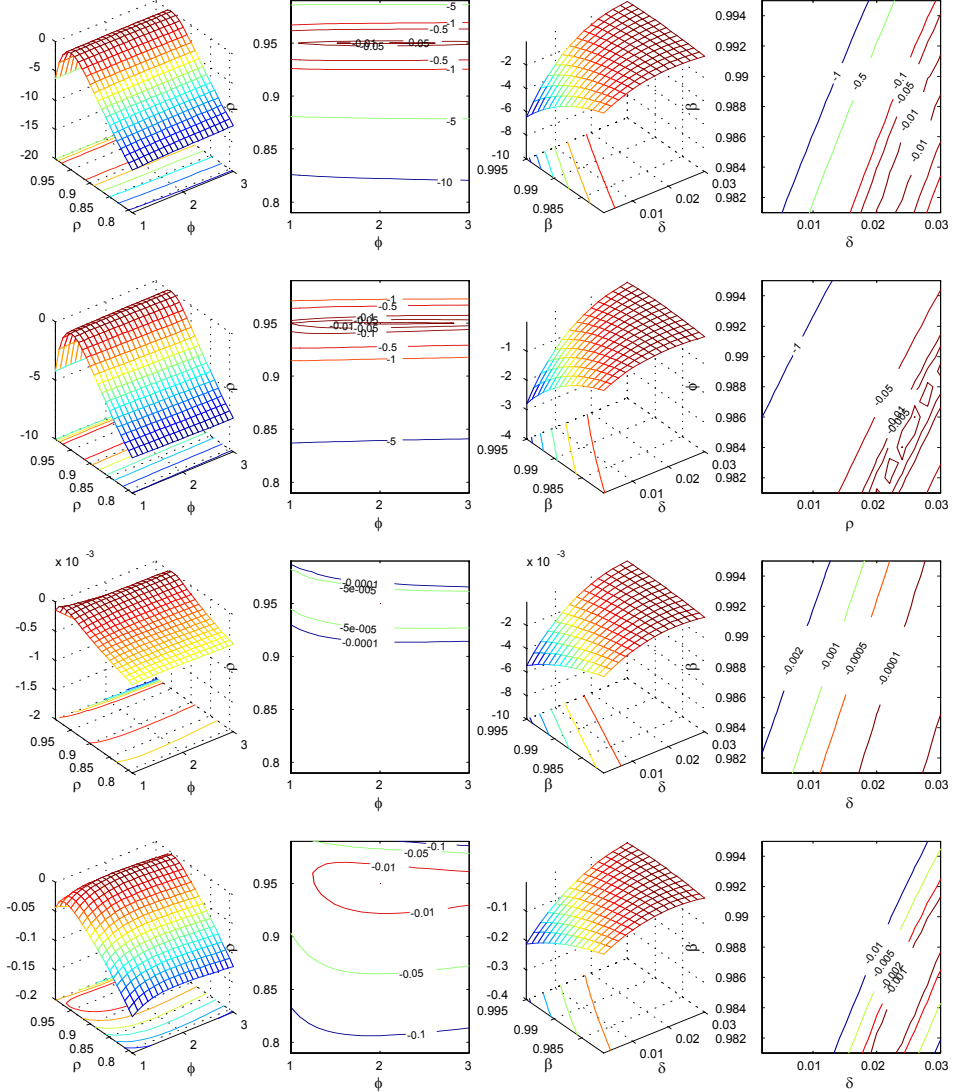
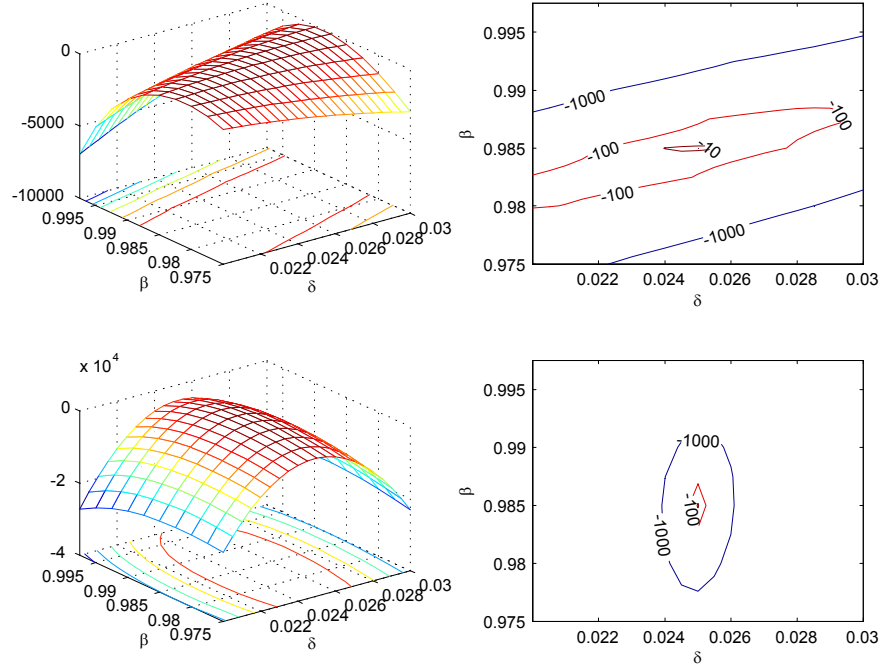


Figure 1: Distance surfaces and contour plots; basic, subset, weighted and matching VAR

For empirical purposes, it is important to know whether identification problems depend on the objective function or are intrinsic to the model, in which case the choice of objective function is irrelevant. To distinguish between these two alternative we have examined the shape of the likelihood of the model, computed by generating 250

observations¹ and assuming z_t to be normally distributed with $\sigma_z = .001$. Since model misspecification is not an issue here, the likelihood function provides a natural upper "identification bound" of the parameters. If the likelihood function displays identification problems, we cannot hope to do better by using limited information approaches. Having a well-behaved likelihood is thus a necessary, but not sufficient condition for proper estimation².

In general, identification problems seem less acute when the likelihood function is used: there are some flat areas but contour plots are much better behaved and, for example, β and δ can be pinned down with much higher precision (see top panel in figure 2). Hence, at least in the context of this model, partial and weak identification problems are to a large extent related to the choice of objective function.



Because the likelihood function of DSGE models is typically ill-behaved, it has now

¹As the economic model is stochastically singular, we have added normally distributed measurement error to each series in order to be able to compute the likelihood function.

²The discussion here excludes the possibility that particular frequencies of the spectrum carry special information about the parameters, in which case the likelihood function of appropriately filtered data may be have better identification properties than the likelihood function of the true data.

became common to employ Bayesian methods for estimation. Given the recent emphasis, a few words contrasting identification in classical and Bayesian frameworks, are in order. Posterior distributions are proportional to the likelihood times the prior. If the parameter space is variation free, that is, there are no implicit constraints on combinations of parameters, the likelihood of the data carries information for the parameters if the prior and posterior have different features (see Poirier (1998)). When this is not the case, there is a simple diagnostic for detecting lack of identification. If prior information becomes more and more diffuse, the posterior of parameters with doubtful identification features will also become more and more diffuse. Hence, using a sequence of prior distributions with larger and larger variances one may detect potential problems ³. Nevertheless, since identification problems have to do with the shape of the likelihood, they do not disappear when a Bayesian approach is employed.

When the parameter space is not variation free, e.g. because stability conditions or economically motivated non-negativity constraints are implicitly imposed, the prior of non-identified parameters may be marginally updated because such restrictions may make the likelihood informative. In this case, finding that prior and posterior differ it does not guarantee that parameters are identified ⁴. If one uses a sequence of prior distributions with increasing spreads, one can still detect potential identification problems. Unfortunately, this simple diagnostic is hardly ever used and often prior distributions are not even reported. This is dangerous: the combination of a tightly specified prior and auxiliary restrictions on the parameter space can in fact produce a well behaved posterior even if the economic model per se contains little information on the parameter of interest. The second row of figure 2 show that this can happen: here a tight prior on δ generates a lot of curvature in the posterior distribution. Hence, uncritical use of Bayesian methods, including employing prior distributions which do not truly reflect the location and spread uncertainty, may hide identification problems rather than highlighting them.

What do one typically do when partial identification problems emerge? The standard practice of fixing β will work here since for any value of β , the impulse based objective function has reasonable curvature in the δ dimension (and viceversa). How-

³There is no problem in eliciting more and more diffuse prior distributions since the parametrization is given by economic theory.

⁴As Matthias Villani has pointed out to us, the reason for why prior and posterior differ in this case is because the prior is not properly marginalized, that is, the fact that the parameter space is not variation free is neglected.

ever, such an approach may also induce serious biases, unless the chosen β happens to be the right one. We show this graphically in Figure 3, where we report contours plots correctly assuming $\beta = 0.985$ and incorrectly assuming $\beta = 0.995$. In the latter case, the maximum of the constrained objective function shifts away from the true value. In addition, the curvature is accentuated around the wrong value thus giving very precise parameter estimates.

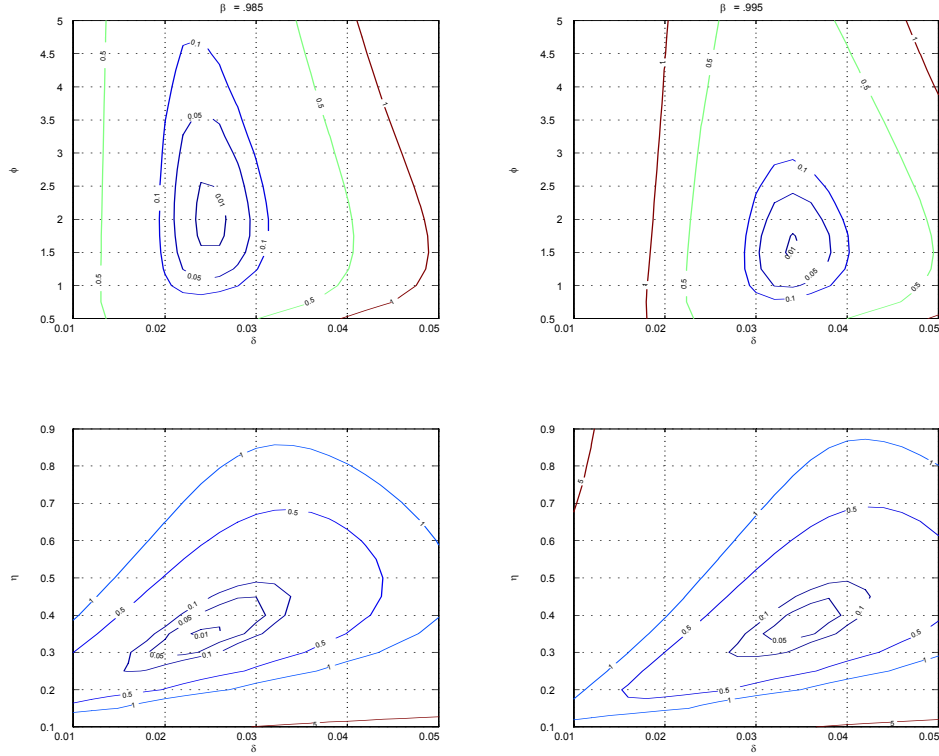


Figure 3: Contour Plots

To conclude, this section suggests four important conclusions. First, for identification purposes, more information is always better than less: neglecting the cross equations restrictions present in variables, in their covariance matrix and/or at long horizons may lead to or exacerbate identification problems. Second, while one could probably be better endowed to answer interesting economic questions if she carefully selects the objective function, identification problems may not vanish if the model is not explicitly parametrized with an eye to estimation. Third, classical and Bayesian

approaches face the same identification problems. Formal use of external and reliable information may give an hedge to the latter in dealing with such problems. Finally, the practice of fixing some parameters and estimating others may lead to important distortions, unless the selected parameters happen to be the true ones.

3.4 Higher order solution methods and identification

Fernandez-Villaverde and Rubio-Ramirez (2005) and An and Schorfheide (2005) have recently shown that the likelihood function appears to have more curvature when a second order approximation is used in place of a log-linear one. In this section we are interested in analyzing whether the distance function obtained by matching responses using a second order solution has better identification features than the one obtained using a first order solution. Two contrasting elements make the outcome unpredictable. First, observation equivalence, under-identification, partial and weak identification may be reduced when higher order terms (and cross equation restrictions) are brought into the problem. Second, since responses in second order system depend on the size and the sign of the shocks and the initial conditions, identification problems could become more acute in this setup.

To illustrate the potential trade-off existing between first and second order solutions when matching impulse responses we take a version of an RBC model with (external) habit in consumption driven by a permanent technology disturbance and a transitory labor supply shock. Lifetime preferences are given by $E_0 \sum_t \beta^t [\log(c_t - b\bar{c}_{t-1}) - a_t N_t]$ where β is a discount factor, b regulates the evolution of consumption habits, \bar{c}_t is the aggregate level of consumption - taken as given by the agents, a_t is a labor supply shock with time series representation $\ln(\frac{a_t}{\rho a_{t-1}}) = u_t^a$, where u_t^a is iid with variance one and E_t denotes the expectation operator, conditional on the information at time t . We assume the production function $y_t = z_t N_t$, where $\ln(\frac{z_t}{z_{t-1}}) \equiv u_t^z$, and u_t^z is iid with variance one, and the resource constraint is $c_t = y_t$.

Detrending the variables by the level of technology, log-linearizing around the steady state, and considering only responses to labor supply shocks we have

$$\hat{N}_t = (b + \rho)\hat{N}_{t-1} - b\rho\hat{N}_{t-2} - (1 - b)\hat{u}_t^a \quad (7)$$

As Sargent (1978) and Kennan (1988) have argued b and ρ are not separately identified from (7) unless $b = \rho$. In fact, the reduced form version of (7) is $\hat{N}_t = \eta_1 \hat{N}_{t-1} - \eta_2 \hat{N}_{t-2} - \eta_3 \hat{u}_t^a$ which has two solutions $b = 0.5(\eta_1 \pm \sqrt{\eta_1^2 - 4\eta_2})$ and $\rho = \eta_1 - b$, where

$\eta_1^2 - 4\eta_2 = (b - \rho^2) \geq 0$. Hence, there are two values of ρ and b consistent with exactly the same dynamics of hours in response to labor supply shocks: a high value of habit and a low value of the persistence of labor supply shocks, and a low value of habit and a high value of the persistence of labor supply shocks. The second order approximation to the equilibrium condition is:

$$\begin{aligned}\hat{N}_t &= b\hat{N}_{t-1} + \frac{b(b-1)}{2}\hat{N}_{t-1}^2 - (1-b)\hat{a}_t - \frac{1}{2}(-(1-b)^2 + 1-b)\hat{a}_t^2 \\ \hat{a}_t &= \rho\hat{a}_{t-1} + u_t^a\end{aligned}$$

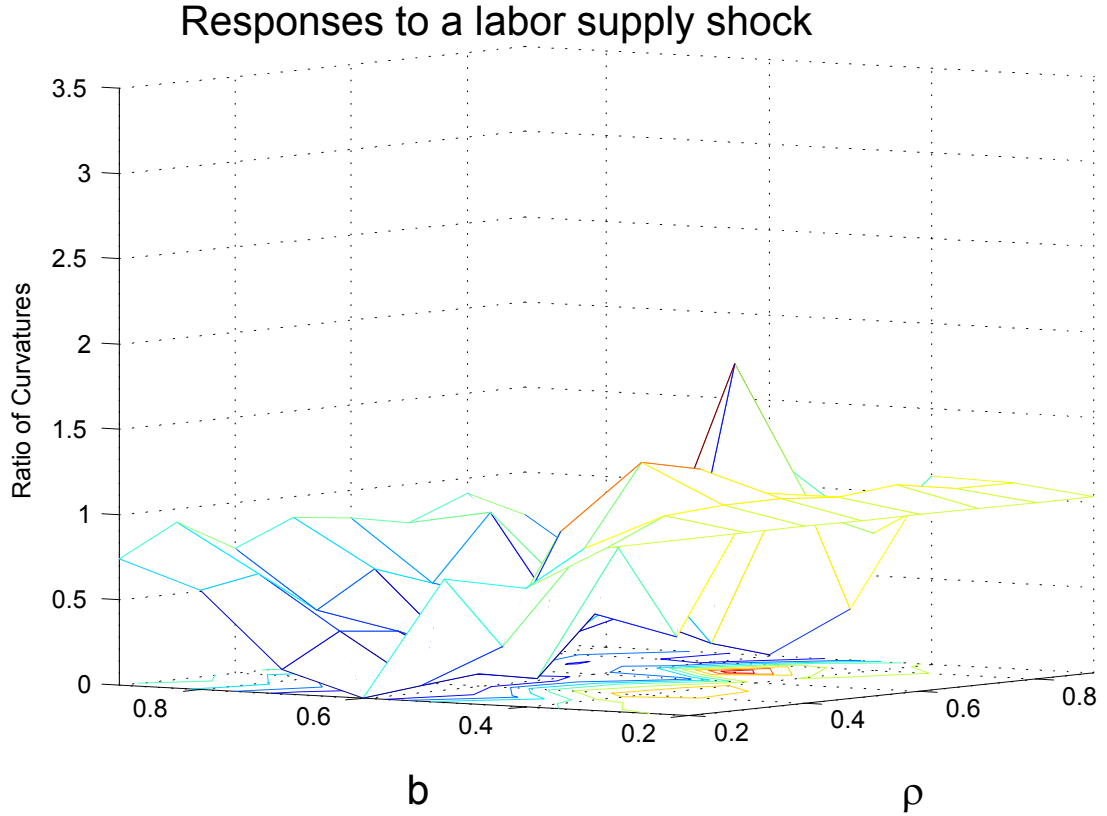


Figure 4: Distance function: ratio of linear to quadratic solution

Figure 4 plots the ratio of the linear to the quadratic distance function when the true parameters are $(b = 0.6, \rho = 0.2)$; twenty equally weighted responses of \hat{N}_t are used to construct the objective functions; and the size of the shock and the initial conditions are both integrated out (see Koop, Pesaran and Potter (1996)). Since a value above one in the vertical scale indicates that the curvature of the linear distance function

is larger than the curvature of the second order one, figure 4 clearly shows that the distance function obtained with a second order approximation is not necessarily better behaved everywhere in the parameter space and that asymmetries could be important.

4 Identification and estimation

Next, we examine what identification problems imply for estimation and inference. Throughout this section we assume that the investigator knows the correct model and, for most of it, assume that no misspecification occurs when computing responses. Initially we endow the researcher with the population responses; later we explore in what way small samples complicate the inferential task.

To make our points transparent, we employ a well known small scale New-Keynesian model. We choose this specification because several authors, including Ma (2002), Beyer and Farmer (2004) and Nason and Smith (2005), have argued that it is liable to some of the problems we have discussed so far. The log-linearized version of the model consists of the following three equations:

$$y_t = \frac{h}{1+h}y_{t-1} + \frac{1}{1+h}E_t y_{t+1} + \frac{1}{\phi}(i_t - E_t \pi_{t+1}) + v_{1t} \quad (8)$$

$$\pi_t = \frac{\omega}{1+\omega\beta}\pi_{t-1} + \frac{\beta}{1+\omega\beta}E_t \pi_{t+1} + \frac{(\phi+1)(1-\zeta\beta)(1-\zeta)}{(1+\omega\beta)\zeta}y_t + v_{2t} \quad (9)$$

$$i_t = \lambda_r i_{t-1} + (1-\lambda_r)(\lambda_\pi \pi_{t-1} + \lambda_y y_{t-1}) + v_{3t} \quad (10)$$

where h is the degree of habit persistence, ϕ is the relative risk aversion coefficient, β is the discount factor, ω the degree of indexation of prices, ζ the degree of price stickiness, while $\lambda_r, \lambda_\pi, \lambda_y$ are policy parameters. The first two shocks follow an AR(1) with parameters ρ_1, ρ_2 , while v_{3t} is iid. The variances of the shocks are denoted by $\sigma_i^2, i = 1, 2, 3$.

The model has 14 parameters: $\theta_1 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ are under-identified from scaled impulse responses, while $\theta_2 = (\beta, \phi, \zeta, \lambda_r, \lambda_\pi, \lambda_y, \rho_1, \rho_2, h, \omega)$ are the structural parameters which are the focus of our attention.

Since the model features three shocks, we can construct several limited information objective functions, obtained considering the responses to only one type of shock, and a full information one. We take the true parameters to be $\beta = 0.985, \phi = 2.0, \zeta = 0.68, \lambda_r = 0.2, \lambda_\pi = 1.55, \lambda_y = 1.1, \rho_1 = 0.65, \rho_2 = 0.65, \omega = 0.25, h = 0.85$, which are standard in calibration exercises and quite close to, e.g., Rabanal and Rubio-Ramirez

(2005) estimates of a version of such a model for the US. Twenty equally weighted responses of the three variables are used to construct the objective function.

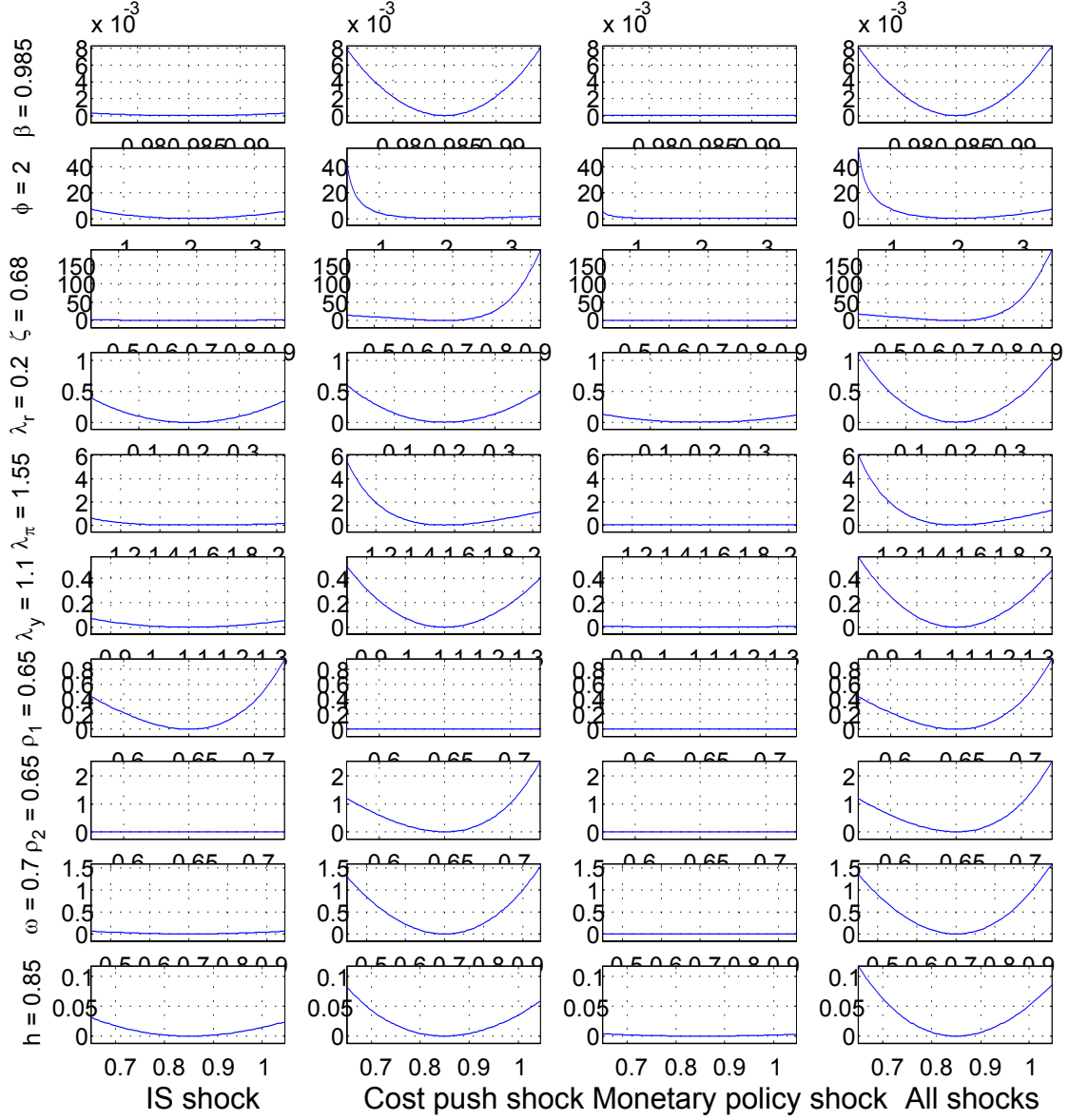


Figure 5: Shape of different objective functions

Figure 5 plots the shape of the objective function in each of the elements of θ_2 . Column 1 presents the distance function obtained using responses to v_{1t} , column 2 the

one obtained using responses to v_{2t} , column 3 the one obtained using responses to v_{3t} and column 4 the one obtained with all the shocks. In each case we vary one parameter at the time within the range presented in the x-axis keeping the others fixed at their "true" values.

The figure displays many interesting features. First, the three limited information objective functions are flat in several dimensions (see e.g. $\lambda_r, \lambda_y, \omega, h$). Second, different shocks have different information about certain parameters (see e.g. ζ, λ_r). Third, the objective functions are asymmetric in certain dimensions see, for example, ϕ, λ_r, ζ when cost push shocks are considered. Fourth, there are parameters which are under-identified by certain shocks: as intuition suggests, the persistence of, say, the cost push shock can not be identified considering responses to other shocks. Fifth, even when responses to all shocks are used, the objective function is still somewhat flat and asymmetric in several dimensions. It is important to stress that these features are independent of the exact values of θ_1 and θ_2 used (see Canova and Sala (2005) for details).

Since Figure 5 considers one dimension at the time, ridges in the objective function are not detectable. Figure 6 shows that indeed ridges are present: responses to monetary shocks carry little information, for example, about the correct combination of λ_y and λ_π . This is a bit surprising: since the policy rule is backward looking, a regression of current interest rates on lagged inflation and lagged output should be able to separately recover λ_π and λ_y . To explain why λ_π and λ_y are poorly identified, note that monetary policy induce small responses in the output gap and inflation. Consequently, responses of interest rate are broadly unaffected by changes in λ_y and λ_π . This is not the case with the other shocks, and the distance function is well behaved when all shocks are used.

In sum, this prototype model displays an array of potential identification problems. Next, we investigate what happens to parameter estimates and to statistical and economic inference in this situation.

4.1 Asymptotic properties

For the sake of presentation, we will focus on estimates obtained matching responses to monetary policy shocks which appear to produce the distance function with the worst identification properties and are those on which the literature has focused most of its attention. Figure 7 reports the histogram of estimates obtained starting the

minimization routine 500 times from different initial conditions uniformly drawn within the ranges considered on the horizontal axis. Superimposed with a vertical bar is the true parameter value.

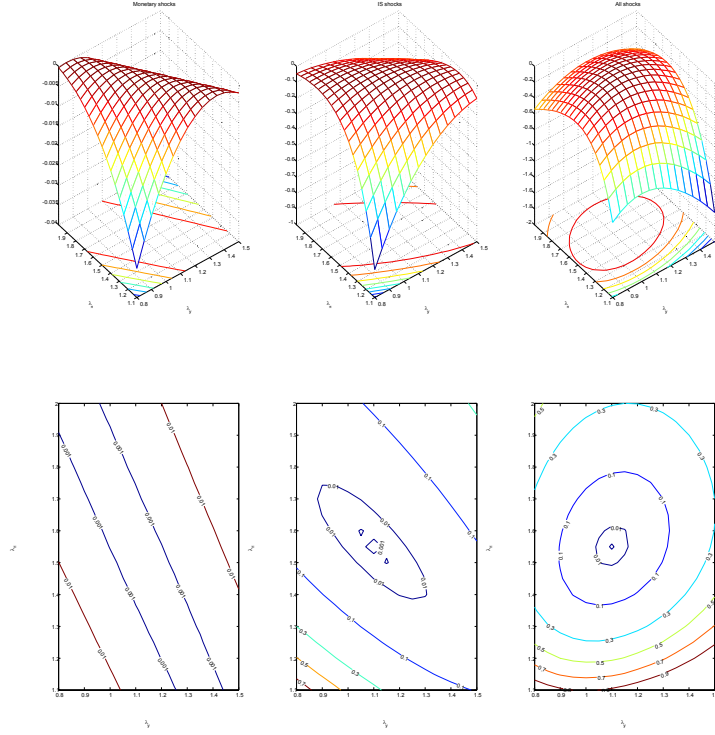


Figure 6: Distance function and contour plots

Histograms are obtained eliminating all cases where convergence failed; or the estimated parameters produce imaginary or indeterminate solutions. The histograms do not capture sampling uncertainty associated with the estimation of structural parameters. Instead, the figure displays the multivariate mapping from impulse responses to structural parameters. If the objective function were free of identification problems, this mapping would be univocal: from any starting point the true vector would be reached and the histograms would all be degenerate ⁵.

⁵Failure to reach the true parameter could of course be the result of a poor minimization routine rather than of identification problems. We have checked for this possibility in a number of ways and found that absent identification problems our routine always finds the true parameter vector from any initial conditions.

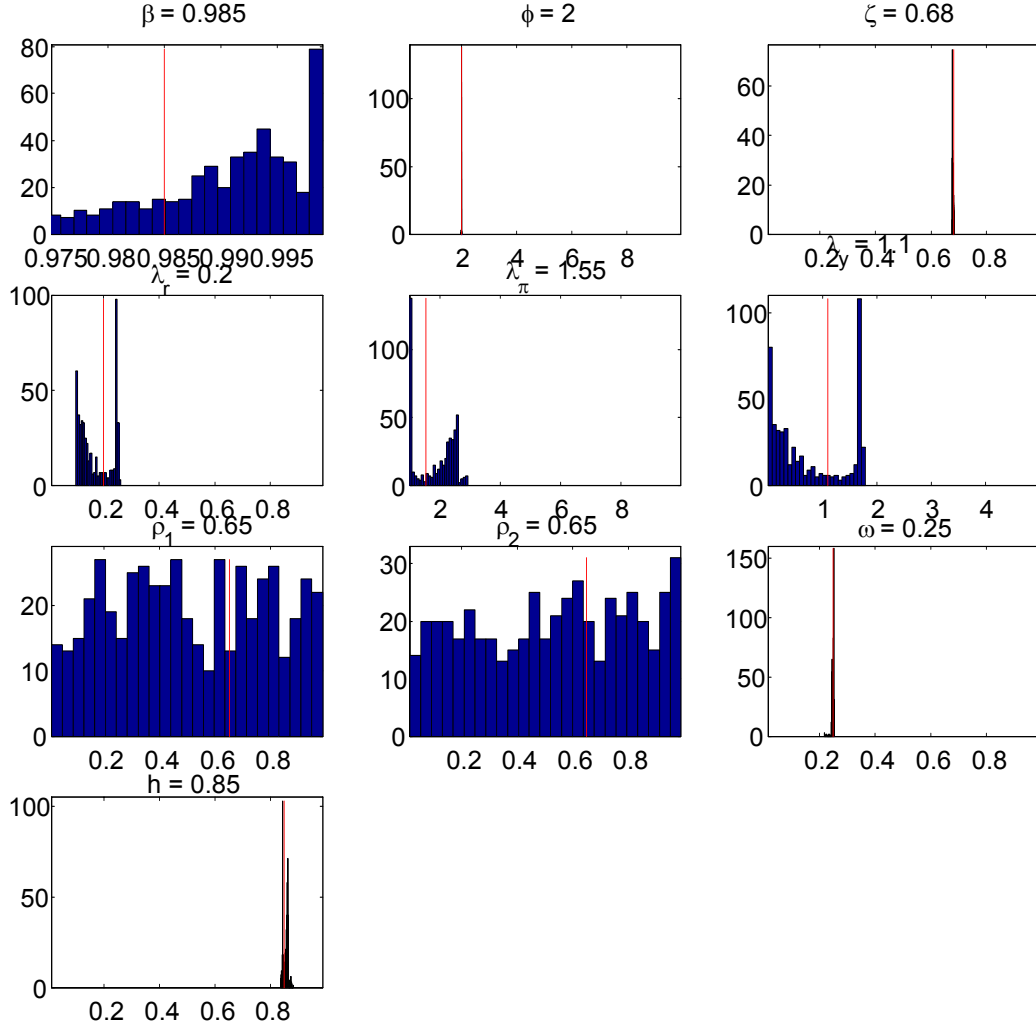


Figure 7: Distribution of estimates

There are three features of figure 7 worth discussing. First, there is a tendency to overestimate β . Second, the mode of the distribution of estimates of λ_π is located at 2.64, well above the true value of 1.55. Third, the histogram of λ_y has two modes, one at around zero and one at 1.85, and the one of λ_r has similar features.

Would it be possible to detect these estimation failures, for example, looking at the minimized value of the objective function or to the resulting impulse responses? The answer is negative. The objective function is within the tolerance level (10^{-7}) for all the parameter combinations generating figure 7 and, as shown in figure 8, population

and estimated responses to monetary shocks are indistinguishable. Interestingly, responses to IS and cost push shocks are also similar to the true ones. Hence, parameter vectors with potentially different economic interpretations are indistinguishable when normalized responses are used to construct objective functions.

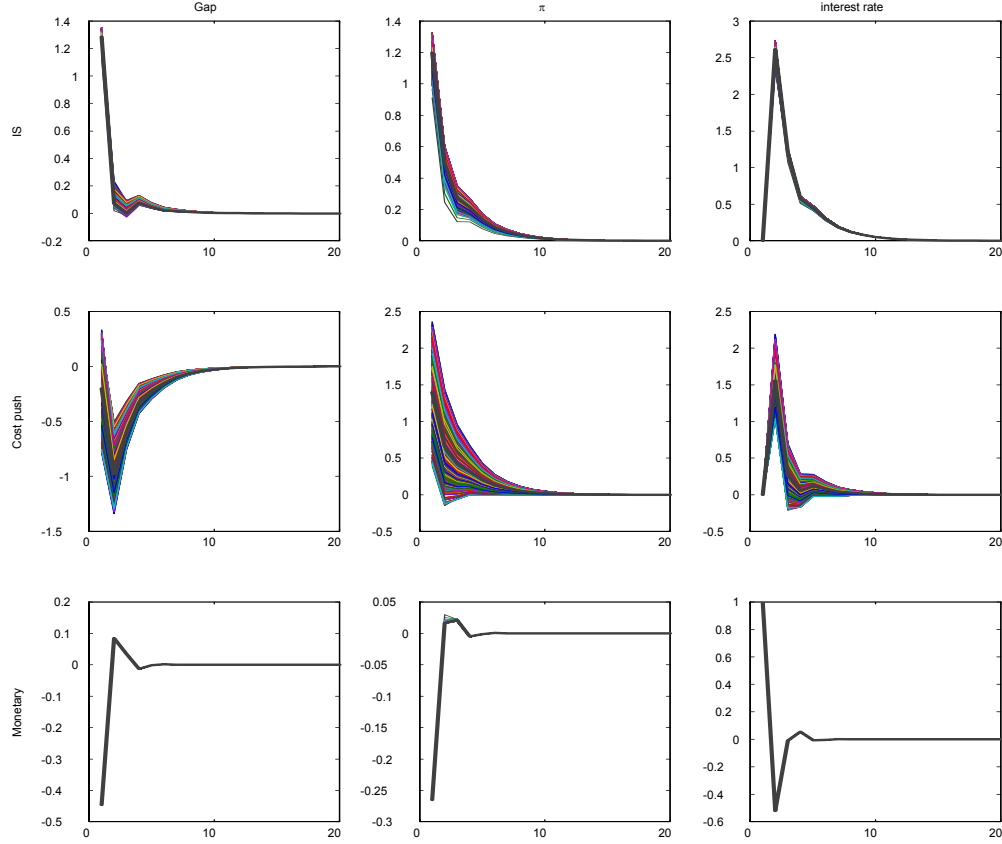


Figure 8: Impulse responses

For forecasting purposes these differences are probably unimportant: as long as the fit and the forecasting performance is the same, the true nature of the DGP does not matter. However, policy analyses and conditional forecasting exercises conducted using the estimated parameter vector may deviate from those obtained with the true one. Hence, it is unwise to attach any economic interpretation to the estimates or draw conclusions about how the economy works from these exercises and this is true even in the ideal situation considered in this subsection.

4.2 Small samples

The distortions present in figure 7 may be magnified when only estimates of impulse responses obtained with samples of small or medium sizes are available. Furthermore, it is conceivable to have situations where the objective function is well behaved but important identification problems emerge just because of small samples. In this subsection we are interested in (a) quantifying the importance of these problems when samples of the size typically used in macroeconomics are employed to compute responses and (b) highlight some of the properties of the estimates of parameters with problematic identification features. We focus again our attention on responses to policy shocks, since the model implies that reduced form interest rate innovations are the true monetary policy shocks. For the majority of this subsection we still assume that the investigator correctly identifies monetary disturbances. Later we examine what happens when shock identification fails. Using the log-linearized solution, we simulate 200 time-series for interest rates, the output gap and inflation for $T = 120, 200, 1000$, estimate an unrestricted VAR⁶ on the simulated data, compute impulse responses and bootstrap confidence bands. We use the resulting confidence bands to build a diagonal weighting matrix: weights are inversely proportional to the uncertainty in the estimates.

Table 1 presents a summary of our estimation results. We report the true parameters, the mean estimate, the numerical standard errors computed across replications (in parenthesis) and the percentage bias (in brackets).

A few features are worth commenting upon. First, biases are evident in the estimates of the partially identified parameters (λ_π, λ_y) , the weakly identified parameters $(\zeta, \omega, h$ and $\lambda_r)$ and the under-identified parameters (ρ_1, ρ_2) . Note that even with 250 years of quarterly data major biases remain. Second, numerical standard errors are large for the partially identified parameters and invariant to sample size for the under-identified ones. Third, parameter estimates do not converge to population values as T increases. Finally, and concentrating on $T = 200$, estimates suggest an economic behavior which is somewhat different from the one characterizing the DGP. For example, it appears that price stickiness is stronger and the Central Bank reaction to the output gap and inflation is equally strong. Once again, armed just with impulse responses, an investigator has little possibility to detect such interpretation problems.

⁶We checked that the VAR is able to correctly estimate the true impulse responses with the correct identification when $T = 5000$.

Table 1: NK model. Matching monetary policy shocks

	True	T = 120	T = 200	T=1000	T=1000 wrong
β	0.985	0.984 (0.007) [0.6]	0.985 (0.007) [0.7]	0.986 (0.008) [0.7]	0.981 (0.004) [0.6]
ϕ	2.00	2.39 (2.81) [95.2]	2.26 (2.17) [70.6]	1.41 (1.19) [48.6]	10 (0) [400]
ζ	0.68	0.76 (0.14) [19.3]	0.76 (0.12) [17.5]	0.83 (0.10) [23.5]	0.84 (0.06) [23.7]
λ_r	0.2	0.47 (0.29) [172.0]	0.43 (0.27) [152.6]	0.41 (0.24) [132.7]	0.02 (0.05) [90.5]
λ_π	1.55	2.60 (1.71) [98.7]	2.22 (1.51) [78.4]	2.18 (1.38) [74.5]	4.92 (0.33) [217.5]
λ_y	1.1	2.82 (2.03) [201.6]	2.56 (2.01) [176.5]	2.16 (1.68) [126.5]	0.67 (0.98) [78.3]
ρ_1	0.65	0.52 (0.20) [30.4]	0.49 (0.21) [34.3]	0.50 (0.19) [31.0]	0.50 (0.19) [31.3]
ρ_2	0.65	0.49 (0.20) [32.9]	0.48 (0.21) [34.8]	0.48 (0.21) [34.7]	0.48 (0.21) [34.7]
ω	0.25	0.76 (0.39) [238.9]	0.73 (0.40) [232.3]	0.65 (0.38) [198.1]	0.92 (0.27) [284.0]
h	0.85	0.79 (0.35) [30.9]	0.76 (0.37) [32.4]	0.90 (0.21) [21.3]	0 (0) [100]

While not very favorable, the results of table 1 are a bit on the optimistic side. Biases can be amplified if, in addition to small samples, shock identification is also subject to errors. We report in the last column of table 1 estimates obtained when $T = 1000$ and monetary shocks are identified wrongly assuming that interest rates contemporaneously responds to the output gap and inflation. Biases are of course evident. More interestingly, standard errors of the estimates are smaller indicating major shifts in the entire distribution of estimates. Since significance of estimates is typically an appreciable feature in applied work, it is possible that an investigator would prefer the (biased) estimates of the last column of the table to the "insignificant" estimates obtained in the case monetary shocks are correctly chosen.

In conclusion, identification problems combined with small samples typically lead to biased estimates of certain structural parameters, to inappropriate inference when conventional asymptotic theory is used to judge the significance of the parameters and, possibly, to wrong economic interpretations. In addition, the practice of showing that model's responses computed using the estimated parameters lie within the confidence bands of responses estimated from the data may be uninformative, as the objective function is close to zero at a variety of different parameter values.

4.3 Linking the results to the literature

Under-identification and weak identification have been recognized to be serious estimation problems. Choi and Phillips (1992), Stock and Wright (2000) have shown the consequences these two phenomena have on the asymptotic properties of estimates in IV and GMM setups. Choi and Phillips showed that under-identification produce asymptotic distributions of estimates which strongly deviate from normal; Stock and Wright that identification problems in GMM frameworks produce inconsistent estimates of weakly or under-identified parameters, that the joint distribution of weakly (or under-identified) and properly identified parameters is non-standard; and that stan-

dard t-statistics are, in general, invalid.

Our results are consistent with many of the theoretical predictions derived by these authors. In particular, we find (i) erratic properties of the estimates of weakly (or under-identified) parameters as T increases; (ii) standard errors which are large and do not necessarily decrease with the sample size; (iii) t-tests which are uninformative about the properties of estimates. However, our results also show that not all the parameters which appeared to be weakly identified from the third column of figure 5, display similar properties, see e.g. ω and h .

Stock and Wright also develop an asymptotic theory which is robust to identification problems. Since our objective function resembles the objective function they use, one may wonder whether identification problems can be sidestepped and distortions eliminated using their approach. Unfortunately, their theory is inapplicable in our case because W is never chosen to be the continuously updating weighting matrix of Hansen et al. (1996). Furthermore, the combination of numerical solutions, large dimensional parameter space and highly non-linear mapping between structural parameters and the objective function renders their theory difficult to implement, even when the appropriate weighting matrix is used.

5 Misspecification and observational equivalence.

The previous section showed that when the investigator knows the model, inference may be difficult as different parameter values are almost equally probable from the point of view of the objective function. When the true model is unknown, one can not a priori exclude that structures with alternative economic features could be equally likely. Since the literature has built-in frictions in standard DSGE models to enhance its fit without caring too much about their identifiability, we want to investigate whether models with different frictions may be indistinguishable when responses to a limited number of shocks are considered and whether it is possible to obtain significant estimates of parameters that are in fact absent from the DGP.

To study this issue we consider a model which is much richer than those employed so far, includes real and nominal frictions, and has been shown to capture reasonably well important features both of the US economy (see Christiano, et al. (2005), Dedola and Neri (2004)) and the EU economy (see Smets and Wouters (2003)). The log linearized model consists of the following 11 equations:

$$\begin{aligned}
0 &= -k_{t+1} + (1 - \delta)k_t + \delta x_t \\
0 &= -u_t + \psi r_t \\
0 &= \frac{\eta \delta}{\bar{r}} x_t + (1 - \frac{\eta \delta}{\bar{r}}) c_t - \eta k_t - (1 - \eta) N_t - \eta u_t - e z_t \\
0 &= -R_t + \lambda_r R_{t-1} + (1 - \lambda_r)(\lambda_\pi \pi_t + \lambda_y y_t) + e r_t \\
0 &= -y_t + \eta k_t + (1 - \eta) N_t + \eta u_t + e z_t \\
0 &= -N_t + k_t - w_t + (1 + \psi) r_t \\
0 &= E_t \left[\frac{h}{1+h} c_{t+1} - c_t + \frac{h}{1+h} c_{t-1} - \frac{1-h}{(1+h)\phi} (R_t - \pi_{t+1}) \right] \\
0 &= E_t \left[\frac{\beta}{1+\beta} x_{t+1} - x_t + \frac{1}{1+\beta} x_{t-1} + \frac{\chi^{-1}}{1+\beta} q_t + \frac{\beta}{1+\beta} e x_{t+1} - \frac{1}{1+\beta} e x_t \right] \\
0 &= E_t [\pi_{t+1} - R_t - q_t + \beta(1 - \delta) q_{t+1} + \beta \bar{r} r_{t+1}] \\
0 &= E_t \left[\frac{\beta}{1+\beta \gamma_p} \pi_{t+1} - \pi_t + \frac{\gamma_p}{1+\beta \gamma_p} \pi_{t-1} + T_p (\eta r_t + (1 - \eta) w_t - e z_t + e p_t) \right] \\
0 &= E_t \left[\frac{\beta}{1+\beta} w_{t+1} - w_t + \frac{1}{1+\beta} w_{t-1} + \frac{\beta}{1+\beta} \pi_{t+1} - \right. \\
&\quad \left. \frac{1+\beta \gamma_w}{1+\beta} \pi_t + \frac{\gamma_w}{1+\beta \gamma_w} \pi_{t-1} - T_w (w_t - \nu N_t - \frac{\varphi}{1-h} (c_t - h c_{t-1}) - e w_t) \right]
\end{aligned}$$

The first equation describes capital accumulation, δ is the depreciation rate, and x_t is current investment; the second equation links capacity utilization u_t to the real rate r_t and ψ is a parameter; the third equation is the resource constraint linking consumption c_t and investment expenditures to output, where \bar{r} is the steady state interest rate and $e z_t$ is a technological disturbance; the fourth equation represents the monetary policy rule and $e r_t$ is a monetary policy disturbance; the fifth equation represents the production function, where η is the capital share; the sixth equation is a labor demand equation, where N_t is hours worked and w_t the real wage rate; the seventh equation is an Euler equation for consumption, where h captures habit persistence, ϕ is the risk aversion coefficient and π_t is the current inflation rate; the eighth equation is an Euler equation for investment, where q_t is Tobin's q, β is the discount factor, χ^{-1} the elasticity of investment with respect to Tobin's q and $e x_t$ an investment shock; the ninth equation describes the dynamics of Tobin's q; the last two equations represent the wage setting and the price setting equations: $\gamma_p(\gamma_w)$ is a price (wage) indexation parameter, $\zeta_p(\zeta_w)$ a price (wage) stickiness parameter, ν is the inverse elasticity of labor supply, $e p_t(e w_t)$ are shocks to the pricing relationships, $T_p \equiv \frac{(1-\beta \zeta_p)(1-\zeta_p)}{(1+\beta \gamma_p) \zeta_p}$ and $T_w \equiv \frac{(1-\beta \zeta_w)(1-\zeta_w)}{(1+\beta)(1+(1+\epsilon_w)\nu \epsilon_w^{-1}) \zeta_w}$, where ϵ_w is a wage markup. The vector of parameters includes the structural ones: $\theta_1 = (\beta, \phi, \nu, h, \delta, \eta, \chi, \psi, \gamma_p, \gamma_w, \zeta_p, \zeta_w, \epsilon_w, \lambda_r, \lambda_\pi, \lambda_y)$ and

the auxiliary ones $\theta_2 = (\rho_z, \rho_x, \sigma_z, \sigma_r, \sigma_p, \sigma_w, \sigma_x)$, where ρ_z, ρ_x represent the persistence of the technology and investment shocks and $\sigma_i, i = 1, \dots, 5$ the standard deviation of the disturbances. As usual σ_i 's are not identified from the normalized responses and the persistence parameters are identified only when own shocks are considered.

This model is sufficiently rich and complicated that it is difficult to know a-priori which parameters are identifiable and which are not. To explore this issue we construct true responses using the posterior mean estimates for the US economy obtained by Dedola and Neri (see table 2) and examine the shape of the distance function in the neighborhood of this vector, one parameter at a time. 20 responses of the 11 variables are used to construct the distance function.

Table 2. Parameter values

$\theta_1 :$	$\beta = 0.991$	$\phi = 3.014$	$\nu = 2.145$	$h = 0.448$	$, \theta_2 :$	$\rho_z = 0.997$	$\sigma_p = 0.221$
	$\delta = 0.0182$	$\eta = 0.209$	$\chi = 6.300$	$\psi = 0.564$		$\rho_x = 0.522$	$\sigma_w = 0.253$
	$\gamma_p = 0.862$	$\gamma_w = 0.221$	$\zeta_p = 0.887$	$\zeta_w = 0.620$		$\sigma_z = 0.0064$	$\sigma_x = 0.557$
	$\epsilon_w = 1.2$	$\lambda_r = 0.779$	$\bar{\pi} = 1.016$	$\lambda_\pi = 1.454$		$\sigma_r = 0.0026$	
	$\lambda_y = 0.234$						

Figure 9, which plots the shape of distance function when monetary and technology shocks are jointly considered, shows that the problems previously noted are present to a much larger degree here. For example, the objective function is very flat in many dimensions (the scale of the graphs is $10e-7$), somewhat asymmetric and this is true for a larger range of parameters' values. Moreover, there is a multidimensional ridge in the price stickiness (ζ_p), price indexation (γ_p), wage stickiness (ζ_w) and wage indexation (γ_w) parameters (see figure 10) - several combinations of these parameters which produce a value for the objective function which is close to the minimum. For these dimensions, the use of responses to technology shocks does not help: identification of these parameters is as problematic considering or disregarding TFP or investment specific disturbances.

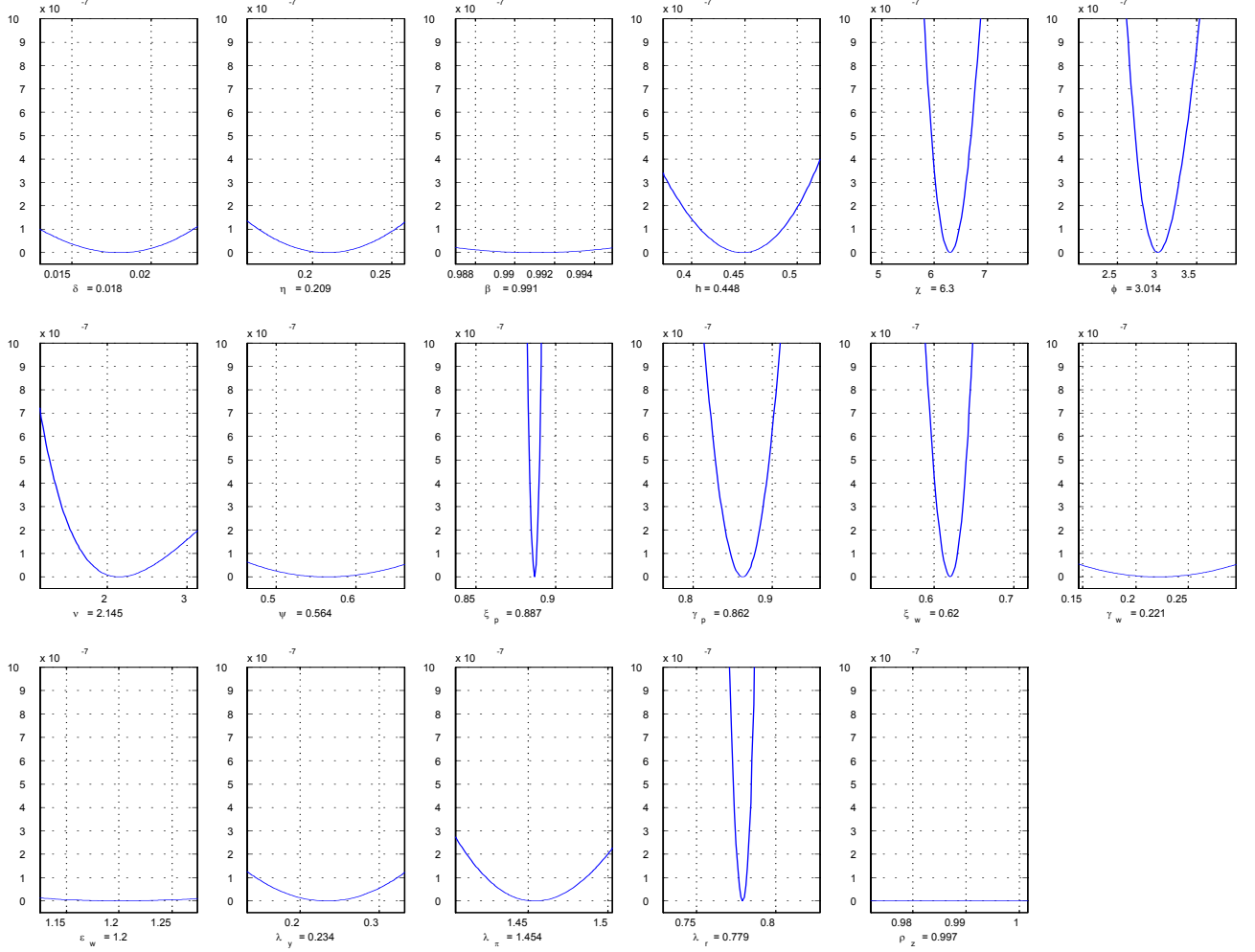


Figure 9. Objective function: monetary and technology shocks

Armed with this preliminary evidence, we consider a few alternative models where either stickiness or indexation in wages or prices is eliminated from the true DGP and estimate the parameters of the fully fledged model. Table 4 reports our estimation results when population responses are used. For each specification considered there are four rows: each report estimates obtained starting the minimization routine at different points⁷. In cases 1 to 5 and 7 only responses to monetary shocks are used; in case 6 responses to monetary and technology shocks are employed.

⁷For each parameter θ_i , we select an economically reasonable interval $[a \ b]$ and assume a uniform distribution on it. The starting values are selected as: $a + j * stderr(\theta_j)$ or $b - j * stderr(\theta_j)$, where $j = 1, 2$.

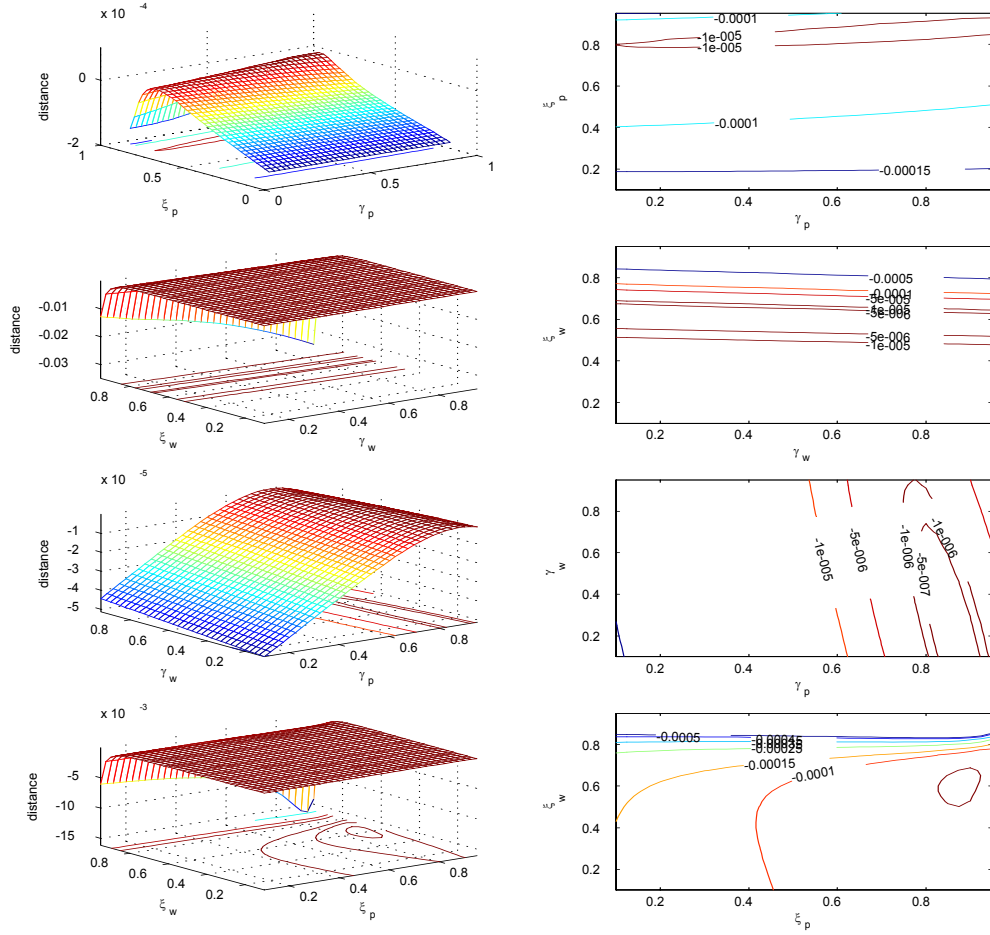


Figure 10: Distance surfaces and Contour Plots

Several interesting features are present in Table 4. First, in the baseline case, when all the features are present, price and wage indexation are estimated to be smaller than the true ones. Second, responses to monetary shocks can not distinguish models featuring price indexation from models missing this feature (compare cases 1 and 2); it is possible to confuse a model with no price stickiness and no wage indexation with a model with these two features but with no price indexation (see case 3); models with no price indexation and high wage indexation are observationally equivalent to models where both features are present and roughly of the same size (see case 4).

Table 4. Estimation results

	ζ_p	γ_p	ζ_w	γ_w	Obj.Fun.
Baseline	0.887	0.862	0.62	0.221	
x0 = lb + 1std	0.8944	0.8251	0.615	0	1.8235E-07
x0 = lb + 2std	0.8924	0.7768	0.6095	0.1005	3.75E-07
x0 = ub - 1std	0.882	0.7957	0.6062	0.1316	2.43E-07
x0 = ub - 2std	0.9044	0.7701	0.6301	0	8.72E-07
Case 1	0	0.862	0.62	0.221	
x0 = lb + 1std	0.1304	0.0038	0.6401	0.245	2.7278E-08
x0 = lb + 2std	0.1015	0.0853	0.6065	0.1791	4.84E-08
x0 = ub - 1std	0.0701	0.1304	0.6128	0.1979	4.72E-08
x0 = ub - 2std	0.0922	0.0749	0.618	0.215	3.05E-08
Case 2	0	0	0.62	0.221	
x0 = lb + 1std	0.1396	0.0072	0.6392	0.2436	3.1902E-08
x0 = lb + 2std	0.0838	0.1193	0.6044	0.1683	4.38E-08
x0 = ub - 1std	0.0539	0.1773	0.6006	0.1575	5.51E-08
x0 = ub - 2std	0.0789	0.0971	0.6114	0.1835	2.61E-08
Case 3	0	0.862	0.62	0	
x0 = lb + 1std	0.0248	0	0.6273	0.029	7.437E-09
x0 = lb + 2std	0.4649	0	0.7443	0.4668	2.10E-06
x0 = ub - 1std	0.0652	0.0004	0.6147	0.0447	7.13E-08
x0 = ub - 2std	0.6463	0.2673	0.8222	0.3811	5.56E-06
Case 4	0.887	0	0.62	0.8	
x0 = lb + 1std	0.9264	0.3701	0.637	0.4919	3.5156E-07
x0 = lb + 2std	0.9076	0.2268	0.6415	0.154	3.51E-07
x0 = ub - 1std	0.9014	0.3945	0.6477	0	6.12E-07
x0 = ub - 2std	0.9263	0.3133	0.6294	0.4252	4.13E-07
Case 5	0.887	0	0	0.221	
x0 = lb + 1std	0.9186	0.3536	0.0023	0	4.7877E-07
x0 = lb + 2std	0.8994	0.234	0	0	3.06E-07
x0 = ub - 1std	0.905	0.3494	0.0021	0	4.14E-07
x0 = ub - 2std	0.9343	0.5409	0.0042	0	9.64E-07
Case 6	0.887	0	0	0.221	
x0 = lb + 1std	0.877	0.0123	0.0229	0	2.4547E-06
x0 = lb + 2std	0.8919	0.0411	0.0003	0	4.26E-07
x0 = ub - 1std	0.907	0.2056	0.001	0.0001	6.58E-07
x0 = ub - 2std	0.8839	0.0499	0.0189	0	2.46E-06
Case 7	0.887	0	0	0.25	
x0 = lb + 1std	0.9056	0.2747	0.0154	0.25	1.60E-06
x0 = lb + 2std	0.9052	0.2805	0	0.25	2.41E-07
x0 = ub - 1std	0.9061	0.3669	0.0003	0.25	4.26E-07
x0 = ub - 2std	0.8985	0.194	0.001	0.25	2.07E-07

Finally, a model where prices are sticky and wages are partially indexed can not be distinguished from a model which features substantial price indexation but no wage stickiness or wage indexation (case 5). Third, in all the cases, the minimized objective function is within the tolerance limit. Fourth, taking the estimates producing the infimum of the objective function across minimizations fails to solve the problem since the ridge in (γ_p, γ_w) is extremely flat. This fact can be clearly appreciated in figure 11, where we report responses to monetary shocks obtained in case 5 with true and estimated parameters: any investigator looking at this graph would have no doubt that she has nailed down the correct model! Can these problems be reduced if responses to a larger number of shocks are considered? Case 6 reports estimates of the parameters obtained jointly using responses to monetary and technology shocks, and little improvements obtain. It is only when responses to all shocks are considered that the range of values consistent with the true DGP shrinks.

Does it matter which parameters values one uses, say, for policy purposes? It clearly does. For example, using the true parameters of case 5 of and those estimated in the first row of the block of case 5 in table 3, we find that the implied variability of output and inflation in the two cases is different. Equally weighting the variability of the two series and computing the resulting loss, welfare turns out to be about 4 times worse with estimated parameters than with the true ones (-0.0022 vs. -0.0005).

Could we reduce the observational equivalence problem using external information to fix some of the parameters? Such a strategy is unlikely to work here, since the ridge in the objective function is multidimensional. Hence, we need to fix three of the four troublesome parameters and at the right value. The last row of Table 3 (case 7) reports estimates obtained for the model of case 5 when γ_w is fixed to 0.25. Clearly, estimates of γ_p are still off the mark.

It is important to stress that the results we present are obtained in the ideal conditions in which the population responses are available. Clearly, observationally equivalent problems could be made considerably worse if the weighting matrix is altered, the number of responses for each variables or the number of variables consider reduced, and only sample responses are available. In models like this where partial, weak and observational equivalence problems are present, one needs to bring a lot of information external to the dynamics, as for example it is done in Christiano et. al. (2005), to be able to interpret estimates. It then becomes crucial where this external information comes from and whether it is credible or not.

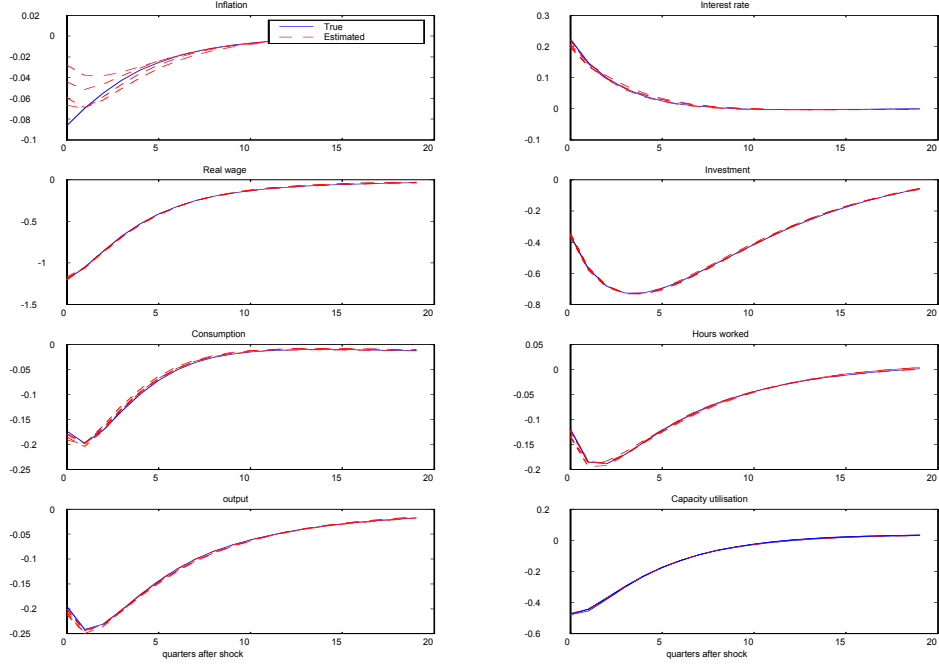


Figure 11: Impulse responses, Case 5

6 Detecting identification problems

Are there ways to detect potential problems and to understand what are the features of the model economy that could lead to them? The graphical analysis we have used could be routinely and costlessly implemented and lots of information gathered this way. However, such an analysis can also be strengthened using formal methods. When the objective function has derivatives up to second - a standard assumption made in the literature - under, partial and weak identification all induce Hessians at the optimum which are rank deficient or fail to have sufficient curvature.

How do one check for the rank of the Hessian? Cragg and Donald (1997) have provided a procedure to do this. Let $h = \text{vec}(H)$ be the vectorized version of H and let $d(L) = (h - p)'(h - p)$, where $p = \text{vec}(P)$ and P is a matrix of rank L . Under regularity conditions, when an estimate \hat{h} is available, $Td(L) \rightarrow \chi^2((K - L)(K - L - 1)/2 - K)$, where $K(K + 1)/2$ is the number of free elements of H and for $L < L_0$, the true rank, $Td(L)$ is divergent, while for $L \geq L_0$, $Td(L) \leq Td(L_0)$.

Alternatively, Anderson (1984, p.475) has shown that estimates of the eigenvalues

of a matrix when properly scaled have an asymptotic standard normal distribution. Therefore, the null hypothesis of full rank can be tested against the alternative of rank deficiency examining whether the smallest of the eigenvalues of the Hessian is statistically not different from zero. Since the magnitude of the eigenvalues may depend on the unit of measurements, Anderson also suggests to test the null that the sum of the smallest k' eigenvalues to the average of all k eigenvalues is large. This ratio is also asymptotically normally distributed with zero mean and unit variance when properly scaled, and it is useful since the alternative accounts for the possibility that none of the first k' eigenvalues is zero but that all of them are small (generating weak identification problems). Since this test requires that the Hessian is consistently estimated under the null and the alternative - which is impossible to do in our case - one could use the insights of the test to diagnose anomalies in the size of the eigenvalues.

Finally, one could use the concentration statistics $\mathcal{C}_{\theta_0}(i) = \int_{j \neq i} \frac{g(\theta) - g(\theta_0)d\theta}{\int (\theta - \theta_0)d\theta}, i = 1, 2, \dots$, to detect identification problems. Stock, Wright and Yogo (2002) showed that this statistics synthetically measures the curvature of the objective function around θ_0 and it is related to the non-centrality parameter of the χ^2 used in testing the hypothesis that the objective function at the optimum is zero. Large values of $\mathcal{C}_{\theta_0}(i)$ imply that it is easy to reject the null if the objective function is not zero; small values imply that the displacement of the χ^2 from its null value are difficult to detect. While there are no critical values for this statistics, one could use the values produced by Stock, Wright and Yogo for linear models to get an idea of potential identification problems.

We apply the last two diagnostics to the Hessian of the objective function of the model of section 5 at the values estimated in case 5. Both confirm the presence of significant rank deficiencies. In fact, the maximal concentration statistic (over i) is 0.25 and thirteen of the eighteen roots of the Hessian are small: the sum of the first 12 roots is only 1.0 percent of the average root, the sum of the first 13 roots is 1.8 percent of the average root and the first root is calculated to be smaller than $1.0e^{-10}$. Therefore, at least 12 of the parameters can not be identified from the responses to monetary shocks. The situation slightly improves when we use both monetary and technology shocks (case 6), but not by much: the sum of the first 12 roots is 2.1 percent of the average root. It is easy to verify that the parameters associated with the 12 small eigenvalues are $(\rho_z, \beta, \phi, \nu, h, \delta, \eta, \gamma_p, \gamma_w, \epsilon_w, \lambda_\pi, \lambda_y)$. Interestingly, many of these parameters were also those creating identification problems in the smaller version of the model considered in section 4. Therefore, adding variables (and responses) does

not necessarily improves the identifiability of e.g., $\beta, \lambda_y, \lambda_\pi$; it is difficult to distinguish backward from forward looking dynamics both in prices and wages; and there is very little information to select production, capacity and depreciation parameters. As in section 3.3, the fact that the low of motion of the states is roughly insensitive to variations of these structural parameters in a neighborhood of the estimated values is responsible for the lack of curvature in the objective function.

7 Conclusions and suggestions for empirical practice

Liu (1960) and Sims (1980) have argued that traditional models of simultaneous equations were hopelessly under-identified and that identification of an economic structure was often achieved not because there was sufficient information in the data but because researchers wanted it to be so - limiting the number of variables in an equation or eschewing a numbers of equations from the model.

Since then models have dramatically evolved, precise microfoundations added, general equilibrium features taken into account, and economic measures of fit designed. Still, it appears that a large class of popular DSGE structures is close to being under-identified; observational equivalence is widespread; and reasonable estimates are obtained not because the data is informative but because of a priori or auxiliary restrictions, which make the likelihood of the data (or a portion of it) informative. In these situations, structural parameter estimation amounts to sophisticated calibration and this makes model evaluation and economic inference hard.

A study of identification issues like ours, besides ringing a warning bell about the potential problems existing in tracing a formal link between DSGE models and the data, is useful in practice only to the extent it gives applied researchers a strategy to detect problems and means to either avoid them in estimation and inference or to develop theoretical specifications which overcome the lack of identifiability of the structural parameters. Providing such a set of tools is complicated since the relationship between parameters and impulse responses is highly non-linear; the mapping is unknown and only an approximation is available; problems are multidimensional and standard diagnostics are unsuitable to understand the sources of identification failure.

This paper provides some hints on how to approach such an issue. We summarize our suggestions as a list of non-exhaustive steps which we recommend applied investigators to check before attempting structural estimation. First, plotting the objective

function, a few dimensions at the time, may provide useful indications for the presence of potential identification problems and point out parameters responsible for them. Second, examining the rank of the Hessian (or the magnitude of its smaller eigenvalues) provides formal evidence for the visual tendencies that plots may deliver. Since such tests are unlikely to be able to distinguish which particular problem is present, they should be used as general specification diagnostic for the presence of information deficiencies. These tests are simple to compute and, in principle, applicable to any point in the parameter space. Hence, exploration of the properties of the Hessian at or around e.g., standard calibrated parameters, should logically precede model estimation. Third, simplified versions of the model may give some economic intuition for why identification problems emerge as could the use of several limited information objective functions. Working with small versions of large models or with portions of their dynamic implications will also help with model respecification. Fourth, mixing calibration and estimation may lead optimization routines to search for the minimum of the function in the wrong portion of the parameter space and researchers to draw inappropriate conclusions about how the economy works. Fifth, the smaller is the number of cross variables, cross equation and cross horizon restrictions used in estimation, the larger is the chance that identification problems will be present. This suggests to use as many implications of the model as possible de facto eliminating the hedge that limited information approaches have over likelihood methods, both of classical or Bayesian flavours. Sixth, while for identification likelihood methods are generally preferable, one should be aware that even the likelihood function is not the cure for all identification problems and that Bayesian methods, if improperly used, may cause researchers to oversee them. Seventh, if identification problems persist even when the full information provided by the model and whatever additional external information is used, one could attempt to obtain estimates via S-sets, as suggested by Stock and Wright (2000), rather than minimize the distance between impulse responses. Alternatively, one should go back to the drawing board. Often identification problems occur because models are not explicitly constructed with an eye to estimation. Finally, scientific honesty demands that the specification of the model is based on prior knowledge of the phenomenon, not on the desire to identify the characteristics a researcher happens to be interested in. Nevertheless, resisting the temptation to arbitrarily induce identifiability is the only way to make DSGE models verifiable and knowledge about them accumulate on solid ground.

References

- [1] An, S. and Schorfheide, F. (2005) Bayesian analysis of DSGE models, forthcoming, *Econometric Reviews*.
- [2] Anderson, T. (1984) *An Introduction to Multivariate Statistical Analysis*, Wiley & Sons, New York, NY.
- [3] Altig, D., Christiano, L., Eichenbaum, M. and Linde, J. (2005) Firm specific capital, nominal rigidities and the business cycle, Fed of Chicago, working paper, 2005-01.
- [4] Beyer, A. and Farmer, R. (2004) On the indeterminacy of new Keynesian Economics, ECB working paper 323.
- [5] Canova, F. (2002) Validating two monetary models via VARs, CEPR working paper 3442
- [6] Canova, F. (2004) Structural changes in the US economy, available at www.econ.upf.edu/crei/canova.
- [7] Canova, F. and Sala, L. (2005) Back to square one: identification issues in DSGE models, ECB working paper.
- [8] Chari, V.V., Kehoe, P. and McGrattan, E. (2005) A critique of SVAR using Real Business Cycle theory, Fed of Minneapolis, working paper 631.
- [9] Choi, I and P.C. Phillips (1992) Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations, *Journal of Econometrics* 51, 113-150.
- [10] Christiano, L., Eichenbaum, M. and Evans, C. (2005) Nominal Rigidities and the Dynamic Effects of a shock to Monetary Policy, *Journal of Political Economy*, 113, 1-45.
- [11] Christiano, L., Eichenbaum, M. and Vigfusson, R. (2005) Assessing structural VARs, NBER working paper 12353.
- [12] Cogley, T., Colacito, R., Sargent, T. (2005) Benefits from US Monetary Policy Experimentation in the days of Samuelson and Solow and Lucas, NYU manuscript.

- [13] Cragg, J. and Donald, S. (1997) Inferring the rank of a matrix, *Journal of Econometrics* 76, 223-250.
- [14] Dedola, L. and Neri, S. (2004) What does a technology shock do? A VAR analysis with model-based sign restrictions, CEPR working paper 4537, forthcoming *Journal of Monetary Economics*.
- [15] Del Negro, M, Schorfheide, F., Smets, F. And R. Wouters (2005) On the fit of New Keynesian models, CEPR working paper 4848.
- [16] Ellison, M. (2005) Discussion to Cogley, Colacito and Sargent's Benefits from US Monetary Policy Experimentation in the days of Samuelson and Solow and Lucas, manuscript.
- [17] Fernandez-Villaverde J. and Rubio-Ramirez J. F. (2005), Estimating dynamic equilibrium economies: linear versus nonlinear likelihood, *Journal of Applied Econometrics*, 20, 891-910
- [18] Gali, J. and Rabanal, P. (2005) Technology shocks and aggregate fluctuations: How well does the RBC model fit Postwar US data?, *NBER Macroeconomic Annual*, 20.
- [19] Fernandez-Villaverde, J., Rubio-Ramirez, J. and Sargent, T. (2005) The ABC (and D) of structural VARs, NBER working paper.
- [20] Hansen, L., Heaton, J. and Yaron, A (1996) Finite Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics*, 14, 262-280.
- [21] Hsiao, C. (1983) Identification in Griliches, Z. and Intrilligator, M. (eds.), *Handbook of Econometrics*, I, 224-283.
- [22] Kennan, J. (1998) An econometric analysis of the fluctuations in aggregate labor supply and demand, *Econometrica*, 45, 969-990.
- [23] Kim, J. (2003) Functional equivalence between intertemporal and multisectoral investment adjustment costs, *Journal of Economic Dynamics and Control*, 27, 533-549.

- [24] Koop G., Pesaran H. and Potter S. (1996), Impulse response analysis in non-linear multivariate models, *Journal of Econometrics*, 74, 119-148.
- [25] Ireland, P. (2004) Technology Shocks in the New Keynesian model, *Review of Economics and Statistics*, 86, 923-936.
- [26] Lindè, J. (2005), Estimating New Keynesian Phillips' curve: A full information maximum likelihood approach, *Journal of Monetary Economics*, 52, 1135-1149.
- [27] Liu, T. (1960) Underidentification, Structural estimation, and Forecasting, *Econometrica*, 28, 855-865.
- [28] Lubik, T. and Schorfheide, F. (2004) Testing for indeterminacy: An application to US Monetary Policy, *American Economic Review*, 94, 190-217.
- [29] Nason J. and Smith G: W. (2005), Identifying the new keynesian Phillips curve, Federal Reserve Bank of Atlanta, wp 2005-07.
- [30] Neely, C., Roy, A., Whiteman, C. (2001) Risk Aversion versus Intertemporal Substitution: A case study of identification failures in the Intertemporal Capital Asset Pricing Model, *Journal of Business and Economic Statistics*, 19, 295-403.
- [31] Ma, A. (2002) GMM estimation of the new Phillips curve, *Economic Letters*, 76, 411-417.
- [32] Nason, J. and Smith, G. (2005) Identifying the New Keynesian Phillips Curve, Federal Reserve of Atlanta, working paper 2005-1.
- [33] Pesaran, H. (1981) identification of Rational Expectations Models, *Journal of Econometrics*, 16, 375-398.
- [34] Poirier, D. (1998), Revising Beliefs in Nonidentified Models, *Econometric Theory*, 14, 183-209
- [35] Rabanal, P. and Rubio-Ramirez, J (2005) Comparing New-Keynesian Models of the Business Cycle: A Bayesian Approach, *Journal of Monetary Economics*, 52, 1150-1162.
- [36] Rothenberg, T. (1971) Identification in parametric models, *Econometrica*, 39, 577-591.

- [37] Rotemberg, J. and Woodford, M. (1997) , An Optimization based Econometric Framework for the evaluation of Monetary Policy, *NBER Macroeconomic Annual*, 12, 297-346.
- [38] Sargent, T. (1978) Estimation of Dynamic Labor Demand Schedules under Rational Expectations, *Journal of Political Economy*, 86, 1009-1044.
- [39] Sims, C. (1980) Macroeconomic and Reality, *Econometrica*, 48, 1-48.
- [40] Smets, F. and Wouters, R. (2003) An estimated Dynamic Stochastic General Equilibrium models of the Euro area, *Journal of the European Economic Association*, I, 1123-1175.
- [41] Smith, A. (1993) Estimating Nonlinear Time series models using simulated VARs, *Journal of Applied Econometrics*, 8, 63-84.
- [42] Stock, J. and Wright, J. (2000) GMM with weak identification, *Econometrica*, 68, 1055-1096.
- [43] Stock, J., Wright, J. and Yogo, M. (2002) A survey of weak instruments and weak identification in Generalized methods of moments, *Journal of Business and Economics Statistics*, 20, 518-529.
- [44] Wright, J. (2003) Detecting lack of identification in GMM, *Econometric Theory*, 19, 322-330.