Back to square one: identification issues in DSGE models*

Fabio Canova
ICREA, UPF, CREI and CEPR

Luca Sala
Istituto di Economia Politica and IGIER - Universitá Bocconi

First Draft: May 2005
This draft: December 2005

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. We show that observational equivalence, partial and weak identification problems are widespread, that 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.

*We would like to thank M. Del Negro, M. Ellison, D. Giannone, J. Linde', A. Levin, S. Johansen, F. Schorfheide, J. Stock and the participants of seminars at Bocconi, Warwick, the ECB and at conferences (ESSIM, Cyprus; New methods for Empirical Macroeconomics, Oslo) for comments and suggestions.

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. In particular, following Rotemberg and Woodford (1997) and others, it has become 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 zero and should display "enough" curvature in all relevant dimensions. Since impulse 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. Two reasons make the problem hard. First, since stationary solutions are typically found with numerical methods, the mapping from structural parameters to impulse responses is not analytically available. Second, since the objective function can be evaluated only at a finite number of points, it is difficult to infer its properties in high dimensional parameter spaces.

This paper investigates identifiability issues in DSGE models and their consequences for parameter estimation and model evaluation. Furthermore, it provides diagnostics to detect identification problems when the objective function measures the distance between (structural) sample and model impulse responses. Since the field is vast and largely unexplored, our analysis focuses on a selected number of important issues.

Section 2 discusses the generics of identification, highlights the problems existing in models which are nonlinear in the parameters and gives definitions for several practically relevant situations. Section 3 provides a few examples of simple structures generating three commonly encountered problems: observational equivalence; under-identification; weak and partial identification of the parameters. We examine three general issues. First, we study what features of the economic environment are responsible for the problems. Second, we examine the consequences of altering the weights responses receive in the objective function and the number of variables considered in the analysis. Third, we evaluate whether and in what way changing the objective function affects identification. We show that observational equivalence, weak and partial identification all lead to objective functions with large flat surfaces in the economically reasonable portion of the parameter space; that identification problems depend on the objective function used - full information methods have at times an hedge over partial information ones - and that Bayesian methods, if properly used, can help to detect identification problems but, if improperly used, may cover them up. We demonstrate that flat objective functions lead to serious biases and that fixing some of the troublesome parameters at arbitrary values may create distortions in the distribution of parameter estimates, unless the chosen value happens to be the correct one.

Section 4 investigates the interaction between parameters' identifiability, shock identification and small samples. We argue in the context of a commonly used three equation New-Keynesian model that many of the structural parameters are only weakly or partially identifiable when impulse responses are used. We show that small samples and incorrect shock identification pile up to induce major distortions in parameter estimates when coupled with identification problems and conclude that parameter identification, in practice, has to do with the structure of the model, the objective function, the sample size and several other auxiliary model specification assumptions.

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 particular, in examining cases 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 additional features generating endogenous persistence are only very weakly identified. We show that the objective function is flat in the parameters characterizing nominal price and wage rigidities and that investigators using responses to monetary and/or technology shocks could be mistakenly induced to 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. Finally, section 7 summarizes the results and provides suggestions for empirical practice.

2 A few definitions

Identification problems has 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, if the population objective function does not have a unique maximum, the mapping between structural parameters and reduced form statistics is not unique. Hence, different structural models having potentially different economic interpretations may be indistinguishable from the point of view of the chosen objective function. Such a statement does not imply that the two models are indistinguishable under *all* objective functions nor that it is impossible to find implications which are different. 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 the log-linearized solution of the model. In this case the objective function will be constant for all values of that parameter in a selected range. 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 is globally concave, its curvature may be "insufficient". This problem could be specific to a neighbor of the zero or concern the entire parameter space. We call this phenomenon *weak identification* problem. One interesting special case arises when the objective function is asymmetric in the neighborhood of the zero and its curvature deficient only in a portion of the parameter space. Fourth, and this applies to objective functions which consider only a subset of the implications (e.g. a limited number of shocks or a subset of the responses to all shocks), different responses (shocks) may carry different information about the parameters. Hence, a parameter could be identifiable if all information is employed, but remains under-identifiable if one shock or one particular set of responses is used. We call this *limited information identification* problem. Finally, partial, weak and limited information identification problems can be exacerbated if only a sample version of the population objective function is available.

We formalize the above concepts as follows. Suppose we want to minimize $g(y, T, m, \theta)$, with respect to a $k \times 1$ vector of structural parameters $\theta \in \Theta$, where y is a vector of data, T the sample size, m a DSGE model and $g(y, T, m, \theta) = (ir^d(y, T) - ir^m(m, \theta))W(T)(ir^d(y, T) - ir^m(m, \theta))'$, where $ir^d(y, T)$ is a vector of data-based structural responses, $ir^d(m, \theta)$ is a vector model-based responses, and $W(T)$ is a weighting matrix, function of the sample size T. Identification has to do with the shape and the curvature of $g(y, T, m, \theta)$.

Two models m_1 and m_2, with parameter vectors θ and ξ, are *observationally equivalent* given y, if $g(y, T, m_1, \theta^*) = g(y, T, m_2, \xi^*) = 0$, some θ^*, ξ^*. In other words, $\frac{\partial g(y,T,m_1,\theta)}{\partial \theta}|_{\theta^*} = \frac{\partial g(y,T,m_2,\xi)}{\partial \xi}|_{\xi^*} = 0$ and both $\frac{\partial^2 g(y,T,m_1,\theta)}{\partial \theta \partial \theta'}|_{\theta^*}$ and $\frac{\partial^2 g(y,T,m_2,\xi)}{\partial \xi \partial \xi'}|_{\xi^*}$ are positive definite.

Let $m_1 = m_2$, $\theta = [\theta_1, \theta_2]$ and partition $\Theta = [\Theta^1, \Theta^2]$. θ_1 is *locally under-identified* if $g(y, T, m, \theta_1, \theta_2) = g(y, T, m, \theta_2)$ $\forall \theta_1 \in \Theta_1 \subset \Theta^1$. On the other hand $g(y, T, m, \theta) = g(y, T, m, \theta_1 f(\theta_2)), \forall \theta_1 \in \Theta_1 \subset \Theta^1, \theta_2 \in \Theta_2 \subset \Theta^2$, for some continuous and differentiable f, then θ is locally *partially identified.* Under-identification and partial identification imply that $\frac{\partial g(y,T,m,\theta)}{\partial \theta}|_{\theta} = 0$ and that $\frac{\partial^2 g(y,T,m,\theta)}{\partial \theta \partial \theta'}|_{\theta}$ is rank deficient. Global under and partial identification occur when $\Theta_1 = \Theta^1$.

A parameter vector θ is *locally weakly identifiable* if there exist a unique θ^* such that $g(y, T, m, \theta^*) = 0$ but $g(y, T, m, \theta) < \epsilon, \forall \theta \in \Theta^{\dagger} \subset \Theta$ and *globally weakly identified* if this occur for all $\theta \in \Theta$. Weak identification implies that $\frac{\partial g(y,T,m,\theta)}{\partial \theta}|_{\theta^*} = 0$, that $\frac{\partial^2 g(y,T,m,\theta)}{\partial \theta \partial \theta'}|_{\theta^*}$ is positive definite but that $\mu_1, \ldots, \mu_{k-n}$ are small, where μ_i, $i = 1, \ldots, k$ are the increasingly ordered eigenvalues of $\frac{\partial^2 g(y,T,m,\theta)}{\partial \theta \partial \theta'}|_{\theta^*}$.

Asymmetric weak identification results if $H_r = \frac{\partial_r^2 g(y,T,m\theta)}{\partial_r \theta \theta'}|_{\theta^*}$, the Hessian computed using the right derivatives, differs from $H_l = \frac{\partial_l^2 g(y,T,m,\theta)}{\partial_l \theta \theta'}|_{\theta^*}$, the Hessian computed using the left derivatives of the function g, and one of the two has some μ_i which is small.

Finally, *limited information* identification definitions are obtained when the weighting matrix $W(T)$ can be factored as $W(T) = \mathcal{S}\mathcal{W}(T)$, where $\mathcal{S}$ is a selection matrix with ones and zeros.

If the objective function to be minimized is (minus) the likelihood function of the data, $L(y, \theta)$, identification is related to the shape and the rank of the information matrix: $\Im(\theta)$, whose (i,j)-th element is $\Im(\theta)_{ij} = E(\frac{\partial^2 log(L(y,\theta))}{\partial\theta_i \partial\theta_j})$. As it is well known, θ is locally identifiable if and only if the rank of $\Im(\theta)$ is constant and equal to k in a neighborhood of θ^* (Rothenberg, 1971). Since the information matrix measures the "curvature" of the likelihood function, curvature deficiencies around θ^*, make some rows or columns of $\Im(\theta)$ (close to) zero or proportional to each other.

Under-identification and weak identification have been recognized to be serious problems. Choi and Phillips (1992), Stock and Wright (2000) have shown the consequences these two phenomena have on the asymptotic properties of parameter estimates in IV and GMM setups. Stock and Wright (2000) also develop an asymptotic theory which is robust to identification problems. Since our function g resembles the objective function used in this literature, one may wonder whether identification problems can be sidestepped using their approach. Unfortunately their theory is inapplicable in our case because $W(T)$ is never chosen to be the continuously updating weighting matrix of Hansen et al. (1996). Nevertheless, the intuition obtained in IV and GMM frameworks carries over, to a large extent, to our case.

Before discussing the practical consequences of identification problems for estimation and inference, we provide a few examples intended to show (a) the pervasiveness of identification problems in DSGE models, (b) the consequences of using a limited information approaches to conduct inference, (c) the advantages/disadvantages of employing different objective functions.

3 Identification problems in practice

3.1 Observational equivalence: two structural models have the same impulse responses.

The example we consider illustrates one of problems often encountered in practice: the inability of impulse responses to distinguish two different economic structures. Suppose a time series x_t is generated from $x_t = \frac{1}{\lambda_2+\lambda_1} E_t x_{t+1} + \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2} x_{t-1} + v_t$, where $\lambda_2 \geq 1 \geq \lambda_1 \geq 0$. It is well known that the unique stable rational expectations solution is $x_t = \lambda_1 x_{t-1} + \frac{\lambda_2+\lambda_1}{\lambda_2} v_t$. Therefore, given $v_t = 1$, the responses of $x_{t+j}, j = 0, 1, \ldots$ are $[\frac{\lambda_2+\lambda_1}{\lambda_2}, \lambda_1 \frac{\lambda_2+\lambda_1}{\lambda_2}, \lambda_1^2 \frac{\lambda_2+\lambda_1}{\lambda_2}, \ldots]$, and using at least two horizons, one can estimate λ_1 and λ_2. It is easy to construct a different process whose stable rational expectation solution has the same impulse response. Consider, for example, $y_t = \lambda_1 y_{t-1} + w_t$, $0 \leq \lambda_1 < 1$. Clearly the process is stable and, as long as $\sigma_w = \frac{\lambda_2+\lambda_1}{\lambda_2} \sigma_v$, the responses of x_t and y_t to shocks will be indistinguishable.

What makes the two processes equivalent in terms of impulse responses? Clearly, the unstable root λ_2 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 it in the second case so as to capture the effects of the unstable root. Since an investigator has one degree of freedom, she can make two processes share both contemporaneous and lagged dynamics. In general, since many refinements of currently used DSGE models add some backward looking component to a standard forward looking one, the range of applicability of this result is quite large (see also Lubik and Schorfheide (2004) and An and Schorfheide (2005) for similar examples).

This example can be extended a larger class of processes, driven by pure expectational errors. Suppose $y_t = \frac{1}{\lambda_1} E_t y_{t+1}$, where $y_{t+1} = E_t y_{t+1} + w_t$ and w_t is an iid shock with zero mean and variance σ_w^2. The stable solution is again $y_t = \lambda_1 y_{t-1} + w_t$. Hence, if $\sigma_w = \frac{\lambda_2+\lambda_1}{\lambda_2} \sigma_v$, a process with (deterministic) forward looking dynamics and expectational errors is observationally equivalent to a process with forward and backward looking dynamics driven by an iid fundamental error. Such an equivalence is 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 it is the cornerstone of Pesaran's (1981) critique of tests of rational vs. adaptive expectations models.

Several other examples of observationally equivalent structures have appeared in the literature. For example, Kim (2003) shows that models with adjustment costs to capital are observationally equivalent to models which assume a nonlinear transformation curve between consumption and investment, at least as far as Euler equations are concerned; Ma (2002) shows that a standard forward looking Phillips curve is consistent with two structural models having different firms' pricing behavior; Altig et al. (2004) construct a model with firm specific capital which produces the same inflation dynamics of a model without firm specific capital; while Ellison (2005) shows that two Phillips curve relationships based on different microeconomic and behavioral assumptions can be made observationally equivalent from the point of view of responses to policy shocks if the policy function is appropriately chosen. In general, observational equivalence is problematic when the zeros occur in correspondence of different vectors of economically reasonable parameters. In this case, information external to the models needs to be brought in to disentangle various structural representations.

3.2 Under-identification: some structural parameters disappear from impulse responses.

Cases of models where structural parameters fail to appear in the impulse response functions are also numerous. Consider the following three equations model:

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

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

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

where y_t is the output gap, π_t the inflation rate, i_t the nominal interest rate and the first equation is the log-linearized Euler condition, the second a forward looking Phillips curve and the third characterizes monetary policy. Since this model features no states, the solution for the three variables is a linear in the three shocks v_{jt} and given by:

$$\begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_2 \\ a_4 & 1 & a_2 a_4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix}$$

Three useful points can be made. First, the parameters a_1, a_3, a_5 disappear from the solution. Interestingly, they are those characterizing the forward looking dynamics of the model. Second, different shocks carry different information for the parameters: responses to v_{1t} allow us to recover only a_4; responses to v_{3t} may be used to back out both a_4 and a_2 while responses to v_{2t} have no information for the two parameters. Similarly, responses of different variables carry different information about the structural parameters. Third, different objective functions may have different information about the parameters. In this simple example, a_1, a_3, a_5 remain underidentified even when the likelihood of the model is used; however, the latter has information about the variances of the shocks, information that normalized responses do not have.

From this discussion it is clear that even when the number of nonzero dynamic coefficients exceeds the number of structural parameters, a condition which is necessary for identification in models which are linear in the parameters, problems may remain. This is reminiscent of the irrelevant instruments problem present in IV setups.

3.3 Weak and partial identification

There is a sense in which the situations considered in the two previous examples are pathological. The objective function is, in fact, ill-behaved in both cases: it either displays multiple zeros or it is constant in some dimension. In practice, there are less extreme but equally interesting situations where the population objective function (locally) has a unique zero, its Hessian is (locally) positive definite but parameters are only weakly or partially identified. To show that both features are relatively common 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 better 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 $c_t + k_{t+1} = k_{t-1}^{\eta} z_t + (1-\delta) k_t$, 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 is the current capital stock, η is the share of capital in production and δ the depreciation rate of capital. The parameters of the model are $\theta = [\beta, \phi, \delta, \eta, \rho, z^{ss}]$. Using the method of undetermined coefficients and letting output be $y_t = k_{t-1}^{\eta} z_t$, 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 $A w_t = B w_{t-1} + C e_t$ 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 standard values for the parameters ($\beta = 0.985, \phi = 2.0, \rho = 0.95, \eta = 0.36, \delta = 0.025, z^{ss} = 1$) and use the model to generate data.

To show the features of the objective function obtained matching impulse responses, we compute the distance between the true impulse responses and the impulse responses obtained varying one or two parameters at a time in an economically reasonable neighborhood of the selected values.

The first row of Figure 1 presents (the negative of) two of these three-dimensional surfaces and the corresponding contour plots. While there is a unique minimum in correspondence of the true parameter vector, the objective function is quite flat either locally around the minimum or globally over the entire parameter range. For example, the persistence parameter ρ is very weakly identified in the interval [0.8,1.0]. Interestingly, the distance function displays a ridge of approximately similar height in the depreciation rate δ and the discount factor β, running from $(\delta = 0.005, \beta = 0.975)$ up to $(\delta = 0.03, \beta = 0.99)$, indicating that the two parameters are only partially identifiable. In the latter case, the one percent contour includes the whole range of economically interesting values of δ and β. Although to save space, we do not show this, the share of capital in the production function η is also only very weakly identifiable in the range [0.3,0.6] and another ridge appears when we plot the objective function against the steady state value of the technology shock z^{ss} and the depreciation rate δ.

Given our solution, we can check which of the coefficients in the matrices A and B is responsible for this state of affairs. It turns out that v_{kk} and v_{kz} are the coefficients responsible for the problems. In fact, the (numerical) local derivative of v_{kz} with respect to ρ equals 0.08 and those of v_{kk} and v_{kz} with respect to η are, respectively, -0.10 and 0.09. In other words, the objective function is flat in ρ and η because the dynamics of the capital stocks are only weakly influenced by these two parameters. Since the law of motion of the capital stock determines the dynamics of the other variables, the responses of other variables carry little information about the structural parameters. The local derivatives of v_{kk} and v_{kz} with respect to β and δ are also small, have similar magnitude but opposite sign. Hence, the dynamics of the capital stock are roughly insensitive to proportional changes in these two parameters.

The distance surface plotted in the first row of figure 1 uses the responses of the vector w_t. Since it is unusual to consider the entire vector of responses produced by the model, we have recalculated the surfaces when only responses to consumption and output are used to construct the distance function. Clearly, one expects some loss of information relative to the baseline case; the question is how large the loss is. The second row of Figure 1, which reports these surfaces, shows that the curvature of the objective function is smaller at any point in the range but that the shape hardly changes. Therefore, there are no obvious distortions, only a uniform loss of curvature. Since there is only one shock and since output and consumption inherit the dynamics of the capital stock and of the technology shocks, excluding these two variables does not distort the information. It should be obvious that when there is more than one shock or more than one state variable, results could dramatically change.

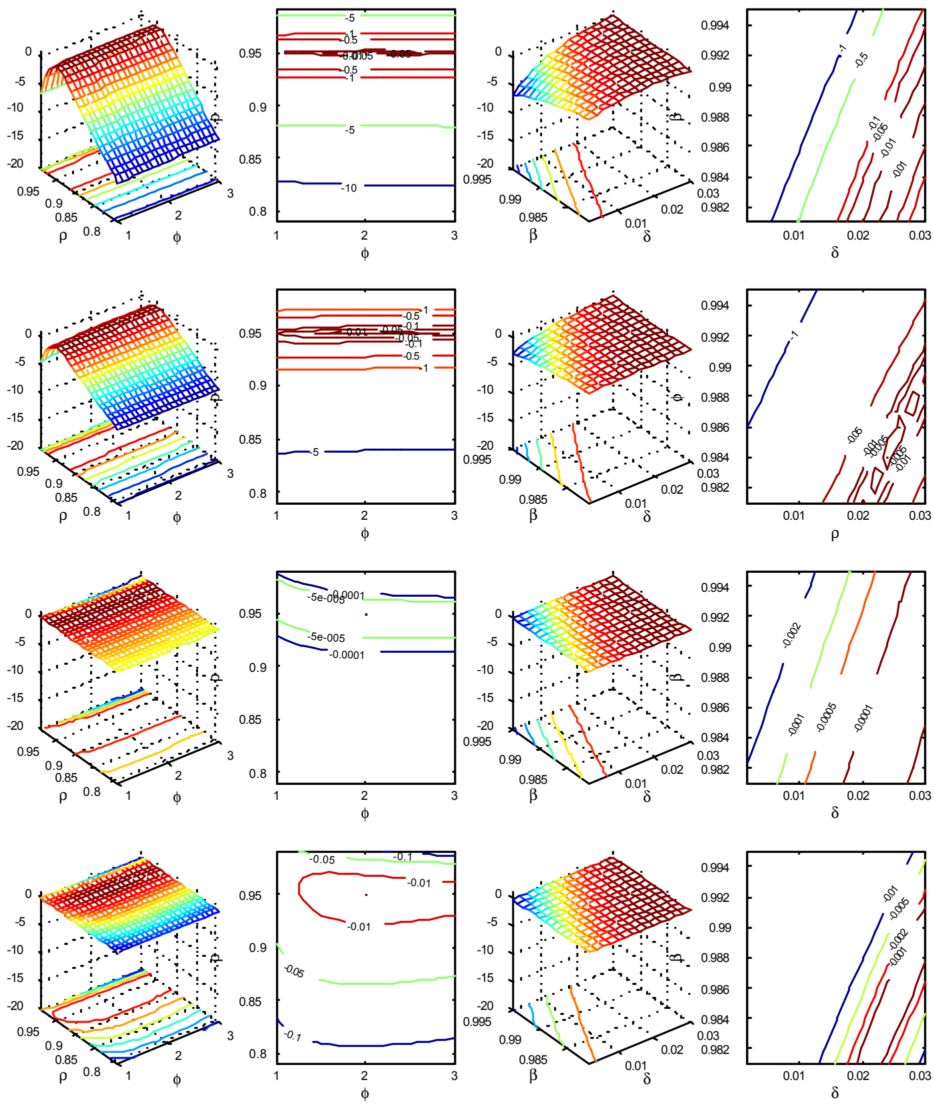

Figure 1: Distance surfaces and contour plots; basic, subset, weighted and matching VAR

When responses are estimated, weak identification problems may arise because responses at long horizons are noisy and may carry little information about the structural parameters. Such a phenomenon is analogous to the weak instrument problem in GMM frameworks where instruments "too lagged" in the past are more likely to satisfy the exogeneity assumption but may also be weakly correlated with the objects of interest (see e.g. Stock, Wright and Yogo (2002)). So far we used 20 horizons of each of the four variables and since we were working with population responses, we set $W(T) = I$. To see how identifiability depends on the choice of horizons and, at the same time, mimic typical situations encountered in applied work where long horizon responses have large standard errors, we computed the surfaces using a weighting matrix with $\frac{1}{h^2}$ on the diagonal, $h = 1, 2, \ldots, 20$ for each variable and zero everywhere else. As it is clear from the third row of figure 1, this choice considerably worsens identification problems: plateaus exist in all dimensions and the objective function is now flatter for a much larger range of values of the parameters. Hence, the larger the number of cross equations restrictions used is, the smaller identification problems are likely to be.

One may wonder if matching the coefficients of the D matrix in the VAR(1) representation of the model: $w_t = Dw_{t-1} + v_t$, where $D = A^{-1}B$ and $v_t = A^{-1}Ce_t$, as suggested by Smith (1993), would make any difference for identification purposes. Intuitively, concentrating on VAR coefficients could be beneficial because shocks' identification is entirely sidestepped. On the other hand, choosing parameters to match only the coefficients of the D matrix could worsen the outcome since information present in v_t is neglected. While it is *a priori* difficult to determine which effect dominates, the fourth row of Figure 1 indicates that the loss of information due to the use of a smaller number of restrictions dominates.

For empirical purposes, it is important to know whether identification problems are specific to one particular objective function or intrinsic to the model. If the former is true, carefully choosing the objective function may avoid headaches. In the latter case, some reparametrization of the original model is needed. To distinguish between these two alternative we have examined the shape of the likelihood of the model, assuming a normally distributed technology shock, in the same bivariate dimensions previously considered. Under correct model specification, the "degree of identification" delivered by the likelihood function is a natural upper bound. 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 estimation. Weak identification problems seem to be less acute when the likelihood function is used. For example, ρ is pinned down with much higher precision (see top panel in figure 2). Also the likelihood shows some flat area but contour plots are much better behaved.

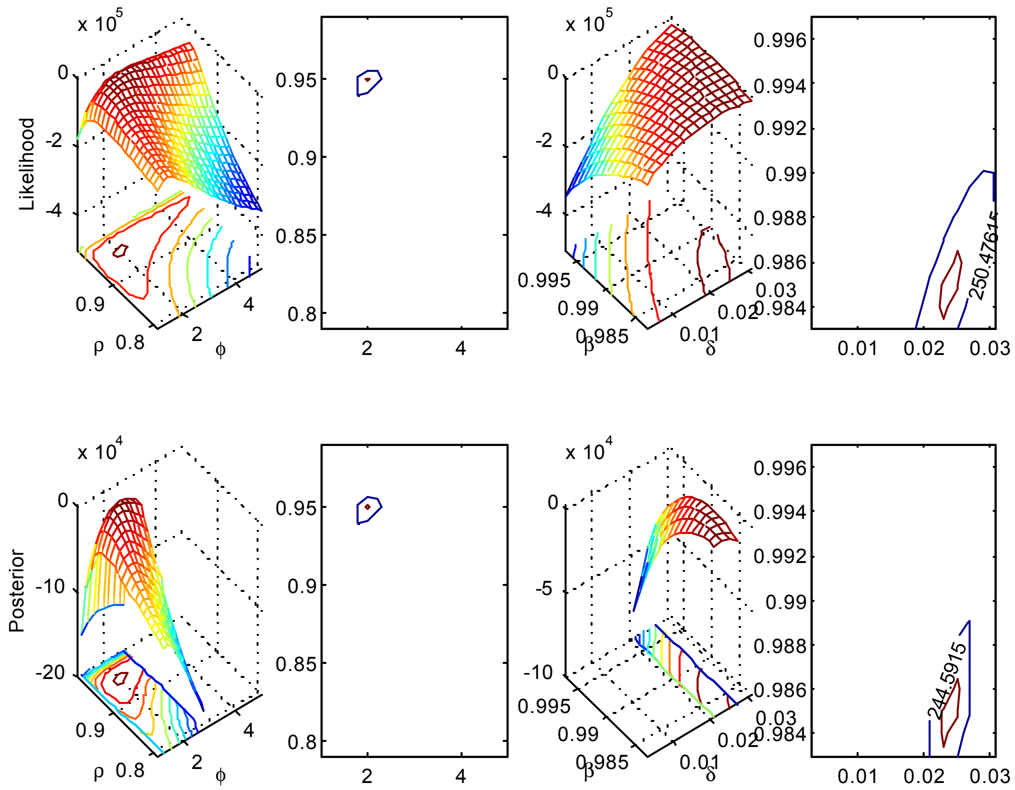

Figure 2: Likelihood and Posterior Surfaces and Contours

Since it is now common to estimate DSGE models with Bayesian methods, few words contrasting identification problems in classical and Bayesian frameworks are in order. Posterior distributions are proportional to the likelihood times the prior. If the space of parameters is variation free, that is, there is no implicit constraint on combinations of parameters, the data carries important information if the prior and posterior have different features. When this is not the case, there is a simple diagnostic for detecting lack of identification, a diagnostic unavailable in the (classical) setup we consider in this paper. In fact, 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.

When the parameter space is not variation free, because stability or non-explosiveness conditions or economically motivated non-negativity constraints are imposed, the prior of non-identified parameters may be marginally updated even if the likelihood has no information (see Poirier (1998)). In this case, finding that prior and posterior differ does not guarantee that the data is informative. Only by using a sequence of prior distributions with increasing spreads, one can detect potential identification problems.

Unfortunately, this simple diagnostic is hardly ever used and often prior distributions are not even reported. This is dangerous. A tightly specified prior can in fact produce a well behaved posterior distribution, even if the likelihood function has little information, giving the illusion of having collected useful evidence, e.g., about important policy parameters. We show this fact in the second row of figure 2: here a tight prior on δ eliminates the partial identification problem previously encountered. Hence, uncritical use of Bayesian methods, including employing prior distributions which do not truly reflect the existing location uncertainty, may hide identification problems instead of highlighting them.

In conclusion, one could probably be better endowed to answer interesting economic questions if she carefully selects the objective function used. However, even in the most favorable conditions, identification problems are likely to remain if the model is not specifically parametrized with an eye to estimation.

What are the practical consequences of having objective functions with large flat areas and ridges? First, the choice of minimization algorithm matters: with a poor one, it is unlikely that final estimates of weakly identified parameters will move much from initial conditions [1]. Second, different values of β and δ could be selected, depending on the initial conditions. Third, as we will see in the next section, since estimates of weakly identified parameters are likely to be inconsistent and their asymptotic distribution non-normal, the practice of reporting point estimates of the model's parameters with standard errors computed under the usual asymptotic assumptions, may be uninformative about the goodness of parameter estimates or the properties of the model.

[1] For example, the Matlab routine FMINUNC is totally unable to explore surfaces with these featuers while the Matlab routine LSQNONLIN seems to be doing a much better job.

How do one solve partial identification problems? The standard practice of fixing β works here since for any value of β, the distance function has reasonable curvature in the δ dimension (and viceversa). However, 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 conditional on correctly assuming $\beta = 0.985$ and on incorrectly assuming $\beta = 0.995$.

It is clear that when β is incorrectly chosen, the location of the objective function shifts away from the maximum so that the estimated distribution of the parameters may fail to include the true value. Hence, even minor errors in setting one of the partially identified parameters may lead routines to search for optimal values in wrong portions of the parameters space and give the wrong impression that estimation is successful, as standard errors and the optimized objective function may be small.

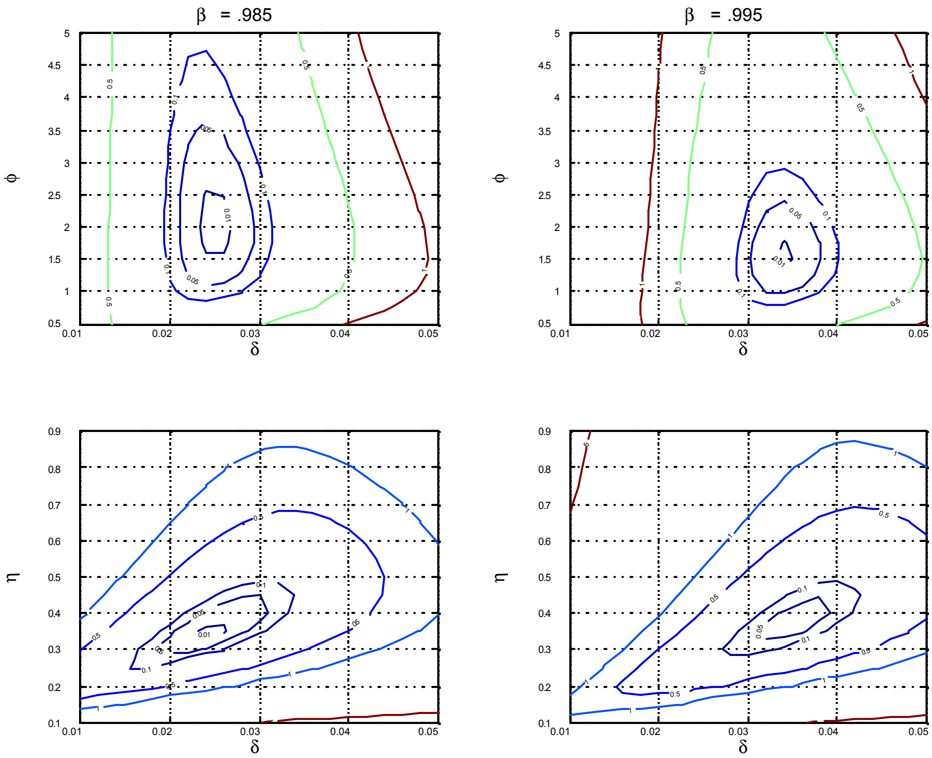

Figure 3: Contour Plots

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 impulse responses. In the first part we endow the researcher with the population responses; in the second we measure in what way small samples complicate the inferential task.

To make our points transparent, we employ a well known New-Keynesian model. We choose such a 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:

$$y_t = \frac{h}{1+h} y_{t-1} + \frac{1}{1+h} E_t y_{t+1} + \frac{1}{\phi}(i_t - E_t \pi_{t+1}) + v_{1t} \tag{4}$$

$$\pi_t = \frac{\omega}{1+\omega\beta}\pi_{t-1} + \frac{\beta}{1+\omega\beta}\pi_{t+1} + \frac{(\phi+\nu)(1-\zeta\beta)(1-\zeta)}{(1+\omega\beta)\zeta} y_t + v_{2t} \tag{5}$$

$$i_t = \lambda_r i_{t-1} + (1-\lambda_r)(\lambda_\pi \pi_{t-1} + \lambda_y y_{t-1}) + v_{3t} \tag{6}$$

where h is the degree of habit persistence, ν is the inverse elasticity of labor supply, 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. As it is standard, the first two shocks follow autoregressive processes of order one with AR parameters ρ_1, ρ_2, while v_{3t} is an iid shock. The variances of the three shocks are denoted by $\sigma_i^2, i = 1, 2, 3$. While other equations can be added to alleviate potential problems, this structure is sufficient to highlight the distortions one is likely to face in the presence of identification problems.

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

The framework we use allows us to construct a number of objective functions - several limited information ones, obtained considering the responses to only one type of shock and a full information one - and therefore assess the importance of limited information identification problems in the context of a concrete example. We take the true parameters to be $\beta = 0.985, \phi = 2.0, \nu = 3.0, \zeta = 0.68, \lambda_r = 0.2, \lambda_\pi = 1.55, \lambda_y = 1.1, \rho_1 = 0.65, \rho_2 = 0.65, \omega = 0.25, h = 0.85$, which are standard in calibration exercises and quite close to Rubio and Rabanal's (2005) estimates.

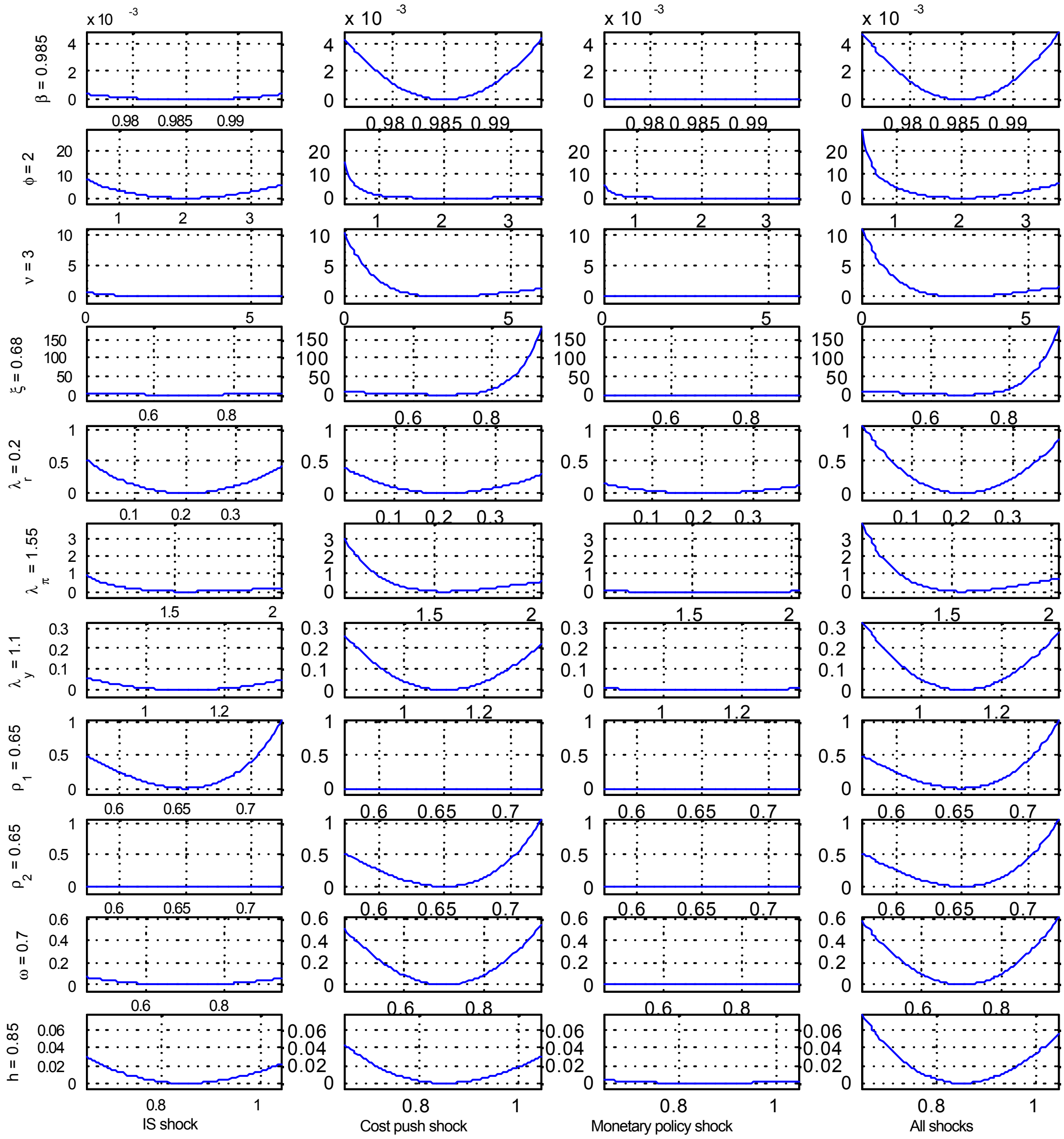

Figure 4: Shape of different objective functions

To start with we plot in figure 4 the shape of the objective function in each of the elements of θ_2. We compute the distance function using responses to v_{1t} (column 1), to v_{2t} (column 2), to v_{3t} (column 3) and to all the shocks (column 4), varying one parameter at the time within an economically reasonable range around the selected values. To be clear: figure 4 shows the curvature of the objective function one dimension at the time, conditional on the other $n-1$ values being fixed at their "true" values.

Many interesting features are present in the figure. First, the first three objective functions are flat in several dimensions (see e.g. $\lambda_\pi, \lambda_y, \omega, h$). Second, different shocks have different information about certain parameters (see e.g. ζ, λ_r). Interestingly, the function measuring the distance in response to monetary shocks is very flat in all the dimensions except ϕ, ζ. Therefore, responses to monetary shocks are unlikely to be informative about many of the structural parameters. Third, the objective functions are asymmetric in certain dimensions. For example, when cost push shocks are considered, the distance function is asymmetric in the risk aversion parameter ϕ, the inverse elasticity of labor supply ν and the price stickiness ζ. 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. Finally, even when responses to all shocks are used, the objective function is still flat and asymmetric in several dimensions. Hence, weak and partial identification problems remain, even when all the available information is used.

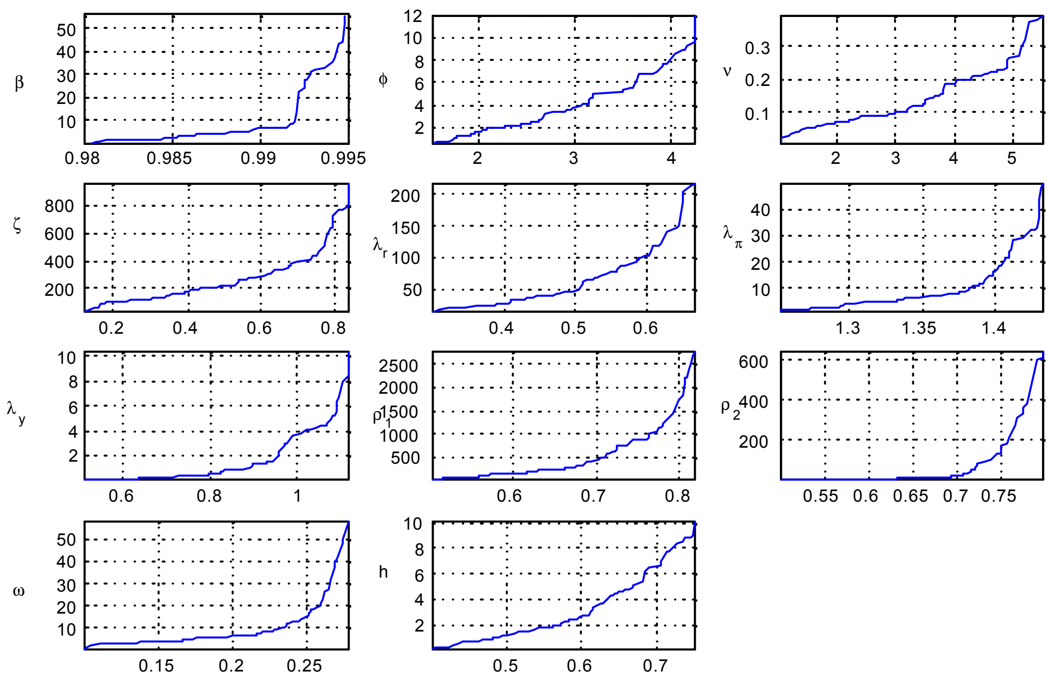

Figure 5: Concentration statistics

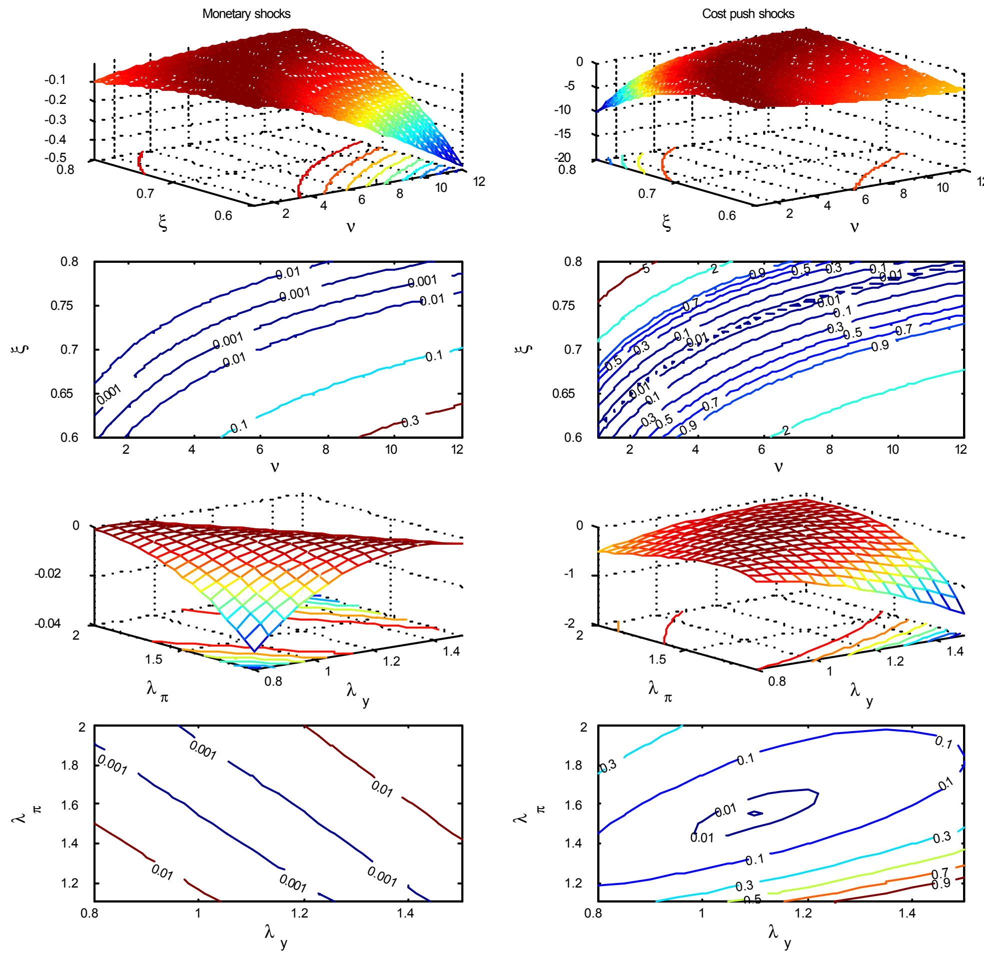

Figure 6: Distance function and contour plots

Since the shapes presented in figure 4 may depend on the choice of the true parameter vector, figure 5 plots the concentration statistics $\mathcal{C}_{\theta_0}(i) = \int_{j\neq i} \frac{g(\theta)-g(\theta_0)d\theta}{(\theta-\theta_0)d\theta}, i = 1, \ldots, 11$, when we vary θ_0 over a reasonable range. Such a statistics synthetically measures the global curvature of the objective function over a selected range of values for the parameters (see Stock, Wight and Yogo (2002)). We present results obtained when we match all impulse responses since, as figure 4 shows, identification can not be improved upon by matching responses to single shocks. For each θ_0 reported on the horizontal axis, we construct the statistics varying $\theta \in [0.5\theta_0, 1.5\theta_0]$ using a grid with 100 values in each of the 11 dimensions. To interpret the figure, note if the objective function had a slope of 1, changing the value of θ_0 would not change $\mathcal{C}_{\theta_0}$. Furthermore, an objective function with a slope of 1 would produce a concentration statistics equal to 1 throughout the range for θ_0. Hence, values for $\mathcal{C}_{\theta_0}$ exceeding 100 indicate "good" curvature in the objective function. Figure 5 confirms that ν, λ_y, ϕ, h are weakly identified no matter what the true value of the parameter is, while for ζ, ρ_2 identification appears to depend on the true parameter value.

Since figures 4 and 5 consider one dimension at the time, they may miss ridges in the objective function: that is, they may miss the presence of observationally equivalent structures, indexed by the size of two parameters. Figure 6 shows that ridges are present: both responses to cost push and to monetary shocks carry little information about the correct combination of λ_y and λ_π or ν and ζ.

In sum, this prototype model displays an array of potential identification problems. In the next subsections, 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. Since the results obtained matching other shocks or all shocks are similar, our focus does not reduce the generality of the conclusions we draw. Figure 7 reports the density 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 in each box is the true parameter value. Histograms are obtained eliminating all cases where (i) convergence failed; (ii) the estimated parameters produce imaginary or (iii) indeterminate solutions. It is worth mentioning that the histograms in Figure 7 do not capture sampling uncertainty associated with the estimation of structural parameters, as the econometrician is here endowed with the population responses. Instead, with this figure we intend to display the multivariate mapping from impulse responses to structural parameters. If the distance function had no ridges, flat regions or local minima, this mapping would be univocal: from any starting point one would reach the true value and the histograms would be degenerate.

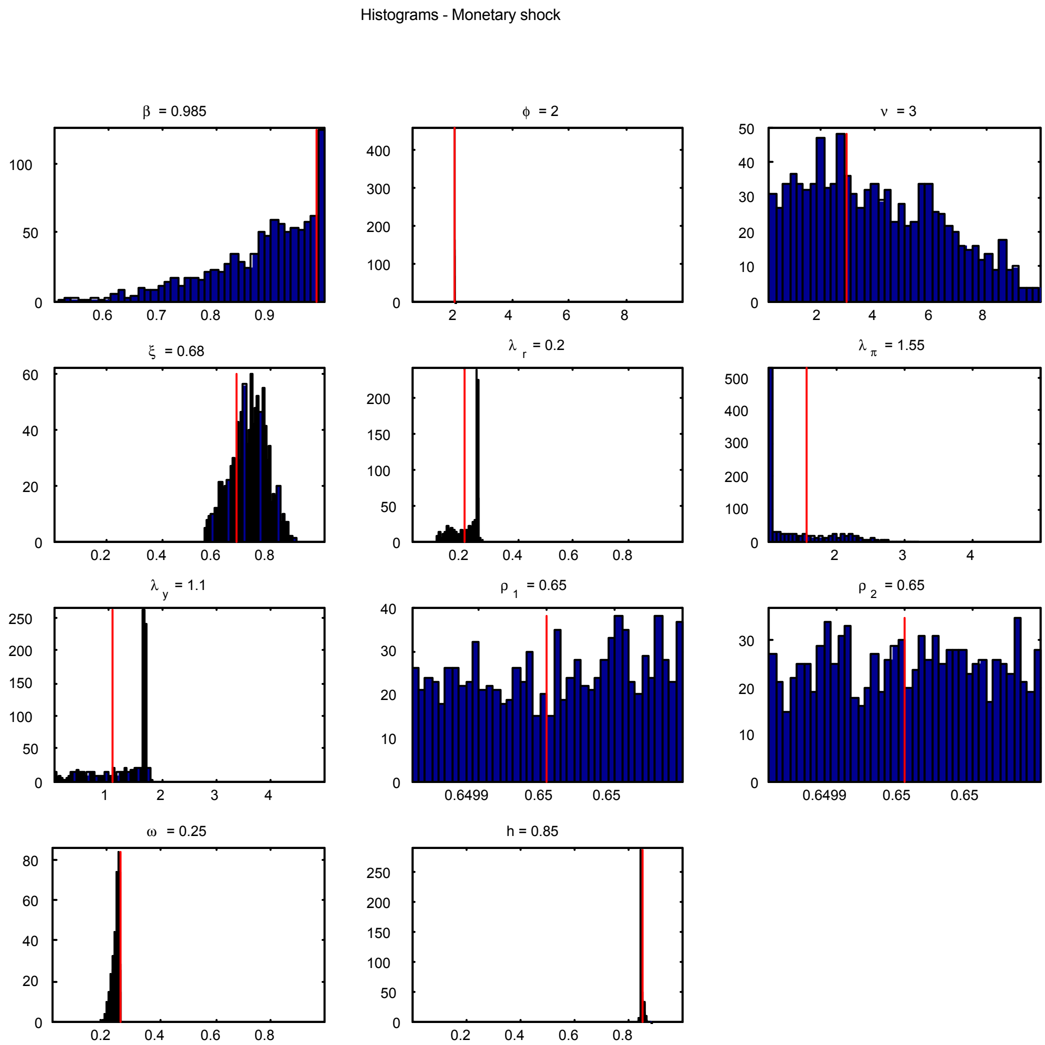

Figure 7: Distribution of estimates

There are few interesting features we would like to comment upon. First, a number of biases are evident. For example, there is a tendency to overestimate β; the mode of the distribution of estimates of λ_π is located at 1.06, well below the true value of 1.55, and the one of λ_y is located at 1.8, well above the true value 1.1. Interestingly, for some parameters (notably ξ and the three λs) it is possible to rule out portions of the parameter space, but it is not possible to pin down precisely the true parameter value. For other, e.g. ν, no parts of the parameter space can be completely excluded. Hence, even disregarding sampling uncertainty, major estimation biases may be induced in parameters with problematic identification features.

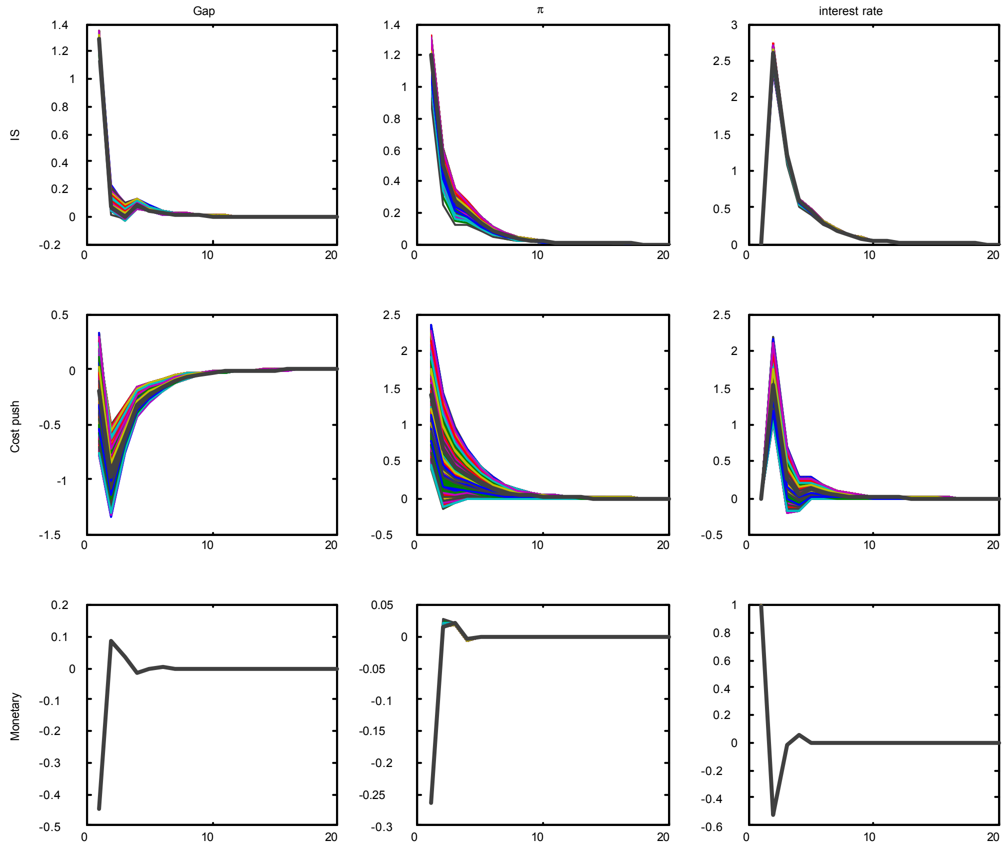

Figure 8: Impulse responses

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 small for all the parameter combinations generating figure 7, and as shown in figure 8, population and implied responses to monetary shocks are indistinguishable. Interestingly, responses to IS and cost push shocks are also very similar to the true ones. Hence, parameter vectors with potentially different economic interpretations are indistinguishable when normalized responses are used to construct objective functions [2].

[2]Linde (2005) has shown that maximum likelihood estimation of the parameters of a model like ours is feasible and succesful. However, his parameters are not truely structural and identification problems are absent. In general, there is no reason to expect maximum likelihood estiamtors to be better endowed to deal with ridges in the parameter space than minimum distance estimators.

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, it is unwise to attach any economic interpretation to the estimates, draw conclusions about how the economy works, conduct policy analyses using the estimated vector and, more importantly, this is true even in the ideal situation considered in this subsection.

4.2 Weak identification and small samples

The distortions we have noted could 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 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) further highlight some of the properties of estimates of parameters with problematic identification features. We focus again attention on responses to policy shocks, since the particular structure we have imposed 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 knows the model and correctly identifies the monetary shock. 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 $T = 120, 200, 1000$, run an unrestricted VAR(2)[3] on the simulated data, compute impulse responses and bootstrapped confidence bands. We use the confidence bands to build the weighting matrix: weights are inversely proportional to the uncertainty in the estimates. Consistent with the theoretical model, we identify the monetary policy shock as the first element of a Choleski decomposition of the covariance matrix where the interest rate is ordered first in the VAR.

Stock and Wright (2000) have shown that identification problems in GMM frameworks produce inconsistent estimate of weakly or under-identified parameters, that the joint distribution of weakly (or under-identified) and properly identified parameters is non-standard; and that standard t-statistics are, in general, invalid. While their conclusions do not necessarily apply to our framework, one should intuitively expect

[3] We checked that the VAR(2) is able to correctly estimate the true impulse responses with the correct identification when $T = 5000$.

similar patterns to emerge when the objective function measures the distance of impulse responses. In particular, one should expect (i) erratic properties of estimates of weakly (or under-identified) parameters as T increases; (ii) standard errors which are large and do not change with the sample size; and (iii) t-tests which are uninformative about the properties of estimates.

Table 1: NK model. Matching monetary policy shocks

	True values	Population	T = 120	T = 200	T=1000	T=1000 wrong
β	.985	.987 (.003)	.98 (.007)	.98 (.006)	.98 (.007)	.999 (.008)
ϕ	2	2 (.003)	1.49 (2.878)	1.504 (1.906)	1.757 (.823)	10 (.420)
ν	3	4.082 (1.653)	4.184 (1.963)	4.269 (1.763)	4.517 (1.634)	1.421 (2.33)
ζ	.68	.702 (.038)	.644 (.156)	.641 (.112)	.621 (.071)	.998(.072)
λ_r	.2	.247 (.026)	.552 (.272)	.481 (.266)	.352 (.253)	.417 (.099)
λ_π	1.55	1.013 (.337)	1.058 (1.527)	1.107 (1.309)	1.345 (1.186)	3.607 (1.281)
λ_y	1.1	1.683 (.333)	4.304 (2.111)	2.924 (2.126)	1.498 (2.088)	2.59 (1.442)
ρ_1	.65	.5 (.212)	.5 (.209)	.5 (.212)	.5 (.167)	.5 (.188)
ρ_2	.65	.5 (.207)	.5 (.208)	.5 (.213)	.5 (.188)	.5 (.193)
ω	.25	.246 (.006)	1 (.360)	1 (.35)	1 (.306)	0 (.384)
h	.85	.844 (.006)	1 (.379)	1 (.321)	1 (.233)	0 (.166)

Table 1 presents a summary of our estimation results. We report true parameters, median estimates and standard errors obtained using population responses and responses estimated with different T. Standard errors are computed across replications.

Few features are worth commenting upon. First, large biases are evident in the estimates of the weakly identified parameters (ν, λ_π, λ_y), the under-identified parameters (ρ_1, ρ_2) and their standard errors are large. Second, parameter estimates of the identified parameters do not necessarily converge to the population ones as T increases (see, for example, ϕ). This is consistent with the idea that the bias present in weakly and under-identified parameters spills to the other parameters and remain significant even in large samples. Third, parameter estimates and standard errors of weakly identified and under-identified parameters are independent of the sample size. Fourth, and even with 250 years of quarterly data major biases in, e.g., the two policy parameters, remain. Finally, and concentrating on $T = 200$, estimates suggest an economic behavior which is substantially different from the one characterizing the DGP. For example, it appears that agents have preferences where the stock of habit plays an extreme role; price indexation is complete and the Central Bank reaction to the output gap is much stronger than the one to inflation. Once again, armed just with impulse responses, an investigation has little possibility to detect such interpretation problems.

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, the identification of monetary shocks is subject to errors. To give a glimpse of how shock and parameter identification interact, we report in the last column of table 2 estimates obtained when $T = 1000$ and monetary shocks are identified as the third element of a Choleski decomposition; that is, 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 tend to be 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 presented in the last column of table 1 to the "insignificant" estimates obtained in the case monetary shocks are correctly chosen.

In conclusion, small samples exacerbate the consequences of identification problems for estimation and inference. Weak identification combined with small samples typically lead to very 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. Furthermore, 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.

5 Misspecification and observational equivalence.

When the investigator knows the model, ridges in the objective function may appear so that combinations of parameters with different economic interpretation are almost equally likely. When the true model is unknown, one can not a-priori exclude that different structures with alternative economic interpretations are almost 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 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:

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

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 ez_t is a technological disturbance; the fourth equation represents the monetary policy rule and er_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 the current inflation rate; the eight 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 ex_t an investment shock; the ninth equation describes the dynamics of the 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, $ep_t(ew_t)$ are shocks to the pricing relationships, and $T_p \equiv \frac{(1-\beta\zeta_p)(1-\zeta_p)}{(1+\beta\gamma_p)\zeta_p}$ 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 $\sigma_i, i = 1, \ldots 5$ 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.

To first show the identification problems a researcher faces in matching the responses of such a model we construct population 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.

Table 2. Parameter values

θ_1 :

$\beta = 0.991$	$\phi = 3.014$	$\nu = 2.145$	$h = 0.448$
$\delta = 0.0182$	$\eta = 0.209$	$\chi = 6.300$	$\psi = 0.564$
$\gamma_p = 0.862$	$\gamma_w = 0.221$	$\zeta_p = 0.887$	$\zeta_w = 0.620$
$\epsilon_w = 1.2$	$\lambda_r = 0.779$	$\bar{\pi} = 1.016$	$\lambda_\pi = 1.454$
$\lambda_y = 0.234$			

, θ_2 :

$\rho_z = 0.997$	$\sigma_p = 0.221$
$\rho_x = 0.522$	$\sigma_w = 0.253$
$\sigma_z = 0.0064$	$\sigma_x = 0.557$
$\sigma_r = 0.0026$	

Figure 9, which plots the distance function when we consider monetary and technology shocks jointly, shows that the problems previously noted are present to a much larger degree here. For example, the local derivative of the objective function with respect to many of the parameters is very flat (the scale of the graphs is 10e-7), somewhat asymmetric and this is true for a even larger range of values for the parameters. 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). That is, there are several combinations of these parameters which produce an objective function which is close to zero. Note that, at least in 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.

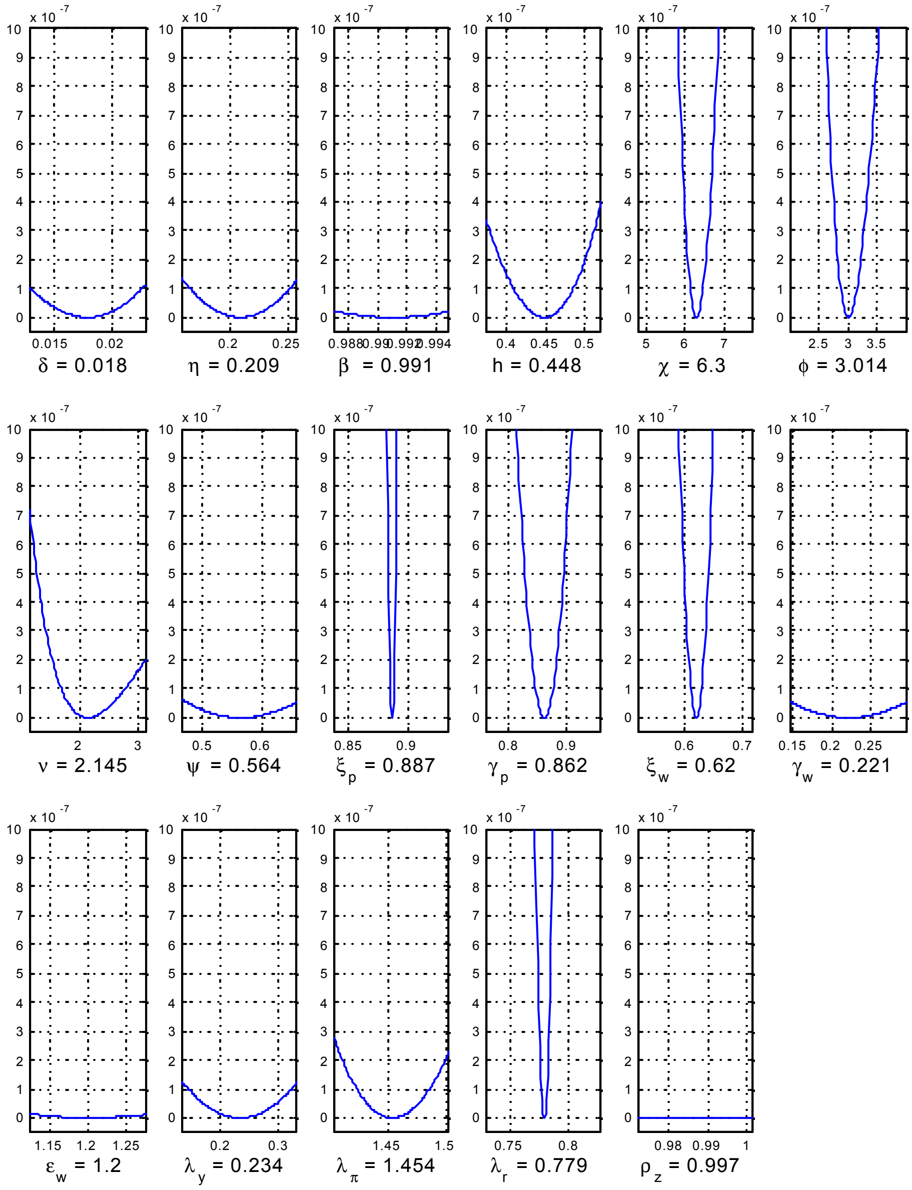

Figure 9. Objective function: monetary and technology shocks

Armed with this evidence, we consider a few alternative models where either stickiness or indexation in wage or prices is eliminated from the true DGP and estimate the parameters of the fully fledged model. Table 3 reports our estimation results when population responses are used and different specification considered. For each specification there are four rows: they report results obtained starting the minimization routine at different points[4]. 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. Estimates appear without standard errors since, as in Neely, Roy and Whiteman (2001), it is impossible to invert the Hessian at the selected estimates as its determinant is zero at machine precision, indicating not only weak but also under-identification of two or more parameters.

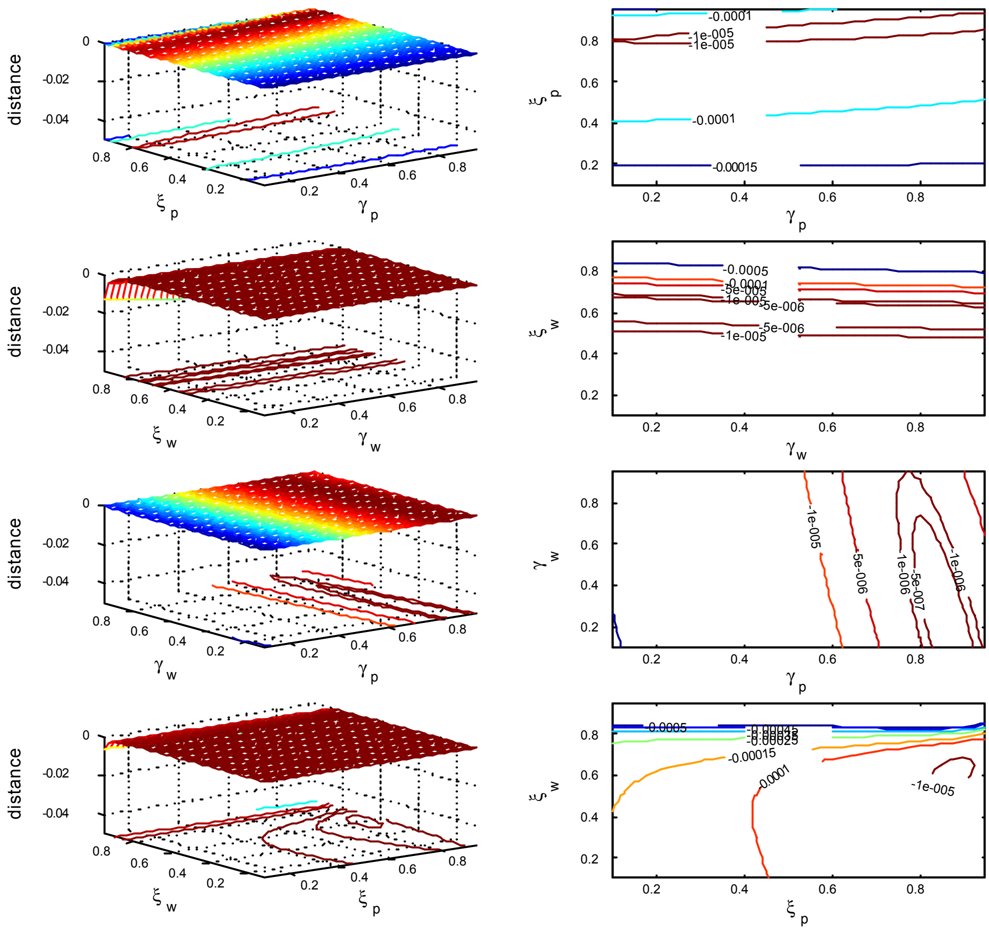

Figure 10: Distance surfaces and Contour Plots

[4] For 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 $j = 1, 2$.

Table 3. 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**	**0.221**	
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

lb is the economic lower bound and ub is the economic upper bound.

Several interesting features are present in Table 3. First, in the baseline case, price and wage indexations are estimated to be smaller than the true ones. Hence, even when the model is correct, the ridge in (γ_p, γ_w) dimensions makes it hard to unbiasedly select these parameters, even though the economic bias is minor. Second, responses to monetary shocks can not distinguish models featuring price indexation from models missing this feature (compare cases 1 and 2); it possible to confuse a model with no price stickiness and no wage indexation with a model where these two features exist but no price indexation is present (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). 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. Also in this case, taking the estimates producing the infimum of the objective function 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 investigators looking at this graph would have no doubt that she has nailed down the correct model! Can these problems can 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 important to stress the results we present are obtained in the ideal conditions in which the population responses are available. Clearly, the observationally equivalent problem 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.

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 producing partial identification 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. The observational equivalence problem has not disappeared: fixing one dimension of indeterminacy (and fixing it about right) does not help in estimating γ_p.

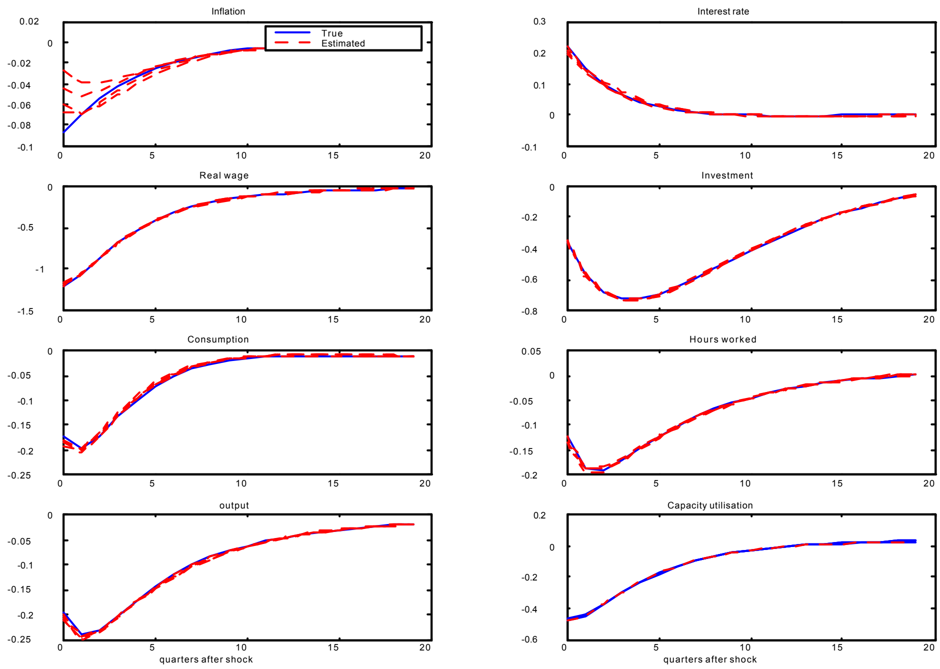

Figure 11: Impulse responses, Case 5

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 of the model, 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 reliable or not.

6 Detecting identification problems

One way to respond to the results we have presented is to argue that the models we considered are not even close to the true DGP of the real world. Therefore, our exercises are irrelevant and the models should not be used to fit the data or to conduct policy analysis. We think this conclusion is unwarranted, first because there are few operational alternatives to the models we have examined and, second, because these setups are likely to face similar problems unless the complexities due to optimizing agents, budget constraints and market clearing conditions are removed from the setup.

Rather than denying their existence, we find it more useful to think about 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 be strengthened using formal methods. As seen in section 2, three conditions need to be satisfied for proper identification. First, the objective function must have a unique zero. Second, the Hessian at the zero must be positive definite and possesses full rank. Third, its curvature must be "sufficient". Since under, partial and weak identification all induce Hessians which are rank deficient or fail to have sufficient curvature, we concentrate on this latter property.

How do one check for the rank of the Hessian? Cragg and Donald (1997) have provided a procedure to do this. Let $h = vech(H)$ and let $d(L) = (h-p)'(h-p)$, where $p = vech(P)$ and P is a matrix of rank L. Under regularity conditions, when an estimate $\hat{h}$ is available, $Td(L) \to \chi^2$, where the degrees of freedom are $(K-L)(K-L-1)/2-K$, $K(K+1)/2$ 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 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).

We apply this last test to the Hessian of the objective function of the model of section 5 at the values estimated in case 5. The Hessian is calculated using the outer product of the gradient produced by the minimization routine. The test confirms the presence of significant rank deficiencies. In fact, 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 of the model 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. Staring at figure 9, 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, several 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., $h, \beta, \lambda_y, \lambda_\pi, \nu$; 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. The data appears to be informative only for the parameters for which we had information in the smaller model, (i.e. the risk aversion coefficient ϕ, the price stickiness ζ_p), for the inverse elasticity of investments with respect to Tobin′s q χ, and, partially, for the wage stickiness ζ_w. Once again, the fact that the low of motion of the states of the model is roughly insensitive to variations of these structural parameters in a neighborhood of the estimated values is responsible for the lack of curvature [5].

7 Conclusions and suggestions for empirical practice

Liu (1960) and, twenty years later, 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 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 microfundations were added, general equilibrium features taken into account, and economic measures of fit designed. Still, it appears that a large class of popular log-linearized 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 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. In fact, lack of information makes models untestable: no experiment will ever be to contradict prior restrictions and, viceversa, prior restrictions appear always to be satisfied in the data.

[5] Anderson's test depends on the true values of the structural parameters. Since it is difficult to produce consistent estimates when identification problems are present, one may want to repeat the test at a number of points and take, e.g., the supremum of the sum of the first k' eigenvalues relative to the average eigenvalue. By usual arguments (see Wright (2003)), the test computed this way is conservative in the sense that the null hypothesis will be rejected less often than the nominal size.

A study of identification issues like ours, beside ringing a warning bell, 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 (or the likelihood function) is highly non-linear; the mapping is unknown and only an approximation is available; problems are multidimensional and simple 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 shape of 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, testing the rank of the Hessian (or the magnitude of its smaller eigenvalues) provides formal statistical 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. Furthermore, since these tests are simple to compute and, in principle, applicable to any point in the parameter space, 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 we have seen in the case of a simple RBC model - as could the use of several limited information objective functions - as we have done with the simple New-Keynesian model. Working with small versions of large models or with portions of their dynamic implications is the only constructive way to understand source of identification failures and help with model respecification. Fourth, and practically speaking, the smaller is the number of 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 - both in terms of variables, number responses and number of structural shocks - de facto eliminating the hedge that limited information approaches have over likelihood methods, both of classical or Bayesian flavours. Fifth, lacking prior information on the structural parameters, one could attempt to obtain estimates via S-sets, as suggested by Stock and Wright (2000), rather than minimize the distance between impulse responses. Such an approach is appealing because estimates are consistent and robust to identification problems. 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. (2004) Firm specific capital etc. Northwestern University, manuscript.

[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] 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.

[8] 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.

[9] Cogley, T., Colacito, R., Sargent, T. (2005) Benefits from US Monetary Policy Experimentation in the days of Samuelson and Solow and Lucas, NYU manuscript.

[10] Cragg, J. and Donald, S. (1997) Inferring the rank of a matrix, *Journal of Econometrics* 76, 223-250.

[11] 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*.

[12] Del Negro, M, Schorfheide, F., Smets, F. And R. Wouters (2005) On the fit of New Keynesian models, CEPR working paper 4848.

[13] 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.

[14] 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.

[15] 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.

[16] Hsiao, C. (1983) Identification in Griliches, Z. and Intrilligator, M. (eds.), *Handbook of Econometrics*, I, 224-283.

[17] Kim, J. (2003) Functional equivalence between intertemporal and multisectoral investment adjustment costs, *Journal of Economic Dynamics and Control*, 27, 533-549.

[18] Ireland, P. (2004) Technology Shocks in the New Keynesian model, *Review of Economics and Statistics*, 86, 923-936.

[19] Linde', J. (2005), Estimating New Keynesian Phillips' curve: A full information maximum likelihood approach, *Journal of Monetary Economics*, 52, 1135-1149.

[20] Liu, T. (1960) Underidentification, Structural estimation, and Forecasting, *Econometrica*, 28, 855-865.

[21] Lubik, T. and Schorfheide, F. (2004) Testing for indeterminacy: An application to US Monetary Policy, *American Economic Review*, 94, 190-217.

[22] 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.

[23] Ma, A. (2002) GMM estimation of the new Phillips curve, *Economic Letters*, 76, 411-417.

[24] Nason, J. and Smith, G. (2005) Identifying the New Keynesian Phillips Curve, Federal Reserve of Atlanta, working paper 2005-1.

[25] Pesaran, H. (1981) identification of Rational Expectations Models, *Journal of Econometrics*, 16, 375-398.

[26] Poirier, D. (1998), Revising Beliefs in Nonidentifed Models, *Econometric Theory*, 14, 183-209

[27] Rothenberg, T. (1971) Identification in parametric models, *Econometrica*, 39, 577-591.

[28] Rotemberg, J. and Woodford, M. (1997) , An Optimization based Econometric Framework for the evaluation of Monetary Policy, *NBER Macroeconomic Annual*, 12, 297-346.

[29] Rubio, J and Rabanal, P. (2005) Comparing New-Keynesian Models of the Business Cycle: A Bayesian Approach, *Journal of Monetary Economics*, 52, 1150-1162.

[30] Sims, C. (1980) Macroeconomic and Reality, *Econometrica*, 48, 1-48.

[31] 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.

[32] Smith, A. (1993) Estimating Nonlinear Time series models using simulated VARs, *Journal of Applied Econometrics*, 8, 63-84.

[33] Stock, J. and Wright, J. (2000) GMM with weak identification, *Econometrica*, 68, 1055-1096.

[34] 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.

[35] Wright, J. (2003) Detecting lack of identification in GMM, *Econometric Theory*, 19, 322-330.