Euro area inflation persistence in an estimated nonlinear dsge model

Gianni Amisano, Oreste Tristani

www.elsevier.com/locate/jedc

Euro area inflation persistence in an estimated nonlinear dsge model

Gianni Amisano, Oreste Tristani

PII: S0165-1889(10)00091-6
DOI: doi:10.1016/j.jedc.2010.05.001
Reference: DYNCON 2412

To appear in: *Journal of Economic Dynamics & Control*

Received date: 3 May 2007
Accepted date: 22 July 2009

Cite this article as: Gianni Amisano and Oreste Tristani, Euro area inflation persistence in an estimated nonlinear dsge model, *Journal of Economic Dynamics & Control*, doi:10.1016/j.jedc.2010.05.001

Euro area inflation persistence in an estimated nonlinear DSGE model

Gianni Amisano *

DG-Research European Central Bank, Kaiserstrasse 29, D - 60311, Frankfurt, Germany. Department of Economics University of Brescia

Oreste Tristani

DG-Research European Central Bank, Kaiserstrasse 29, D - 60311, Frankfurt, Germany

JEL classification: C11, C15, E31, E32, E52

Abstract

We estimate the approximate nonlinear solution of a small DSGE model on euro area data, using the *conditional* particle filter to compute the model likelihood. Our results are consistent with previous findings, based on simulated data, suggesting that this approach delivers sharper inference compared to the estimation of the linearised model. We also show that the nonlinear model can account for richer economic dynamics: the impulse responses to structural shocks vary depending on initial conditions selected within our estimation sample

Key words: DSGE models, inflation persistence, second order approximations, sequential Monte Carlo, Bayesian estimation.

* Corresponding author
Email addresses: gianni.amisano@ecb.europa.eu, amisano@eco.unibs.it (Gianni Amisano), oreste.tristani@ecb.europa.eu (Oreste Tristani).

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have become popular tools for monetary policy analysis. The central feature of these models, emphasized in the theoretical work of Yun (1996) and Woodford (2003), is the presence of nominal rigidities in the adjustment of goods prices. More recently, a number of additional frictions have been introduced in the basic sticky-price framework and the resulting models have been successfully taken to the data (see e.g. Christiano, Eichenbaum and Evans, 2005; Smets and Wouters, 2003).

In all cases, however, what is estimated is only the reduced form emerging from the solution of a *linearized* version of those models. This approach has obvious advantages in terms of simplicity and possibility of comparison to other well-known empirical tools, such as VARs. There are a number of reasons, however, to also be interested in exploring the implications of the many *nonlinear* features built in DSGE models.

The first one is that they are more suited to characterize macroeconomic dynamics in presence of large deviations from the steady state. Since 1970, average euro area inflation has reached a maximum and minimum of 14.56 and 0.69 percent, respectively, compared to an average of 5.83 percent [1]. By construction, a linearized model is ill-suited to explain such large deviations and it might deliver distorted estimates, at least in principle, if forced to do so.

More specifically, it is conceivable that the dynamic properties of inflation should depend on its distance from the steady state. Small deviations could be characterized by a relatively small degree of persistence and/or amplitude of inflation responses to shocks. Persistence and amplitude could become more pronounced in case of larger deviations, when the inflationary shock could more easily become entrenched in expectations. These economic features can be captured by higher-order terms in a nonlinear solution, terms which, by construction, would start playing a non-negligible role only when large deviations from steady state do take place. A linearized model, on the contrary, would be forced to account for all observed dynamics with linear terms, thus possibly delivering incorrect estimates.

To test this conjecture, one would ideally solve the model to a high order of accuracy using global solution methods. Unfortunately, however, these methods tend to be too slow for their output to be amenable to econometric estimation. A compromise between the desire to capture the bulk of the nonlinear dynamics of the model and the interest in exploring its empirical performance

[1] These statistics are computed by using the year-on-year percentage changes of the euro area GDP deflator.

is to focus on second-order perturbation methods. These methods produce asymptotically correct expressions around the deterministic steady state of the model**, but there is no guarantee that these solutions will be accurate away from the approximation point. Nevertheless,** second order perturbations are often found to produce more accurate results than linearisations for a range of values of the state variables around the steady state (see e.g. Aruoba, Fernández-Villaverde and Rubio-Ramírez, 2006).

The second reason to be interested in nonlinear models is empirical. Nonlinear models have been argued to provide sharper estimates of the structural parameters than their linearized counterparts. The nonlinearities induce additional testable implications, compared to those characterizing the linearized version of the same model. A straightforward example can be made for the case of solutions obtained through second-order perturbation methods. These approximate solutions imply that the variance of exogenous shocks have an impact on the unconditional means of observable variables. The link amounts to a restriction on the size of those variances, which is ignored in linearized solutions. [2] A number of authors have therefore reported a superior performance of estimates based on the nonlinear model, compared to estimates based on the linearized model (e.g. An and Schorfheide, 2007; Fernandez-Villaverde and Rubio-Ramirez, 2006, 2007). However, these results are mostly based on simulated data, drawn by construction from the "true", nonlinear model. It is obviously interesting to test if nonlinear estimates can also do better on actual data, where the model is only an approximation of reality. It is in fact conceivable that the tighter theoretical constraints imposed in the estimation of a nonlinear model may result in a worse fit, when compared to a linearized version of the same model.

We provide new evidence on these issue, based a relatively standard DSGE model solved using second-order perturbation methods. We estimate the model on euro area data over the 1970-2004 period using sequential Monte Carlo methods to construct the likelihood. Our results highlight that nonlinearities have played a non-negligible role over the past three decades.The nonlinear model tends to perform consistently better than the linear over the whole sample and especially when inflation is high.

Concerning the dynamic features of euro area inflation, our preferred specification shows that notable differences can be found between linear and nonlinear estimates of our DSGE model. The amplitude and persistence of the responses of inflation to shocks differ, depending on whether they are computed starting from the "high inflation" values of the seventies, or from the "low inflation"

[2] Fernandez-Villaverde, Rubio-Ramirez and Santos (2006) highlight a more general empirical advantage of the estimation of nonlinear models, which has to do with the approximation errors made when computing the the likelihood function.

levels observed in recent years. For example, a positive inflation target shock has temporarily expansionary effects on output, if it occurs in a low-inflation environment, while it has contractionary effects if it takes place when inflation is high.

The rest of the paper is organized as follows. Section 2 provides a broad description of the two theoretical models employed in the empirical exercise. The main difference between those models concerns the behaviour of monetary policy. While always following a Taylor-type rule, the central bank is assumed to have a stationary stochastic inflation target in the first case and an integrated target in the second case. Section 3 discusses briefly the solution method. It is well-known that approximate nonlinear solutions can be computed using a variety of methods (see Aruoba, Fernandez-Villaverde and Rubio-Ramirez, 2006). We focus on second-order perturbation methods, because they are a direct extension of the standard linearisation and because they are fast to implement. The estimation methodology is presented next, in Section 4, with particular emphasis on the construction of the likelihood function, which is performed using a method not previously used in macroeconomic applications: the conditional particle filter. We also discuss briefly some of the choices available to the researcher in this context and the importance of a plausible specification of the priors for the variance of the shocks. Section 5 presents the estimation results, including posterior means of the parameters and comparisons of stationary and nonstationary, linear and nonlinear models. We also look at nonlinear impulse responses and document the differences which can be observed starting from different points in time. Section 6 concludes. Appendix (A) describes the structures of the models being estimated and and Appendiz (B) explains how the solution method is applied.

2 The theoretical framework

One of the conclusions of the "Inflation Persistence Network" (IPN) coordinated by the European Central Bank is that different estimates of the persistence of aggregate euro area inflation are obtained depending on whether the researcher allows, or not, for shifts in the inflation mean (see e.g. Angeloni *et al.*, 2005). Empirical estimates of inflation persistence are high if a single inflation mean is assumed, while they fall considerably in the second case. For example, Bilke (2005) argues that a structural break in French CPI inflation occurred in the mid-eighties. Controlling for this break, both aggregate and sectoral inflation persistence are stable and low. Levin and Piger (2004) also find strong evidence for a break in the mean of inflation in the late 1980s or early 1990s for twelve industrial countries. Allowing for such break, the inflation measures generally exhibit relatively low inflation persistence. Similar results are obtained by Corvoisier and Mojon (2005) for most OECD coun-

tries. Dossche and Everaert (2005) find similar results when they allow for shifts in the inflation target in the form of a random walk.

By and large, the existence of shifts in the mean of inflation has been tested within statistical or reduced-form frameworks (see e.g. Levin and Piger, 2004; Corvoisier and Mojon, 2005). As a result, it could be argued that there are two difficulties with the interpretation of these results. First, it remains unclear whether the hypothesis of one or more shifts in the mean of inflation would be rejected within a richer model. Secondly, the reasons for a potential shift in the inflation mean are left unspecified, while it would be interesting to understand their determinants.

To shed further light on the first issues, we explore the empirical plausibility of two variants of a simple DSGE model of inflation and output dynamics. The first one is a benchmark model which embodies the assumption of no permanent shifts in the average inflation rate. The second model allows instead for smooth shifts in the mean of inflation through an integrated inflation target. Comparing the empirical performance of these two specifications, we will be able to assess the plausibility of the structural break hypothesis [3].

In the rest of this section, we present in more detail the main features of the microeconomic environment and the two alternative policy rules.

2.1 A simple DSGE model

The model is based on the framework developed by Woodford (2003) and extended in a number of directions by Christiano, Eichenbaum and Evans (2005).

Consumers maximize the discounted sum of the period utility

$$U\left(C_t, H_t, L_t\right) = \frac{\left(C_t - hC_{t-1}\right)^{1-\gamma}}{1-\gamma} - \int_0^1 \chi L_t\left(i\right)^{\phi} di \tag{1}$$

where C is a consumption index satisfying

$$C = \left(\int_0^1 C\left(i\right)^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}}, \tag{2}$$

$H_t = hC_{t-1}$ is the habit stock, $L\left(i\right)$ are hours of labour provided to firm i.

[3] Our results are obviously contingent on our particular model of the evolution of the inflation mean.

For consistency with Smets and Wouters (2003) and Christiano, Eichenbaum and Evans (2005), habit formation is modelled in difference form. However, habit is internal, so that households care about their own lagged consumption.

The household's budget constraint is given by

$$P_t C_t + B_t \leqslant \int_0^1 w_t(i) H_t(i) di + \int_0^1 \Pi_t(i) di + W_t \tag{3}$$

where $\Pi_t(i)$ are profits received from investment in firm i, B_t denotes end of period holdings of a complete portfolio of state contingent assets, W_t denotes the beginning of period value of the assets and $w_t(i)$ is the nominal wage rate. The price level P_t defined as the minimal cost of buying one unit of C_t, hence equal to

$$P_t = \left(\int_0^1 p(i)^{1-\theta} \right)^{\frac{1}{1-\theta}}. \tag{4}$$

In the budget constraint, B_t denotes end of period holdings of a complete portfolio of state contingent assets. W_t denotes the beginning of period value of the assets, $w_t(i)$ is the nominal wage rate, $L_t(i)$ is the supply of labor to firm i and $\Xi_t(i)$ are the profits received from investment in firm i. Following Steinsson (2003), we also introduce a stochastic income tax, which will lead to a trade-off between inflation and the output gap. We write the tax rate as $\frac{\tau_t}{1+\tau_t}$ to ensure that the total tax is bounded between 0 and 1, given that

$$\log \tau_t = (1-\rho_\tau)\overline{\tau} + \rho_\tau \log \tau_{t-1} + v_t^\tau, \qquad v_t^\tau \sim N\left(0, \sigma_\tau^2\right). \tag{5}$$

The first order conditions w.r.t intertemporal aggregate consumption allocations and labour supply can be written as

$$\begin{aligned} \left(1 - \frac{\tau_t}{1+\tau_t}\right) \frac{w_t(i)}{P_t} &= \frac{\phi \chi L(i)^{\phi-1}}{\Lambda_t} \\ \Lambda_t &= (C_t - hC_{t-1})^{-\gamma} - \beta h \mathrm{E}_t\left[(C_{t+1} - hC_t)^{-\gamma}\right] \\ \frac{1}{I_t} &= \mathrm{E}_t\left[\beta \frac{P_t}{P_{t+1}} \frac{\Lambda_{t+1}}{\Lambda_t}\right]. \end{aligned} \tag{6}$$

where I_t is the gross nominal interest rate.

Turning to the firm's problem, the production function is given by

$$Y_t(i) = A_t L(i)^\alpha, \; A_t = A_{t-1}^{\rho_a} e^{v_t^a} \tag{7}$$

where A_t is a technology shock and v_t^a is a normally distributed innovation with constant variance σ_a^2.

We assume Calvo (1983) contracts, so that firms face a constant probability ζ of being unable to change their price at each point in time t. Firms will take this constraint into account when trying to maximize expected profits, namely

$$\max_{P_t^i} \mathrm{E}_t \sum_{s=t}^{\infty} \zeta^{s-t} \beta^s \frac{P_t}{P_{t+s}} \frac{\Lambda_{t+s}}{\Lambda_t} \left(P_s^i Y_s^i - TC_s^i \right), \tag{8}$$

where TC denotes total costs and, as in Smets and Wouters (2003), firms not changing prices optimally are assumed to modify them using a rule of thumb that indexes them partly to lagged inflation and partly to steady-state inflation $\overline{\Pi}$, namely $P_t^i \left(\overline{\Pi}\right)^{1-\iota} \left(\frac{P_{s-1}}{P_{t-1}}\right)^{\iota}$, where $0 \leq \iota \leq 1$. When we assume an integrated inflation target, steady state inflation is not defined and we set $\iota = 1$. We introduce indexation in the model for two reasons. First, aggregate inflation will be driven to some extent by lagged inflation, which is an empirically plausible hypothesis – though not immediately consistent with the microeconomic evidence. Second, firms not allowed to update their prices optimally for a long time will still find themselves with a price which is not too far from the optimum.

Under the assumption that firms are perfectly symmetric in all other respects than the ability to change prices, all firms that do get to change their price will set it at the same optimal level P_t^*. The first order conditions of the firms' problem can be written recursively as implying

$$\begin{aligned}
\left(\frac{P_t^*}{P_t}\right)^{1-\theta\left(1-\frac{\phi}{\alpha}\right)} &= \frac{\phi\chi\theta}{\alpha\left(\theta-1\right)} \frac{\Upsilon_{2,t}}{\Upsilon_{1,t}} \\
\Upsilon_{2,t} &= \frac{A_t^{-\frac{\phi}{\alpha}}}{\left(1-\frac{\tau_t}{1+\tau_t}\right)\Lambda_t} Y_t^{\frac{\phi}{\alpha}} + \mathrm{E}_t \zeta \overline{\Pi}^{-\theta\frac{\phi}{\alpha}(1-\iota)} \beta \frac{\Lambda_{t+1}}{\Lambda_t} \Upsilon_{2,t+1} \Pi_t^{-\theta\frac{\phi}{\alpha}\iota} \Pi_{t+1}^{\theta\frac{\phi}{\alpha}} \\
\Upsilon_{1,t} &= Y_t + \mathrm{E}_t \zeta \overline{\Pi}^{(1-\theta)(1-\iota)} \beta \frac{\Lambda_{t+1}}{\Lambda_t} \Upsilon_{1,t+1} \Pi_t^{(1-\theta)\iota} \Pi_{t+1}^{\theta-1}
\end{aligned} \tag{9}$$

where Π_t is the inflation rate defined as $\Pi_t \equiv \frac{P_t}{P_{t-1}}$ and

$$\frac{P_t^*}{P_t} = \left(\frac{1-\zeta\left(\overline{\Pi}^{1-\iota} \frac{\Pi_{t-1}^{\iota}}{\Pi_t}\right)^{1-\theta}}{(1-\zeta)} \right)^{\frac{1}{1-\theta}} \tag{10}$$

expresses the optimal price at time t as a function of aggregate variables. [4]

Note that we can use equation (10) in the system (9) to write aggregate

[4] Similar nonlinear expressions are used, amongst others, by Ascari (2004) and Benigno and Woodford (2005).

inflation as an implicit function of expected future inflation

$$\Pi_t = \overline{\Pi}^{1-\iota}\Pi_{t-1}^{\iota}\left(\frac{1}{\zeta} - \frac{1-\zeta}{\zeta}\left(\frac{\phi\chi\theta}{\alpha\left(\theta-1\right)}\frac{\Upsilon_{2,t}}{\Upsilon_{1,t}}\right)^{\frac{1-\theta}{1-\theta\left(1-\frac{\phi}{\alpha}\right)}}\right)^{\frac{1}{\theta-1}} \tag{11}$$

It is well known that a first order approximation of this equation yields the new-Keynesian Phillips curve, where inflation is positively related to expected future inflation. The second-order approximation of equation (11) is more elaborate, so that the relationship between current and future inflation is not immediately apparent (see Benigno and Woodford, 2005, for an example in the simpler case without habits nor inflation indexation). Nevertheless, equation (11) is suggestive of two features.

First, past inflation only enters log-linearly in the equation, since it never appears in the $\Upsilon_{2,t}$ and $\Upsilon_{1,t}$ terms. Even with indexation, the fact that past inflation is high does not *per se* matter in inducing a nonlinearity in inflation as a function of the state of the economy. Indexation does, however, matter in changing expectations of future inflation.

The second known feature of equation (11) is that its quadratic approximation will be either concave or convex, regardless of whether inflation deviations from the long run mean are positive or negative. The effects of the second order terms in the solution will therefore be asymmetric. If, *ceteris paribus*, inflation is a convex function of expected future inflation, firms will try to increase current prices more and more aggressively, the larger is the expected future deviation of inflation from steady state. They will however cut their prices less than they would in the linear case, in case of negative inflation deviations from the steady state.

2.2 Two Taylor rules

Equations (5), (7), (9) and (10) describe aggregate economic dynamics. We close the model with a Taylor rule with interest rate smoothing. A key decision that has to be taken in the specification of the rule concerns the inflation target. Since inflation displays a noticeable downward trend over the sample period, the assumption of a constant target is not very appealing. In empirical applications, it is therefore often assumed that the decline in inflation corresponds to a decline in the inflation target. This is also what we do here. However, this assumption is likely to have important implications in terms of the persistence of inflation. In order to explore this issue, we analyse two variants of the policy rule.

The first rule assumes that the inflation target follows a stationary AR(1)

process. In this case, the idea is that the long run target of the central bank is actually constant, but that there are shifts in the horizon at which the central bank tries to get inflation back to that long run level. If the target is temporarily high when inflation is high, then the central bank is willing to tolerate a slow return to the long run target. If, instead, there are no changes in the long run target when inflation is high, inflation will be brought back on target more quickly.

In logarithmic terms (lower case letters), the first rule takes the form

$$i_t = (1-\rho_I)\left((\overline{\pi} - \ln\beta) + \psi_\pi\left(\pi_t - \pi_t^*\right) + \psi_y\left(y_t - y_t^n\right)\right) + \rho_I i_{t-1} + v_t^i \quad (12)$$

$$\pi_t^* = (1-\rho_\pi)\,\overline{\pi} + \rho_\pi \pi_{t-1}^* + v_t^{\pi^*} \quad (13)$$

where i_t is the logarithm of the gross nominal interest rate, π_t^* is the inflation target, v_t^i is a policy shock and y_t^n is the logarithm of the level of natural output. The innovations v_t^i and $v_t^{\pi^*}$ are white noise with variances σ_i^2 and $\sigma_{\pi^*}^2$, respectively. In this model, considerable deviations from the mean of inflation can arise from short-term movements in the inflation target. The model solved using the first policy rule is dubbed M1.

The second policy rule is identical to the first, except for the property that the inflation target becomes integrated (and the steady state level of the interest rate is modified accordingly)

$$i_t = (1-\rho_I)\left((\pi_t^* - \ln\beta) + \psi_\pi\left(\pi_t - \pi_t^*\right) + \psi_y\left(y_t - y_t^n\right)\right) + \rho_I i_{t-1} + v_t^i \quad (14)$$

$$\pi_t^* = \pi_{t-1}^* + v_t^{\pi^*} \quad (15)$$

In this case, smooth changes of the inflation mean occur over time as the central bank target is revised. The idea here is that the inflation target process captures true shifts in the objective of the central bank. Given the slow decline in inflation over our sample period, this should supposedly reflect a shift in public preferences in favour of lower and lower inflation levels. The integrated inflation target induces a non-stationary behaviour also in actual inflation and the nominal interest rate. These nominal variables are also co-integrated, so that the model can be written in stationary form in terms of the rate of growth of inflation, $\Delta\pi_t = \pi_t - \pi_{t-1}$, and the deflated inflation target and interest rate, defined as $\widetilde{\pi}_t^* = \pi_t^* - \pi_t$ and $\widetilde{i}_t = i_t - \pi_t$, respectively (see Appendix A). This model is dubbed M2.

3 Second-order approximate solution

We solve the model using a second order approximation around the non-stochastic steady state. The model dynamics will then be described by two systems of equations: a quadratic law of motion for the predetermined variables of the model and a quadratic relationship linking each non-predetermined variable to the predetermined variables.

The solution is obtained numerically. A few methods have been proposed in the literature, including those in Schmitt-Grohé and Uribe (2004), Kim, Kim, Schaumburg and Sims (2003, henceforth KKSS), Lombardo and Sutherland (2007). For our applications we select the implementation proposed by Gomme and Klein (2008), that has the advantage of being relatively faster. Speed is particularly important for estimation, since the model needs to be solved at every evaluation of the likelihood. For this reason, we also rely on analytical derivatives to evaluate the second order terms of the approximation.

The solution can be written as follows. The vector $\mathbf{x}_t$ of predetermined variables, expressed in terms as deviations from its non-stochastic steady state value, follows the quadratic law of motion

$$\mathbf{x}_{t+1} = \frac{1}{2}\mathbf{h}_{\sigma\sigma} + \mathbf{H}_x\mathbf{x}_t + \frac{1}{2}\mathbf{H}_{xx}\left(\mathbf{x}_t \otimes \mathbf{x}_t\right) + \sigma\mathbf{J}\mathbf{v}_{t+1} \tag{16}$$

$$\mathbf{v}_{t+1} \backsim NID(0, \mathbf{I}_{n_s}) \tag{17}$$

where $\mathbf{h}_{\sigma\sigma}, \mathbf{H}_x$ and $\mathbf{H}_{xx}$ are $n_x \times 1, n_x \times n_x$, and $n_x \times n_x^2$ matrices, respectively. The vector of shocks has variance covariance matrix $\mathbf{I}_{n_s}$, where n_s is typically different from n_x. The scalar σ is the perturbation parameter: when $\sigma = 0$ the system becomes deterministic. Non-predetermined variables, $\mathbf{y}_t$, also expressed as deviations from their non-stochastic steady-state values, are linked to predetermined variables by the solution

$$\mathbf{y}_t = \frac{1}{2}\mathbf{g}_{\sigma\sigma} + \mathbf{G}_x\mathbf{x}_t + \frac{1}{2}\mathbf{G}_{xx}\left(\mathbf{x}_t \otimes \mathbf{x}_t\right) \tag{18}$$

where $\mathbf{g}_{\sigma\sigma}$, $\mathbf{G}_x$ and $\mathbf{G}_{xx}$ are $n_y \times 1$, $n_y \times n_x$ and $n_x \times n_x n_y$ matrices, respectively. The definitions of state vectors in the models analysed in this paper are contained in Appendix (B).

We tested the accuracy of the solution obtained using our second order approximations focusing on 1-step ahead errors in the Euler equation and on our favourite econometric specification (i.e. the one referred to as "model M1" below). There are different dimensions over which Euler equation errors can be computed and we explored many of them at the posterior mean of the parameter values. First,

we computed errors at each point in time, starting from the filtered value of the state vector. Second, we computed Euler equation errors: (a) along a simulation path, (b) on an ellipse around the mean of the state variables, and (c) for innovations of various size. [5] **Overall, the second order approximation tends to be more accurate than the linearisation of the model, but the errors it produces can become non-negligible for some particular realisations of the state vector. More specifically, the maximum of the base-10 logarithm of the normalised, absolute Euler equation error can reach the value -1.1, which corresponds to an error of almost 8%.**

3.1 Simulation and impulse responses

Some care needs to be taken when simulating the approximate second-order solution (16)-(18). KKSS emphasise that a standard simulation procedure would introduce undesired higher order elements in the simulated path. Such elements are compounded over the simulation period and could conceivably lead to explosionary paths.

We therefore follow the alternative recursive approach suggestion by KKSS. This approach is also the basic intuition for the solution method proposed in Lombardo and Sutherland (2007). The approach amounts to using jointly both the second order solution (16)-(18) and the first order solution

$$\mathbf{x}_{t+1} = \mathbf{H}_x \mathbf{x}_t + \sigma \mathbf{J} \mathbf{v}_{t+1}. \tag{19}$$

Given past realisations of the first order state vector, $\mathbf{x}_t^L$, and second order state vector, $\mathbf{x}_t^Q$, we proceed as follows:

(1) draw $\mathbf{v}_{t+1}^L$ and simulate $\mathbf{x}_{t+1}^L$

$$\mathbf{x}_{t+1}^L = \mathbf{H}_x \mathbf{x}_t^L + \sigma \mathbf{J} \mathbf{v}_{t+1}^L$$

(2) construct $\mathbf{x}_{t+1}^Q$ without further simulations (i.e. using $\mathbf{v}_{t+1}^L$) as

$$\mathbf{x}_{t+1}^Q = \frac{1}{2}\mathbf{h}_{\sigma\sigma} + \mathbf{H}_x \mathbf{x}_t^Q + \frac{1}{2}\mathbf{H}_{xx}\left(\mathbf{x}_t^L \otimes \mathbf{x}_t^L\right) + \sigma \mathbf{J} \mathbf{v}_{t+1}^L$$

(3) construct the second order jump vector

$$\mathbf{y}_{t+1} = \frac{1}{2}\mathbf{g}_{\sigma\sigma} + \mathbf{G}_x \mathbf{x}_{t+1}^Q + \frac{1}{2}\mathbf{G}_{xx}\left(\mathbf{x}_{t+1}^L \otimes \mathbf{x}_{t+1}^L\right)$$

[5] We relied on an implementation of our model in dynare++ to compute the second three methods.

(4) Go back to 1.

This issue is particularly important for us, since we make extensive use of simulation methods in the rest of the paper. More specifically, we use these methods to:

- compute the likelihood, relying on the conditional particle filter (see next section);
- compute nonlinear impulse responses, which are obtained simply going through steps 2-3 described above, starting from a certain value $\mathbf{x}_t = \overline{\mathbf{x}}$, and given two different paths for the structural shocks $[\mathbf{v}_{t+1}^{(1)}, \mathbf{v}_{t+2}, ..., \mathbf{v}_{t+h}]$ and $[\mathbf{v}_{t+1}^{(2)}, \mathbf{v}_{t+2}, ..., \mathbf{v}_{t+h}]$ (see section 5.4).

4 Estimation method

4.1 Non linear-non Gaussian state space models

The system (16)-(18) can be cast in the general form

$$\text{(measurement equation) } \mathbf{y}_t^o = G(\mathbf{x}_t, \mathbf{w}_t, \boldsymbol{\theta}) \tag{20}$$

$$\text{(state equation) } \mathbf{x}_t = H(\mathbf{x}_{t-1}, \mathbf{v}_t, \boldsymbol{\theta}) \tag{21}$$

where $\mathbf{y}_t^o$ is the subset of imperfectly observable elements of the vector $\mathbf{y}_t$, $\boldsymbol{\theta}$ is the parameter vector, $\mathbf{v}_t \equiv \left[v_t^a, v_t^{\pi^*}, v_t^\tau, v_t^i\right]'$ is the vector of structural shocks and $\mathbf{w}_t$ are measurement errors.

In order to be able to do inference on the unobservables (parameters and state vector) we need to solve a filtering problem, i.e. given $p(\mathbf{x}_t|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta})$ obtain $p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_{t+1}^o, \boldsymbol{\theta}), t = 0, 1, ...T-1$, where

$$\underline{\mathbf{y}}_t^o = \left[\mathbf{y}_1^{o\prime} \ \mathbf{y}_2^{o\prime} \ ... \ \mathbf{y}_t^{o\prime}\right]' \tag{22}$$

collects all the data evidence up to time t.

The filtering problem is conceptually straightforward and consists of two steps:

- projection

$$p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta}) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t, \boldsymbol{\theta}) p(\mathbf{x}_t|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta}) d\mathbf{x}_t \tag{23}$$

- update

$$p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_{t+1}^o, \boldsymbol{\theta}) = \frac{p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta}) p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1}, \boldsymbol{\theta})}{p(\mathbf{y}_{t+1}^o|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta})} \tag{24}$$

The filtering recursion yields the likelihood of each observation as

$$p(\mathbf{y}_{t+1}^o|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta}) = \int p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1},\boldsymbol{\theta})d\mathbf{x}_{t+1} \tag{25}$$

When the state space is linear and the shocks are Gaussian, the integral required by the projection and the update steps can be performed analytically giving rise to the Kalman filtering recursion. In our context, with non-linear state and measurement equations, it is necessary to compute those integrals by using either some approximation or numerical methods.

In this paper, the integration steps which are inherent in the filtering recursion are performed using Sequential Monte Carlo methods. For concise and effective introductions to these methods, see Arulampalam *et al.* (2002), Doucet *et al.* (2001). To date, sequential Monte Carlo methods have been used in statistics and (only marginally) in financial econometrics, while in macroeconomic applications they have been used very seldom. Fernandez-Villaverde and Rubio-Ramirez (2006, 2007a) and An and Schorfheide (2007) are the first studies in which these techniques are used for DSGE models, while in Casarin and Trecroci (2006) these methods are used to investigate the dynamics of univariate volatilities of macroeconomic aggregates.

Sequential Monte Carlo techniques are numerically very expensive and quite involved but they offer the advantage to deal with more realistic models. In a DSGE framework, these methods can be used also to deal with non-Gaussian shocks, with non-constant volatility (see Fernandez-Villaverde 2007a, Amisano and Tristani 2009b), or with time-varying parameters (Fernandez-Villaverde and Rubio-Ramirez 2007b)

The intuition behind the simplest version of these methods, which is called the *particle filter* is to perform the filtering recursion and compute the likelihood $p(\mathbf{y}_{t+1}^o|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})$. The particle filter is based on the following identity

$$\begin{aligned} p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_{t+1}^o,\boldsymbol{\theta}) &= \int p(\mathbf{x}_{t+1}|\mathbf{x}_t,\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})p(\mathbf{x}_t|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta}) \cdot \frac{p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1},\mathbf{x}_t,\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})}{p(\mathbf{y}_{t+1}^o|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})}d\mathbf{x}_t \\ &= \int p(\mathbf{x}_{t+1}|\mathbf{x}_t,\boldsymbol{\theta})p(\mathbf{x}_t|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})\frac{p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1},\boldsymbol{\theta})}{p(\mathbf{y}_{t+1}^o|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})}d\mathbf{x}_t \end{aligned} \tag{26}$$

Suppose we have a sample of size N of draws from the distribution $p(\mathbf{x}_t|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta})$ (a *swarm* of N *particles*)

$$\mathbf{x}_t^{(i)} \backsim p(\mathbf{x}_t|\underline{\mathbf{y}}_t^o,\boldsymbol{\theta}), i = 1,2,...,N \tag{27}$$

then it is possible to obtain a sample of N draws from the distribution $p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_{t+1}^o,\boldsymbol{\theta})$

applying the following three steps:

(1) (projection) draw a large number of realisations from the distribution of $\mathbf{x}_{t+1}$ conditioned on $\underline{\mathbf{y}}_t^o$; this amounts to simulating the state equation

$$\mathbf{x}_{t+1}^{(i)} \backsim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)}, \boldsymbol{\theta}), i = 1, 2, ..., N;$$

(2) (update) assign to each draw a weight which is determined by its "distance" from (compatibility with) $\mathbf{y}_{t+1}^o$. The weight assigned to each of the draws is $p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1}, \boldsymbol{\theta})$

$$w_{t+1}^{(i)} = p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1}, \boldsymbol{\theta}), i = 1, 2, ..., N; \tag{28}$$

(3) resample (with re-immission) the draws $\mathbf{x}_{t+1}^{(i)}$ using probabilities

$$p_{t+1}^{(i)} = \frac{w_{t+1}^{(i)}}{\sum_{j=1}^{N} w_{t+1}^{(j)}}. \tag{29}$$

Note that the unnormalised weights (28) are very important for inference: their sample mean is the t^{th} observation conditional density:

$$\frac{1}{N}\sum_{j=1}^{N} p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1}^{(j)}, \boldsymbol{\theta}) \approx \iint p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1}, \boldsymbol{\theta}) p(\mathbf{x}_{t+1}|\mathbf{x}_t, \boldsymbol{\theta}) p(\mathbf{x}_t|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta}) d\mathbf{x}_{t+1} d\mathbf{x}_t =$$
$$= p(\mathbf{y}_{t+1}^o|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta}) \tag{30}$$

This likelihood can be used as a basis for full information inference (Bayesian or not) on the parameters of the model, while the whole filtering procedure can be used for carrying out smoothed or filtered inference on the unobservable variables.

If we call $p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}_t^o, \boldsymbol{\theta})$ the prior distribution (prior to observing $\mathbf{y}_{t+1}^o$) and $p(\mathbf{y}_{t+1}^o|\mathbf{x}_{t+1}, \boldsymbol{\theta})$ the "likelihood", the particle filter algorithm can be given a very simple Bayesian interpretation which immediately clarifies its limitations: it is as if we were doing posterior simulation drawing from the prior and then using the likelihood as weights. This is a very straightforward procedure to implement but hardly a computationally efficient one in the case the "likelihood" is much more concentrated than the "prior".

It is well-known that the particle filter can be quite inefficient, especially in the presence of outliers in the data or in situations in which the measurement error is nearly absent. A few variants have therefore been proposed in the literature, including the auxiliary-variable particle filter and the conditional particle filter. Details on the relative merits of each of them in a DSGE framework can be found in Amisano and Tristani (2009a). Here we focus on the

conditional particle filter, proposed by Ionides (2007), which displays useful properties when dealing with data characterised by a low signal-to-noise ratio (see Fernandez-Villaverde and Rubio-Ramirez, 2006b, on the difficulties posed by the absence of measurement errors for the particle filter).

The conditional particle filter is based on the following identity

$$\begin{aligned} p(\mathbf{x}_{t+1}|\underline{\mathbf{y}}^o_{t+1},\boldsymbol{\theta}) &= \int p(\mathbf{x}_{t+1}|\mathbf{x}_t,\underline{\mathbf{y}}^o_{t+1},\boldsymbol{\theta})\cdot p(\mathbf{x}_t|\underline{\mathbf{y}}^o_{t+1},\boldsymbol{\theta})d\mathbf{x}_t = \\ &= \int p(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}^o_{t+1},\boldsymbol{\theta})\cdot p(\mathbf{x}_t|\underline{\mathbf{y}}^o_{t},\boldsymbol{\theta})\cdot \frac{p(\mathbf{y}^o_{t+1}|\mathbf{x}_t,\underline{\mathbf{y}}^o_{t},\boldsymbol{\theta})}{p(\mathbf{y}^o_{t+1}|\underline{\mathbf{y}}^o_{t},\boldsymbol{\theta})}d\mathbf{x}_t \\ &= \int p(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}^o_{t+1},\boldsymbol{\theta})\cdot p(\mathbf{x}_t|\underline{\mathbf{y}}^o_{t},\boldsymbol{\theta})\cdot \frac{p(\mathbf{y}^o_{t+1}|\mathbf{x}_t,\boldsymbol{\theta})}{p(\mathbf{y}^o_{t+1}|\underline{\mathbf{y}}^o_{t},\boldsymbol{\theta})}d\mathbf{x}_t \quad (31) \end{aligned}$$

The conditional particle filter is applied by repeating the following steps:

(1) (projection) draw $\mathbf{x}^{(i)}_{t+1}$ from the distribution of $\mathbf{x}_{t+1}$ conditioned on $\underline{\mathbf{y}}^o_{t+1}$

$$\mathbf{x}^{(i)}_{t+1} \backsim p(\mathbf{x}_{t+1}|\mathbf{y}^o_{t+1},\mathbf{x}^{(i)}_t,\boldsymbol{\theta}), i = 1,2,...,N;$$

(2) (update) assign weights

$$w(\mathbf{x}^{(i)}_{t+1}) = p(\mathbf{y}^o_{t+1}|\mathbf{x}^{(i)}_t,\boldsymbol{\theta});$$

(3) resample (with re-immission) the draws $\mathbf{x}^{(i)}_{t+1}$ using probabilities $p^{(i)}_{t+1} = \frac{w^{(i)}_{t+1}}{\sum_{j=1}^{N} w^{(j)}_{t+1}}$.

Note that, once again, the sample mean of the weights $w^{(i)}_{t+1}$ is the conditional likelihood of $\mathbf{y}^o_{t+1}$

$$\begin{aligned} \frac{1}{N}\sum_{j=1}^{N} p(\mathbf{y}^o_{t+1}|\mathbf{x}^{(j)}_t,\boldsymbol{\theta}) \approx \iint p(\mathbf{y}^o_{t+1}|\mathbf{x}_t,\boldsymbol{\theta})p(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}^o_{t+1},\boldsymbol{\theta})p(\mathbf{x}_t|\underline{\mathbf{y}}^o_t,\boldsymbol{\theta})d\mathbf{x}_{t+1}d\mathbf{x}_t = \\ = p(\mathbf{y}^o_{t+1}|\underline{\mathbf{y}}^o_t,\boldsymbol{\theta}) \quad (32) \end{aligned}$$

It is easy to see why this filtering procedure works more efficiently than the particle filter: in drawing $\mathbf{x}_{t+1}$ from $p(\mathbf{x}_{t+1}|\mathbf{y}^o_{t+1},\mathbf{x}^{(i)}_t,\boldsymbol{\theta})$ we use already the information contained in $\mathbf{y}^o_{t+1}$. This feature is called "adaption" and it is the starkest difference with respect to the particle filter algorithm, which, on the other hand, does not use any information on $\mathbf{y}^o_{t+1}$ to draw $\mathbf{x}_{t+1}$. Table (1) conceptually compares how the particle filter and the conditional particle filter work.

The main difficulty with the conditional particle filter is that the distribution $p(\mathbf{x}_{t+1}|\mathbf{y}^o_{t+1}, \mathbf{x}^{(i)}_t, \boldsymbol{\theta})$ is not known analytically when the measurement equation is nonlinear. An approximate solution to this problem is to use a linearization of the measurement equation around the expected future value of the state vector. In this way, we can draw from an approximate probability distribution $\widetilde{p}(\mathbf{x}_{t+1}|\mathbf{y}^o_{t+1}, \mathbf{x}^{(i)}_t, \boldsymbol{\theta})$, the distribution implied by the linearization procedure, and compute the weights $\widetilde{w}(\mathbf{x}^{(i)}_{t+1})$ as proportional to $\widetilde{p}(\mathbf{y}^o_{t+1}|\mathbf{x}^{(i)}_t, \boldsymbol{\theta})$.

To be more specific, consider again our quadratic DSGE state space model with measurement errors

$$\begin{aligned}\mathbf{y}^o_{t+1} &= \frac{1}{2}\mathbf{g}_{\sigma\sigma} + \mathbf{G}_x\left(\mathbf{x}_{t+1}\right) + \frac{1}{2}\mathbf{G}_{xx}\left(\mathbf{x}_{t+1} \otimes \mathbf{x}_{t+1}\right) + \boldsymbol{\Sigma}^{1/2}\mathbf{w}_{t+1} \\ \mathbf{x}_{t+1} &= \frac{1}{2}\mathbf{h}_{\sigma\sigma} + \mathbf{H}_x\left(\mathbf{x}_t\right) + \frac{1}{2}\mathbf{H}_{xx}\left(\mathbf{x}_t \otimes \mathbf{x}_t\right) + \sigma\mathbf{J}\mathbf{v}_{t+1}\end{aligned} \tag{33}$$

At each t, we compute the value of the state vector expected at $t+1$

$$\overline{\mathbf{x}}_{t+1|t} \approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{2}\mathbf{h}_{\sigma\sigma} + \mathbf{H}_x\left(\mathbf{x}^{(i)}_t\right) + \frac{1}{2}\mathbf{H}_{xx}\left(\mathbf{x}^{(i)}_t \otimes \mathbf{x}^{(i)}_t\right)\right] \tag{34}$$

and linearise the measurement equation around this value as

$$\mathbf{y}_{t+1} = \mathbf{y}_{t+1|t} + \mathbf{w}_{t+1|t} \tag{35}$$

where

$$\mathbf{y}_{t+1|t} = \frac{1}{2}\mathbf{g}_{\sigma\sigma} + \left[\mathbf{G}_x + \frac{1}{2}\mathbf{G}_{xx}\overline{\mathbf{D}}_k\right]\mathbf{x}_{t+1|t} + \frac{1}{2}\mathbf{G}_{xx}\left[\left(\overline{\mathbf{x}_{t+1|t}} \otimes \overline{\mathbf{x}_{t+1|t}}\right) - \overline{\mathbf{D}}_k\overline{\mathbf{x}_{t+1|t}}\right] \tag{36}$$

$$\mathbf{w}_{t+1|t} = \sigma\overline{\mathbf{G}}_x\mathbf{J}\mathbf{v}_{t+1} + \boldsymbol{\Sigma}^{1/2}\mathbf{w}_{t+1} \backsim \mathbf{N}(\mathbf{0}, \overline{\mathbf{G}}_x\boldsymbol{\Omega}\overline{\mathbf{G}}'_x + \boldsymbol{\Sigma}), \boldsymbol{\Omega} = \boldsymbol{\sigma}^2\mathbf{J}\mathbf{J}' \tag{37}$$

$$\overline{\mathbf{G}}_x = \left[\mathbf{G}_x + \frac{1}{2}\mathbf{G}_{xx}\overline{\mathbf{D}}_k\right], \tag{38}$$

$$\overline{\mathbf{D}}_k = \left[\frac{\partial\left(\mathbf{x}_{t+1} \otimes \mathbf{x}_{t+1}\right)}{\partial\mathbf{x}_{t+1}}\right]_{\mathbf{x}_{t+1}=\overline{\mathbf{x}}_{t+1|t}} = \left[\left(\mathbf{I}_{n_x} \otimes \mathbf{x}_{t+1}\right) + \left(\mathbf{x}_{t+1} \otimes \mathbf{I}_{n_x}\right)\right]_{\mathbf{x}_{t+1}=\overline{\mathbf{x}}_{t+1|t}} \tag{39}$$

Hence, the joint distribution of $\mathbf{x}_{t+1}$ and $\mathbf{y}^o_{t+1}$ conditioned on $\mathbf{x}_t$ is Gaussian and we can use the standard multivariate Normal formulae for conditional moments to write

$$\left(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}^o_{t+1}, \boldsymbol{\theta}\right) \backsim N\left[E(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}^o_{t+1}), V(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{t+1})\right] \tag{40}$$

$$\mathrm{E}(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}^o_{t+1}, \boldsymbol{\theta}) = \mathbf{x}_{t+1|t} + \boldsymbol{\Omega}\overline{\mathbf{G}}'_x\left[\overline{\mathbf{G}}_x\boldsymbol{\Omega}\overline{\mathbf{G}}'_x + \boldsymbol{\Sigma}\right]^{-1}\left(\mathbf{y}^o_{t+1} - \mathbf{y}^o_{t+1|t}\right) \tag{41}$$

$$\mathrm{V}(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}^o_{t+1}, \boldsymbol{\theta}) = \left\{\boldsymbol{\Omega} - \boldsymbol{\Omega}\overline{\mathbf{G}}'_x\left[\overline{\mathbf{G}}_x\boldsymbol{\Omega}\overline{\mathbf{G}}'_x + \boldsymbol{\Sigma}\right]^{-1}\overline{\mathbf{G}}_x\boldsymbol{\Omega}\right\}$$

These are approximate results but, using the importance sampling principle, it is possible to correct for the approximation induced by the linearization when constructing the likelihood, by assigning weights

$$w(\mathbf{x}_{t+1}^{(i)}) = \frac{p(\mathbf{x}_{t+1}^{(i)}|\mathbf{x}_t^{(i)}, \boldsymbol{\theta}) \times p(\mathbf{y}_{t+1}^o|\mathbf{x}_t^{(i)}, \boldsymbol{\theta})}{\widetilde{p}(\mathbf{x}_{t+1}^{(i)}|\mathbf{y}_{t+1}^o \mathbf{x}_t^{(i)}, \boldsymbol{\theta})} \tag{42}$$

In this paper, however, we neglect the approximation error and directly assign the draws weights equal to

$$\widetilde{w}(\mathbf{x}_{t+1}^{(i)}) = \widetilde{p}(\mathbf{y}_{t+1}^o|\mathbf{x}_t^{(i)}, \boldsymbol{\theta}) \tag{43}$$

In Amisano and Tristani (2008a) we document that the two procedures yield very similar results in this type of model.

Note that the linearised conditional particle filter algorithm that we use is connected with the *extended Kalman filter* (see Arulampalam et al., 2002, section IV). The difference is that both the measurement and the state equations are linearised at any point in time in the extended Kalman filter.

It is important to emphasise that the relative performance of the conditional particle filter with respect to the simple particle filter (or other sequential Monte Carlo filtering algorithms) cannot be established in a general way, and have to be assessed on a case by case basis: there are contexts in which the conditional particle filter fares worse than other alternatives. We investigate this issue in details in Amisano and Tristani (2008a).

4.2 Inference on the parameters of the model

Once the likelihood has been obtained, it can be used either in a maximum likelihood estimation framework or in a Bayesian posterior simulation algorithm.

In this paper we use a random walk Metropolis Hastings algorithm (see Chib, 2001, An and Schorfheide, 2007) In other words, we use sequential Monte Carlo methods to compute the likelihood of the model and we plug this likelihood in a MCMC framework. As is customary, we chose the candidate density to be a multivariate Gaussian distribution centered on the previous draw and with covariance matrix proportional to the empirical sample covariance matrix obtained from preliminary long simulations from the linearised model. This procedure is appealing because, unlike alternative approaches to the choice of the covariance matrix of the candidate distribution, it does not require any

log-posterior maximization. The tuning parameter on the covariance matrix was calibrated to achieve acceptance rates between 20% and 40%. In order to avoid numerical problems, we transformed the parameters in order to get rid of the constraints on their domain.

Some details are important in the practical implementation of the filtering procedure used to obtain the likelihood of the model.

The first issue is how to initialise the filter. We draw N realisations from the initial state vector ($\mathbf{x}_0$) from a multivariate Gaussian distribution with first two moments given by the ergodic mean and covariance matrix implied by the quadratic approximation. [6]

Another important aspect is the choice of M, the number of MCMC replications and N, i.e. the number of particles to be drawn at each observations. We run several chains for each model (M1 and M2) and for each approximation (linear and quadratic). At the outset we run $M = 55,000$ MCMC replications, the first 5000 of which are discarded, and in the nonlinear models $N = 10,000$ particles were used to construct the likelihood. For M1, the model that seems clearly preferred by the data, we rerun estimation with a higher number of MCMC replications ($M = 275000$, discarding the first $27,500$) and with a higher number of particles ($N = 20,000$).

We use an efficient filtering technique, the conditional particle filter, and we choose $N = 20,000$ particles in order to keep computation time at manageable levels. We checked the precision achieved in evaluating the sample likelihood by comparing values obtained using higher numbers of particles (50000). In Figure (1) we report the empirical distribution of the relative absolute differences in log likelihood evaluation for each draw the parameter vector taken from the quadratic posterior distribution of M1:

$$\frac{|p_{N_1}(\underline{\mathbf{y}}^o|\boldsymbol{\theta}^{(i)}) - p_{N_2}(\underline{\mathbf{y}}^o|\boldsymbol{\theta}^{(i)})|}{|p_{N_2}(\underline{\mathbf{y}}^o|\boldsymbol{\theta}^{(i)})|} \tag{44}$$

$$_{N_1} = 20000, N_2 = 50000 \tag{45}$$

$$\boldsymbol{\theta}^{(i)} \backsim p(\boldsymbol{\theta}^{(i)}|\underline{\mathbf{y}}^o), \tag{46}$$

It is easy to see that for most of parameter values, using 20000 particles instead of 50000 generates percentage errors smaller than 0.02%. This seems

[6] In a previous version of the paper, we initialized the filter by using a distribution centered at zero. This had the effect of unduly penalising the quadratic model in a marginal likelihood based model comparison setting against the linear model. The posterior distribution of the parameters seem instead not greatly influenced by filter initialisation. We thank an anonymous referee to have suggested this to us.

acceptable for practical estimation.

In order to test the general stability of our results, we run several different simulation rounds. Estimates are quite stable for the M1 model, but a bit less so for the M2 model, for which across different simulation, one observes quite some variability in the results.

4.3 Prior elicitation

One of the hardest parts in implementing Bayesian techniques is how to specify sensible priors. There are parameters for which this task is less difficult, and these are parameters such as those describing preferences or technology, for which there are well grounded beliefs which can be cast in probabilistic terms to form priors. For some others (typically the second order parameters, i.e. the standard errors of shocks) this task is more difficult. For most of the macro parameters in the first group, we have adopted priors consistent with those of Smets and Wouters (2003), while for parameters associated to second order moments, which play a more determinant role for the second order approximation, we have resorted to prior predictive analysis (see Geweke, 2005, section 8.3.1): we draw parameter values from the joint prior, we solve the model and we compute the moments of the stationary distribution of the data. We obtain in this way a prior distribution of these model-based features. We calibrated the prior hyperparameters in order to have a prior distribution of the first and second moments of the model-based ergodic distribution centered around resonable values, i.e. of the same order of magnitude of the unconditional sample data moments. We have experienced that a bit of thought in the specification of the prior usually helps in eliminating some of the numerical problems encountered by the sequential Monte Carlo filtering procedures.

We decided to dogmatically set measurement standard errors equal to 10^{-6} to concentrate on the role of the four different structural shocks.

The prior used in estimation are described in Table (2). We decided to take into considerations constraints on the parameter domain by aptly specifying prior distributions which automatically satisfy these constraints: non negative parameters were given a Gamma prior, parameters constrained on the unit simplex were given a Beta prior, and parameter which cannot be smaller than 1 were given a Gamma distribution for their difference with respect to one. The standard errors of the shocks were also assigned Gamma distributions. [7]

[7] The inverse Gamma distribution is a more customary choice for standard errors, as it generates conjugate priors in particular models. Since we compute posterior distributions by simulation in any case, there is no reason for us to use an inverse Gamma distribution.

5 Results

All our results are based on output, nominal interest rate and inflation data taken from the Area Wide Model database (see Fagan, Henry and Mestre, 2005). Following Smets and Wouters (2003), we remove a deterministic trend from the GDP series prior to estimation. No transformations are applied to inflation and interest rate data. The estimation period runs from 1970Q1 to 2004Q4. The data are shown in Figure 2. Note that Model M2 uses the changes of inflation and the difference between interest rate and inflation in order to eliminate the unit root behaviour induced by the random walk hypothesis on the inflation target process.

We highlight four main features of our results. First, we briefly discuss our parameter estimates, focusing on differences across models/specifications and compared to the existing literature. Secondly, we compare the estimates based on the two specifications M1 and M2 and show that the first model is overwhelmingly preferred by the data. This conclusion is also informally supported by the fact that the 95% Highest Posterior Density credible sets (henceforth HPD sets, see Geweke, 2005, Section 2.5) constructed using the marginal posterior distribution of ρ_π and ι (see Table 3) in Model M1 do not contain the unit value. We therefore focus on M1 for the rest of our analysis. Next, we compare the linear and nonlinear specifications for model M1 and conclude that the nonlinear version is superior to the linear one. Finally, we discuss the implications of the nonlinear M1 model for the dynamics of inflation, in particular looking at the way in which initial conditions affect the magnitude and the persistence of the effects of shocks in a nonlinear world.

5.1 Parameter estimates

Tables 3 and 4 present the results of the estimation of first and second order versions of M1 and M2. The evidence can be summarised as follows.

- In both models and both specifications (linear and quadratic), posterior distributions tend to have a mean which is far from the prior mean. As an example, in M2 the RRA parameter has a posterior distribution that hovers around 4.0 while the prior distribution is centered around 2.0. Not all parameters, however, have marginal posterior distributions which are tighter than the corresponding priors. See for example γ (RRA), ϕ (labor disutility), and the parameters describing the properties of the tax shock (ρ_τ, σ_τ and τ).
- The posterior means of the deep parameters are mostly stable across the different specification. See for example β (discount factor), γ (RRA), h

(habit persistence), ρ_i (interest rate smoothing).

- For both models, linear and nonlinear specifications tend to produce similar parameter estimates, with some important exceptions: γ and θ are estimated quite differently. As an example, the elasticity of substitution across goods, θ, is higher in the quadratic model than in the linear model. In M1 the posterior mean of γ based on the quadratic model is 4.2, while it is only 3.3 in the linear case. This parameter is crucial for determining precautionary savings and the quadratic approximation allows us to obtain estimates which are different from their linear counterparts.
- In general it seems that the quadratic estimation procedure is capable of generating somehow sharper estimates. Looking just at the univariate marginal posterior of model M1, 4 parameters out of 19 have posterior HPD sets based on the quadratic approximation which are narrower than their counterparts based on linear estimation. For these parameters, the quadratic approximation posterior standard errors are smaller than their counterparts based on the linear model. These parameters are $\phi, \theta, \iota, \psi_\pi$ and ρ_τ. Similar considerations attain to the estimation of model M2: tighter posterior distributions are obtained for 9 out of 16 parameters. In synthesis, the documented benefits in using a higher order solution to estimate the parameters (see Fernandez-Villaverde and Rubio-Ramirez, 2006a, b) are confirmed in our evidence even if not for all parameters.
- For model M1, we observe that mean posteriors are consistent with a very reasonable degree of price stickiness, implying average price durations of 1.8 quarters. Our estimates of the habit formation parameter h and of the parameters of the policy rule are also broadly in line with other existing results, notably those in Smets and Wouters (2003). The main, important exception concerns the inflation indexation parameter. Irrespective of the specification (linear or nonlinear), our estimates ($\iota \simeq 0.1$) are particularly small and imply a very minor degree of inflation persistence. This result is quite surprising in view of the high serial correlation of actual inflation, and also if compared to findings of existing studies. Nevertheless, this is an appealing property of our results given that the assumption of indexation is an ad-hoc feature of the model.
- The tax shock plays only a limited role in the model (see the forecast error variance decomposition in Table 6): at impact tax shocks explain 7%, 3% and 2% of inflation, interest rate and output variability.
- In terms of overall fit both models do quite well. Figure 3 shows the observed series used for estimation and the posterior mean of their filtered one step ahead forecasts for model M1. This seems to indicate a good fit of the data, and similar evidence holds for model M2.
- in order to check the reliability of the estimated models, we have also looked at the posterior distribution of the latent variables implied by the estimated model. As an example, Figure 4 reports the smoothed posterior mean of the inflation target for M1, showing that its range of values and its dynamics are not unreasonable. The target is higher during the seventies and early

eighties, but not as high as actual inflation, then declines to values around 2% (annualised) during the EMU period.

5.2 Model comparison

To test whether the assumption of permanent shifts in the average inflation rate is borne in the data, we compare formally models M1 and M2.

It is useful to note that model M2 is almost nested in M1. With the exception of the intercept term in the Taylor rule, it amounts to fixing two parameters in model M1: $\iota = 1$ (the inflation indexation parameter) and $\rho_\pi = 1$ (the persistence of the inflation target). The second restriction is unlikely to have a strong impact on the marginal likelihood, given that it is in any case estimated to be very close to 1 in model M1 (even if its HPD set does not contain 1). Given the estimates of ι in M1, however, the first restriction is likely to be more binding.

In the literature, model comparison exercises are often based on the marginal likelihood. We also follow this approach here, even if it must be kept in mind that marginal likelihoods are subject to a number of caveats (see for instance Gelman et al., 2004, section 6.7, Del Negro and Schorfheide, 2008, Sims, 2003). An alternative model-evaluation criterion that has been proposed in the literature is to compare the predictive densities implied by the competing models with a recursive estimation approach (see Geweke, 2005, section 2.6.2). We therefore apply a variant of this approach to shed further light on the comparison between models that are similar in terms of marginal likelihood.

The marginal likelihood (ML) of each model $\mathcal{M}_j$ is defined as

$$\ln(p(\mathbf{y}|\mathcal{M}_j) = \ln\left(\int p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{M}_j) p(\boldsymbol{\theta}|\mathbf{y}, \mathcal{M}_j) d\boldsymbol{\theta}\right) \tag{47}$$

The difference between these two quantities for models $\mathcal{M}_j$ and $\mathcal{M}_i$ gives the log Bayes Factor of one model versus the other. Computed values largely different from zero suggest dominance of one model vs the other. The MLs are computed based on the modified Gelfand and Dey approach described in Geweke (1999). This method is very accurate when the posterior PDF is unimodal. Here we use the MLs to compare models M1 and M2 and also the linear versus the quadratic specifications.

There is an issue of detail which has to be emphasised when comparing marginal likelihoods across our two different models. The two models do not use the same observable variables, since in M2 data on inflation and interest rate are transformed to achieve stationarity (we use differenced inflation

and the real interest rate). Nevertheless, conditional on past information, the Jacobian determinant of the transformation from the variables in M1 to the ones in M2 is unity. This can be shown if we define $\mathbf{y}_t^{(M_1)} = [\pi_t, r_t, y_t]'$ and $\mathbf{y}_t^{(M_2)} = [\Delta\pi_t, r_t - \pi_t, y_t]'$. Then $\mathbf{y}_t^{(M_2)}$ can be rewritten as

$$\mathbf{y}_t^{(M_2)} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{y}_t^{(M_1)} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1}^{(M_1)}$$

so that

$$\left| \frac{\partial \mathbf{y}_t^{(M_2)}}{\partial \mathbf{y}_t^{(M_1)'}} | \mathbf{y}_{t-1}^{(M_1)} \right| = \left| \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right| = 1$$

It follows that marginal likelihoods are directly comparable if we condition on $\mathbf{y}_0$. [8]

Looking at Table 5, it seems that model M1, in either specification (linear or quadratic) is superior to model M2. Given that the log Bayes factor is very large (around 27 points comparing linear versions and 36 comparing quadratic versions) and that in the estimation of M1 the posterior 95% HPD sets for ρ_π and ι do not contain respectively the unit and zero values, we conclude that in this application model M1 is clearly preferred to M2. We thus focus on model M1 for the rest of the paper.

In terms of the euro area inflation process, the aforementioned result seems to imply that that process is best characterised by a constant mean, even if considerable and persistent deviations from the mean have occurred over the years. This is however only a tentative conclusion, given that both the M1 and the M2 models can only capture the cross-covariances of the data to a limited extent.

5.3 Linear vs nonlinear

Conclusions on the superiority between the linear and quadratic versions of model M1 are not so stark as the results from comparing M1 to M2, but

[8] A side issue concerns the sample size. Model M1 is estimated on a sample size that includes 1970:1 whereas M2 starts from 1970:2. To make the comparison completely fair, we should re-estimate M1 excluding the first observation. Given the strongly superior performance of M1 with respect to M2, this is unlikely to make a substantive difference.

equally conceptually interesting. Table 5 shows that the ML of these two specifications are very close and the quadratic model prevails marginally (with a ML around 1715, compared to 1714 for the linear case).

It is important to note that, given the specification of the models being compared, the quadratic representation is going to be anyway implicitly penalised, since the model does not include capital. Inclusion of capital would give more importance to precautionary savings decisions and therefore to the quadratic approximation in which precautionary savings are allowed to have a role. [9]

In order to have further elements for the comparison between the linear and the quadratic representations, we also compute conditional predictive densities. To compute conditional predictive densities, the model has to be reestimated at each point in time. In other words we need to compute

$$p(\mathbf{y}_{t+1}^o, \mathbf{y}_{t+2}^o ..., \mathbf{y}_T^o | \underline{\mathbf{y}_t^o}, \mathcal{M}_j) = \prod_{\tau=t}^{T} p(\mathbf{y}_{\tau+1}^o | \underline{\mathbf{y}_\tau^o}, \mathcal{M}_j) \tag{48}$$

for some date $0 \leq t \leq T-1$. Doing this by brute force, i.e. re-estimating the model at each point in time is computationally infeasible for the quadratic model. A much faster alternative is to use the output of the full sample MCMC estimation to construct an estimate based on the harmonic mean of the relevant conditional densities as follows:

$$\frac{1}{M} \sum_{i=1}^{M} \frac{1}{p(\mathbf{y}_{T_0+1}^o, \mathbf{y}_{T_0+2}^o ..., \mathbf{y}_T^o | \underline{\mathbf{y}_{T_0}^o}, \boldsymbol{\theta}^{(i)}, \mathcal{M}_j)} \tag{49}$$

$$\xrightarrow{p} \int \frac{p(\boldsymbol{\theta} | \underline{\mathbf{y}_T^o}, \mathcal{M}_j)}{p(\mathbf{y}_{T_0+1}^o, \mathbf{y}_{T_0+2}^o ..., \mathbf{y}_T^o | \underline{\mathbf{y}_{T_0}^o}, \boldsymbol{\theta}, \mathcal{M}_j)} d\boldsymbol{\theta}$$

$$= \int \frac{p(\boldsymbol{\theta} | \underline{\mathbf{y}_{T_0}^o}, \mathcal{M}_j)}{p(\mathbf{y}_{T_0+1}^o, \mathbf{y}_{T_0+2}^o ..., \mathbf{y}_T^o | \underline{\mathbf{y}_{T_0}^o}, \mathcal{M}_j)} d\boldsymbol{\theta} = \frac{1}{p(\mathbf{y}_{T_0+1}^o, \mathbf{y}_{T_0+2}^o ..., \mathbf{y}_T^o | \underline{\mathbf{y}_{T_0}^o}, \mathcal{M}_j)} \tag{50}$$

This approach is computationally not demanding (it uses the output of the posterior simulation of the model based on the whole sample size), but is has the drawback that the accuracy of the approximation in the simulation based integral (49) can be affected by large errors for a finite number of simulations M .

The results of using this approach are presented in Figure 5, where we plot

$$\ln \left(\frac{p(\mathbf{y}_{t+1}^o, \mathbf{y}_{t+2}^o ..., \mathbf{y}_T^o | \underline{\mathbf{y}_t^o}, \mathcal{M}_{linear})}{p(\mathbf{y}_{t+1}^o, \mathbf{y}_{t+2}^o ..., \mathbf{y}_T^o | \underline{\mathbf{y}_t^o}, \mathcal{M}_{quadratic})} \right), t = 0, 1, ...T-1 \tag{51}$$

[9] We thank a referee for pointing out this to us.

Each point in the figure shows the log predictive density from that point until the end of the sample.

From this comparison we see that the quadratic model seems to be superior to the linear one for most of the possible partitions of the sample. Consistently with the results of the marginal likelihood comparison, the conditional predictive distribution ratio is favourable to the quadratic specification over the full sample (the first observation in the figure). The comparison would then continue to favour the quadratic model for most of the sample. Only at the very end, namely as of the beginning of EMU, the conditional predictive distribution ratio becomes favourable to the linear specification. This suggests that nonlinearities are important in case of very large and persistent shocks, but tend to be less relevant during periods of moderate fluctuations.

On the basis of these results, we conclude that the quadratic version of M1 is superior to the linear one for most of the sample, and especially over the years where inflation is more distant from its steady state value.

These results are broadly consistent with those obtained on the basis of simulated data. Recent literature has emphasised that estimates of the second order model tend to be more precise – e.g. An and Schorfheide (2007). Canova and Sala (2006) has emphasised the chronic under-identification of many DSGE models. It is possible to verify that resorting to higher order approximation induces sensibly more curvature in the likelihood function hence increases identifiability of the parameters. We have verified this feature also for the models that we estimate in this paper and it does generally hold on simulated data (see Amisano and Tristani, 2008a).

Nevertheless, the quadratic model is not clearly and always superior to the linear specification. This result suggests the existence of a trade-off between parameter identification and mis-specification in nonlinear DSGE models, similar to the one encountered when increasing the information set in the estimation of linearised models. More information increases the ability of the researcher to pin down various parameters, but it tends to highlight any weaknesses of the model at the same time. Similarly, estimating a nonlinear model amounts to extracting more testable implications from the theory, hence achieving more efficient, or even less biased, parameter estimates when the model is approximately correct. If the model is only a rough approximation of reality, however, its nonlinear implications are likely to make it more at odds with the data (compared to its linearised counterpart). The finding of more spread-out posterior parameter distributions may be a signal of the latter phenomenon.

5.4 Euro area inflation dynamics

In this section, we discuss the dynamic implications of the model focusing in particular on the persistence and the amplitude of the responses of inflation to shocks. All the discussion is based on the posterior simulation of model M1 in its quadratic version.

First of all, in order to understand the relative importance of the different shocks hitting the system (technology, target, tax and policy shocks), we look at the forecast error variance decomposition (FEVD) coefficients which are reported in Table 6 and graphed in Figure 6. Four main features immediately stand out:

(1) the tax shock has declining importance (over the horizon) for inflation and output but increasing importance for interest rate.
(2) Technology explains the bulk of output variability also at short horizons: the posterior mean of the 1 step ahead FEVD coefficient for output is 88%.
(3) Inflation is mainly driven by target shocks and policy shocks: respectively 76% and 9% at 1-step ahead, 83% and 6% after one year and 91% and 3% after 3 years.
(4) The policy rate is moved by the same shocks as inflation but with reversed relative importance: in the short run it is nearly all (94%) policy shocks and much less the target shocks (7%). As the horizon increases the target shocks become more and more important: at 3 years target shocks account for 65% and policy shocks 8%.

Next, we turn to the impulse response functions and test whether responses starting from a high-inflation level are significantly more persistent than those starting from a low level. At the same time, we analyse whether the amplitude of the response of inflation after a shock of a given size varies depending on the starting value.

The dependence of nonlinear impulse response functions on initial conditions is well-known (see e.g. Gallant, Rossi and Tauchen, 1993). Our aim, however, is exactly to point out the extent to which economic dynamics are different over time, depending on cyclical conditions. We therefore study standard nonlinear impulse response functions, defined as the difference between the expected future sample path of a variable conditional on the state $\mathbf{x}_t$, and the expected future path conditional on $\mathbf{x}'_t$, where $\mathbf{x}_t$ is equal to $\mathbf{x}'_t$ except for an individual element which is perturbed by a known amount. [10]

[10] This is also the definition used in Gallant, Rossi and Tauchen (1993). See also Koop, Pesaran and Potter (1996).

Rather than selecting arbitrarily various initial configurations of the state vector, we focus on its two realisations estimated at the extreme values of inflation observed in our sample. Looking at Table 3, the maximum and minimum of inflation are equal to 16.58% in 1976:01 and 0.59% in 1998:03, respectively.

It is important to keep in mind that we aim at describing the dynamic properties of the system starting from initial conditions which are indeed very far from the state state. This is a hard task when using perturbation methods whose accuracy quickly deteriorates when we move away from the point around which the approximation is taken. On the other hand, we cannot use global approximation methods which are not fast enough to allow us to be used for estimation of the parameters, and we use a quadratic approximation which improves on the linear one.

In order to highlight how the impulse responses vary over time, we calculate them starting from the filtered state vectors on these dates, referred to as $t_{\max}$ (high inflation) and $t_{\min}$ (low inflation). IRFs are then computed using the KKSS simulation strategy illustrated in Section 3.1, and by integrating out future values of shocks. However, different ways to compute IRFs, namely by projecting the quadratic laws of motions and/or by setting to zero all shocks but one, led to virtually the same results. Posterior median responses and the bounds corresponding to a 95% posterior coverage are reported in Figures 7 and 9.

The response of inflation to a technology shock (row 1, column 1 of Figure 7), for which the posterior mean of the standard error is equal to 1.49%, follows the broad pattern typically observed in linearised models, if the shock occurs when inflation is low (the "Low inflation" line). Inflation falls for a few quarters and returns to baseline thereafter. The initial fall is also statistically significant at the 95% level for 2 quarters. The nonlinear effects triggered by the technology shock are quantitatively modest. Starting from a high inflation level (the "High inflation" line), the fall in inflation is reduced and ceases to be significant after 2 periods, but the differences are not statistically significant.

The response of inflation to a positive inflation target shocks (posterior mean of $\sigma_\pi = 0.17\%$) is more markedly dependent on the starting point (see first row, second column of Figure 7). The low estimated value of the inflation indexation parameter, ι, implies that inflation is highly forward-looking. As a result, a positive and highly persistent increase in the inflation target has immediate consequences on current inflation, which rises more than the target itself (around .35%). In turn, this implies that the policy interest rate increases on impact to counter the rising inflationary pressure. Nevertheless, the shock continues to produce expansionary effects on output, if it occurs when inflation is low.

This result, however, changes dramatically if the shock takes place when inflation is already high. In this case, there is a much bigger upward increase in inflation, nearly twice as big as in the previous case, and also more persistent in terms of median half-life. As a result, the policy tightening must be much more severe, so as to progressively *contract* aggregate demand. Hence, the impulse response of output to an inflation target shock changes sign depending on the state of the world prevailing when the shock occurs.

The response of inflation to a monetary policy shock (posterior mean of $\sigma_i = 0.18\%$) is again more marked when inflation at the starting point is high. For given size of the shock, the top row, second column of Figure 9 shows that it falls on impact by around 0.15 percent in this case, compared to a fall of 0.10 after a low-inflation starting point. These differences are nevertheless short-lived and they disappear completely after 5-6 quarters.

In synthesis, the two main shocks driving inflation (target and policy) have quantitatively different impact and persistence behaviours depending on the initial conditions. Our results are important, for example because they suggest that sacrifice-ratios derived from a linearised model may provide a misleading picture. In the case we analyze, the benefits on expectations of cutting a high inflation target are so large, that the cut would have an expansionary effect. This is not the conclusion that one would reach focusing solely on the linearised model.

6 Conclusions

We have presented the results of an empirical analysis of the nonlinear features of a relatively standard, small DSGE model. With the limitations posed by the simplicity of the model, a few main results emerge.

First, the nonlinear macroeconomic dynamics intrinsic in the model can have pronounced and statistically significant effects in case of moderately large movements in the inflation rate. The amplitude and persistence of the responses of inflation to shocks differ at different points in the sample. For example, a given surprise increase in the inflation target produces stronger inflationary consequences if it occurs in a high inflation environment, compared to an environment where price stability is maintained. Even starker differences can be observed for the response of output, which can change sign depending on initial conditions.

When comparing formally linear and nonlinear models, we tend to conclude slightly in favour of the latter specification. We show that this result has an intuitive interpretation in terms of better performance when observed variables

are furthest away from their steady state levels.

From a more general viewpoint, our results illustrate some of the promises of exploring estimated version of nonlinear DSGE models, including the possibility to increase the identifiability of parameters.

Nevertheless, we wish to end with a word of caution, since the estimation on nonlinear models does have drawbacks. The first one, induced by the need of resorting to simulation filtering to carry out likelihood based inference, is that much more computation time is required. A second drawback is that sequential Monte Carlo methods are sensitive to outliers and degeneracies which can arise in actual data. Nevertheless, the conditional particle filter has proven to be a robust tool in our application.

Acknowledgement

The opinions expressed are personal and should not be attributed to the European Central Bank. We acknowledge very useful suggestions from the editor and two anonymous referees on a previous version of the paper. We wish to thank Sumru Altug, Sungbae An, Roberto Casarin, Michael Dueker, Jesus Fernandez-Villaverde, John Landon-Lane, Gianni Lombardo, Juan Francisco Rubio-Ramirez, Ingvar Strid, Ken Wallis and Ida Wolden Bache for useful discussions and comments. We thank also seminars partecipants at the following institutions/conferences: Universities of Brescia, Palermo, Warwick, Zurich, EABCN/CEPR 2006 workshop, Central Bank Workshop on Macroeconomic Modelling 2006, Society for Computational Economics Conference on Computing in Economics and Finance 2006.

A The complete model(s)

The models are composed of the following equations

$$\begin{aligned}
\Upsilon_{2,t} &= \frac{\alpha\left(\theta-1\right)}{\phi\chi\theta}\left(\frac{1-\zeta\left(\frac{\Pi_{t-1}}{\Pi_t}\right)^{1-\theta}}{\left(1-\zeta\right)}\right)^{1+\frac{\theta}{1-\theta}\frac{\phi}{\alpha}}\Upsilon_{1,t}\\
\Upsilon_{2,t} &= \left(1+\tau_t\right)\frac{A_t^{-\frac{\phi}{\alpha}}}{\lambda_t}Y_t^{\frac{\phi}{\alpha}}+\mathrm{E}_t\zeta Q_{t,t+1}\Upsilon_{2,t+1}\Pi_t^{-\theta\frac{\phi}{\alpha}}\Pi_{t+1}^{1+\theta\frac{\phi}{\alpha}}\\
\Upsilon_{1,t} &= Y_t+\mathrm{E}_t\zeta Q_{t,t+1}\Upsilon_{1,t+1}\Pi_t^{(1-\theta)}\Pi_{t+1}^{\theta}\\
\left(Y_t^n\right)^{\frac{\phi-\alpha}{\alpha}} &= \frac{\alpha}{\phi\chi\mu\left(1+\tau_t\right)}A_t^{\frac{\phi}{\alpha}}\left(\left(Y_t^n-hY_{t-1}^n\right)^{-\gamma}-\beta h\mathrm{E}_t\left[\left(Y_{t+1}^n-hY_t^n\right)^{-\gamma}\right]\right)\\
\frac{1}{R_t} &= E_t\left[\beta\left(\frac{Y_{t+1}^n}{Y_t^n}\right)^{\frac{\phi-\alpha}{\alpha}}\left(\frac{A_{t+1}}{A_t}\right)^{-\frac{\phi}{\alpha}}\frac{1+\tau_{t+1}}{1+\tau_t}\right]\\
\lambda_t &= \left(Y_t-hY_{t-1}\right)^{-\gamma}-\beta h\mathrm{E}_t\left[\left(Y_{t+1}-hY_t\right)^{-\gamma}\right]\\
Q_{t,t+1} &= \beta\frac{1}{\Pi_{t+1}}\frac{\lambda_{t+1}}{\lambda_t}\\
\frac{1}{I_t} &= E_t\left(Q_{t,t+1}\right)\\
A_{t+1} &= \overline{A}^{1-\rho}A_t^{\rho}e^{v_{t+1}}\\
\tau_t &= \left(1-\rho_\tau\right)\overline{\tau}+\rho_\tau\tau_{t-1}+v_t^{\tau}
\end{aligned}$$

plus either of the policy rules (12)-(13) or (14)-(15).

In the case of M1 the solution is standard. For M2, we first remove the stochastic trend from nominal variables. More precisely, we define the detrended variables

$$\begin{aligned}
\widetilde{\Pi}_t^* &\equiv \frac{\Pi_t^*}{\Pi_t}\\
\widetilde{I}_t &\equiv \frac{I_t}{\Pi_t}\\
\widetilde{Q}_{t,t+1} &\equiv Q_{t,t+1}\Pi_{t+1}
\end{aligned}$$

and rewrite the system as

$$\Upsilon_{2,t} = \frac{\alpha(\theta-1)}{\phi\chi\theta}\left(\frac{1-\zeta\left(\frac{1}{\Delta\Pi_t}\right)^{1-\theta}}{(1-\zeta)}\right)^{1+\frac{\theta}{1-\theta}\frac{\phi}{\alpha}}\Upsilon_{1,t}$$

$$\Upsilon_{2,t} = (1+\tau_t)\frac{A_t^{-\frac{\phi}{\alpha}}}{\lambda_t}Y_t^{\frac{\phi}{\alpha}} + \mathrm{E}_t\zeta\widetilde{Q}_{t,t+1}\Upsilon_{2,t+1}\left(\Delta\Pi_{t+1}\right)^{\theta\frac{\phi}{\alpha}}$$

$$\Upsilon_{1,t} = Y_t + \mathrm{E}_t\widetilde{Q}_{t,t+1}\Upsilon_{1,t+1}\left(\Delta\Pi_{t+1}\right)^{\theta-1}$$

$$(Y_t^n)^{\frac{\phi-\alpha}{\alpha}} = \frac{\alpha}{\phi\chi\mu(1+\tau_t)}A_t^{\frac{\phi}{\alpha}}\left(\left(Y_t^n - hY_{t-1}^n\right)^{-\gamma} - \beta h\mathrm{E}_t\left[\left(Y_{t+1}^n - hY_t^n\right)^{-\gamma}\right]\right)$$

$$\frac{1}{R_t} = E_t\left[\beta\left(\frac{Y_{t+1}^n}{Y_t^n}\right)^{\frac{\phi-\alpha}{\alpha}}\left(\frac{A_{t+1}}{A_t}\right)^{-\frac{\phi}{\alpha}}\frac{1+\tau_{t+1}}{1+\tau_t}\right]$$

$$\lambda_t = (Y_t - hY_{t-1})^{-\gamma} - \beta h\mathrm{E}_t\left[(Y_{t+1} - hY_t)^{-\gamma}\right]$$

$$\widetilde{Q}_{t,t+1} = \beta\frac{\lambda_{t+1}}{\lambda_t}$$

$$\frac{1}{\widetilde{I}_t} = E_t\left(\widetilde{Q}_{t,t+1}\frac{1}{\Delta\Pi_{t+1}}\right)$$

$$\widetilde{I}_t = \left(\frac{1}{\beta}\left(\widetilde{\Pi}_t^*\right)^{1-\psi_\Pi}\left(\frac{Y_t}{Y_t^n}\right)^{\psi_Y}\right)^{1-\rho_I}\frac{\widetilde{I}_{t-1}^{\rho_I}}{(\Delta\Pi_t)^{\rho_I}}e^{\eta_t}$$

$$A_{t+1} = \overline{A}^{1-\rho}A_t^\rho e^{v_{t+1}}$$

$$\widetilde{\Pi}_t^* = \widetilde{\Pi}_{t-1}^*\frac{1}{\Delta\Pi_t}e^{v_t^\pi}$$

$$\tau_t = (1-\rho_\tau)\overline{\tau} + \rho_\tau\tau_{t-1} + v_t^\tau$$

B Model solution

The approximate solution of the model is computed following Gomme and Klein (2006). First, we collect all first order conditions in a vector function F such that

$$F_t(\mathbf{x}_t, \sigma) \equiv \mathrm{E}_t f(\mathbf{y}_{t+1}, \mathbf{y}_t, \mathbf{x}_{t+1}, \mathbf{x}_t) = 0$$

where $\mathbf{x}_t$ is the vector of (natural logarithms of the) predetermined variables and $\mathbf{y}_t$ is the vector of (natural logarithms of the) non-predetermined variables. More specifically, in the case of M1 $\mathbf{x}_t = \left[\pi_{t-1}, y_{t-1}^{nat}, y_{t-1}, i_{t-1}, a_t, \pi_t^*, \tau_t, v_t^i\right]'$ and $\mathbf{y}_t = [y_t^{nat}, \Upsilon_{1,t}, \Upsilon_{2,t}, \pi_t, i_t, y_t, \lambda_t]'$, while for M2 $\mathbf{x}_t = \left[\widetilde{\pi}_{t-1}^*, y_{t-1}^{nat}, y_{t-1}, \widetilde{i}_{t-1}, a_t, v_t^\pi, \tau_t, v_t^i\right]'$ and $\mathbf{y}_t = \left[y_t^{nat}, \Upsilon_{1,t}, \Upsilon_{2,t}, \Delta\pi_t, \widetilde{i}_t, y_t, \lambda_t, \widetilde{\pi}_t^*\right]'$. In F_t, σ denotes a scalar perturbation parameter, such that the law of motion of the exogenous state variables x_t^{exog} (where $\mathbf{x}_t^{exog} = [a_t, \pi_t^*, \tau_t, v_t^i]'$) can be written as $\mathbf{x}_{t+1}^{exog} = \mathbf{H}_x^{exog}\mathbf{x}_t^{exog} + \sigma\boldsymbol{\eta}\mathbf{v}_{t+1}$, where the variance-covariance matrix of $\mathbf{v}_{t+1}$ is the identity matrix.

References

[1] Amisano, G. , Tristani O. , 2009a. Simulation based filtering for nonlinear DSGE models: problems and solutions. Mimeo.

[2] Amisano, G. , Tristani O. , 2009b. A DSGE model for the term structure with regime shifts. Mimeo.

[3] An, S., Schorfheide F., 2007. Bayesian analysis of DSGE models. Econometric Reviews 26, 2, 113-172.

[4] Angeloni, I., Aucremanne L., Ehrmann M., Galí J., Levin A., Smets F., 2005). Inflation persistence in the euro area: preliminary summary of findings. Mimeo.

[5] Arulampalam, M.S., Maskell, S., Gordon N., Clapp T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 50, 174-188.

[6] Aruoba B., Fernandez-Villaverde F., Rubio-Ramirez J.F., 2006. Comparing solution methods for dynamic equilibrium economies". Journal of Economic Dynamics and Control 30, 2477-508.

[7] Ascari, G., 2004. Staggered prices and trend inflation: some nuisances. Review of Economic Dynamics 7, 642-667.

[8] Benigno, P., Woodford M., 2005. Inflation stabilisation and welfare: the case of a distorted steady state. Journal of the European Economic Association 3, 1185-236.

[9] Bilke, L. , 2005. Break in the mean and persistence of inflation: a sectoral analysis of French CPI. ECB Working Paper No. 463.

[10] Aruoba, S. B., Fernández-Villaverde F., Rubio-Ramírez J. F., 2006. Comparing solution methods for dynamic equilibrium economies, Journal of Economic Dynamics & Control 30, 2477–2508

[11] Calvo, G. A. , 1983. Staggered prices in a utility-maximising framework. Journal of Monetary Economics 12, 983-98.

[12] Canova, F. Sala L., 2006. Back to square one: identification issues in DSGE models. ECB Working Paper No. 583.

[13] Casarin, R., Trecroci C., 2006. Business cycle and stock market volatility: a particle filter approach, University of Brescia Dept. of Economics Working Paper # 0603.

[14] Chib, S. , 2001. Markov Chain Monte Carlo methods: computation and inference, in: Heckman, J.J., Leamer E. (Eds.), Handbook of Econometrics, Vol. 5, North Holland.,

[15] Christiano, L., Eichenbaum, M., Evans, C.L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113, 1-45.

[16] Clarida, R., Galí J., Gertler, M., 1998. Monetary policy rules in practice: some international evidence. European Economic Review 42, 1033-1067.

[17] Clarida, R., Galí J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. Quarterly Journal of Economics, 147-180.

[18] Corvoisier, S., Mojon B., 2005. Breaks in the mean of inflation: how they happen and what to do with them. ECB Working Paper No. 451.

[19] Del Negro, M., Schorfheide F., 2008. Forming priors for DSGE models, and how it affects the assessment of nominal rigidities, Journal of Monetary Economic 55, 7, 1191-1208.

[20] Dossche, M. , Everaert G., 2005. Measuring inflation persistence: a structural time-series approach. ECB Working Paper No. 495.

[21] Doucet, A., de Freitas N., Gordon N.(Eds.), 2001. Sequential Monte Carlo methods in practice. Springer-Verlag, New York.

[22] Fagan, G., Henry J., Mestre R., 2005. An area-wide model (AWM) for the euro area. Economic Modelling 22, 39-59.

[23] Fernandez-Villaverde J., Rubio-Ramirez J.F., Santos M.S., 2006. Convergence properties of the likelihood of computed dynamic models. Econometrica 74, 93-119.

[24] Fernandez-Villaverde J., Rubio-Ramirez J.F., 2007a. Estimating macroeconomic models: a likelihood approach. Review of Economic Studies 74, 1059-1087.

[25] Fernandez-Villaverde J., Rubio-Ramirez J.F., 2007b. How structural are structural parameters?" 2007 NBER Macroeconomics Annual, 83-137.

[26] Gallant, A.R., Rossi P.E., Tauchen G., 1993. Nonlinear dynamic structures", Econometrica 57, 357-85.

[27] Gelman, A., Carlin J.B., Stern H.S., Rubin D.B., 2004. Bayesian data analysis, 2nd edition. Chapman & Hall, Boca Raton, Florida.

[28] Geweke, J., 1999. Using simulation methods for Bayesian econometric modeling: inference, development and communication, Econometric Reviews, 18, 1-74.

[29] Geweke, J. , 2005. Contemporary Bayesian econometrics and statistics. J. Wiley & Sons, Noboken, NJ.

[30] Gomme, P. Klein P., 2008. Second-order approximation of dynamic models without the use of tensors. Mimeo, University of Western Ontario.

[31] Hamilton, J. , 1994. Time series analysis. Princeton University Press, Princeton.

[32] Ionides, E. L., 2007. Inference and filtering for partially observed diffusion processes via sequential Monte Carlo. Mimeo, University of Michigan

[33] Kim, J., Kim, S. Schaumburg E., Sims C., 2003. Calculating and using second order accurate solutions of discrete time dynamic equilibrium models. Mimeo.

[34] Klein, P., 2000. Using the Generalized Schur Form to solve a multivariate linear rational expectations model. Journal Of Economic Dynamics and Control 24, 1405-23.

[35] Koop, G., Pesaran H., Potter S., 1996. Impulse response analysis in nonlinear multivariate models. Journal of Econometrics 74, 119-48.

[36] Levin, A., Piger, J., 2004. Is inflation persistence intrinsic in industrial economies. ECB Working Paper 334.

[37] Liu, J., West M., 2001. Combined parameter and state estimation in simulation-based filtering. in Doucet, A., de Freitas N., Gordon N.(Eds.), 2001. Sequential Monte Carlo methods in practice. Springer-Verlag, New York, pp. 197-223.

[38] Lombardo, G., Sutherland A., 2007. Computing second-order-accurate solutions for rational expectation models using linear solution methods", Journal of Economic Dynamics and Control 31, 515-30.

[39] Schmitt-Grohé S., Uribe M., 2004. Solving dynamic general equilibrium models using a second-order approximation to the policy function. Journal of Economic Dynamics and Control 28, 755-75.

[40] Schorfheide, F. , 2005. Learning and Monetary Policy Shifts. Review of Economic Dynamics 8, 392-419.

[41] Sims, C.A. , 2003. Probability models for monetary policy decisions. Mimeo, Princeton University.

[42] Smets, F., Wouters R., 2003. An estimated stochastic dynamic general equilibrium model of the euro area. Journal of European Economic Association 1, 1123-75

[43] Smets, F., Wouters R., 2005. Comparing shocks and frictions in US and euro area business cycles: a Bayesian DSGE approach. Journal of Applied Econometrics 20, 2, 161-183.

[44] Steinsson, J. , 2003. Optimal monetary policy in an economy with inflation persistence. Journal of Monetary Economics 50, 1425-56.

[45] Sutherland, A. , 2002. A simple second-order solution method for dynamic general equilibrium models. CEPR Discussion Paper No. 3554.

[46] Woodford, M. , 2003. Interest and Prices. Princeton University Press, Princeton.

[47] Yun, T., 1996. Nominal price rigidity, money supply endogeneity, and business cycles. Journal of Monetary Economics 37, 345-370

Table 1. Comparison between Particle Filter and Conditional Particle Filter

	Particle Filter	**Conditional Particle Filter**
draw $\mathbf{x}_{t+1}^{(i)}$ from	$p(\mathbf{x}_{t+1}\|\mathbf{x}_t^{(i)}, \underline{\mathbf{y}}_t^o, \boldsymbol{\theta})$	$p(\mathbf{x}_{t+1}\|\mathbf{x}_t^{(i)}, \mathbf{y}_{t+1}^o, \boldsymbol{\theta})$
compute weight $w_{t+1}^{(i)}$ as	$p(\mathbf{y}_{t+1}^o\|\mathbf{x}_{t+1}^{(i)}, \boldsymbol{\theta})$	$\mathrm{p}(\mathbf{y}_{t+1}^o\|\mathbf{x}_t^{(i)}, \boldsymbol{\theta})$
use information on $\mathbf{y}_{t+1}^o$?	in weights assigned to particles (*blind* proposal)	in drawing $\mathbf{x}_{t+1}^{(i)}$ (*adapted* proposal)

Table 2. Priors for models M1 and M2

n.	par.	description	prior type	prior mean	prior sd	prior low q	prior up q	in Model M2
1	β	intertemporal discount	Beta	0.901	0.089	0.665	0.997	same prior
2	γ	RRA	Gamma*	1.999	0.704	1.125	3.807	same prior
3	h	habit persistence	Beta	0.699	0.137	0.398	0.924	same prior
4	ϕ	disutility of labor	Gamma	4.006	1.008	2.373	6.313	same prior
5	θ	goods elasticity of substitution	Gamma	7.989	2.635	3.792	14.022	same prior
6	ζ	Calvo parameter	Beta	0.600	0.147	0.301	0.863	same prior
7	ι	degree of indexation	Beta	0.669	0.176	0.290	0.948	fixed to 1
8	ψ_π	inflation coefficient in policy rule	Gamma*	2.000	0.182	1.675	2.389	same prior
9	ψ_y	output coefficient in policy rule	Gamma	0.050	0.035	0.006	0.139	same prior
10	ρ_i	lagged interest rate in policy rule	Gaussian	0.801	0.100	0.606	0.996	same prior
11	ρ_τ	tax shock persistence	Beta	0.501	0.151	0.208	0.789	same prior
12	ρ_a	technology shock persistence	Beta	0.900	0.091	0.659	0.997	same prior
13	ρ_π	persistence of target shock	Beta	0.899	0.091	0.659	0.997	fixed to 1
14	$100 \times \sigma_\tau$	standard error of tax shock	Gamma	3.989	1.258	1.922	6.830	same prior
15	$100 \times \sigma_a$	standard error of technology shock	Gamma	0.334	0.150	0.108	0.681	same prior
16	$100 \times \sigma_\pi$	standard error of target shock	Gamma	0.125	0.055	0.040	0.254	same prior
17	$100 \times \sigma_i$	standard error of policy shock	Gamma	0.075	0.033	0.024	0.153	same prior
18	τ	tax shock mean	Gamma	0.399	0.281	0.048	1.119	same prior
19	π	target mean	Gamma*	1.005	0.003	1.001	1.013	not defined

Note: $\sigma_d = 0$, $\alpha = 0.76$, $\chi = 0.275$. All measurement standard errors set to 10^{-6} Gamma* for a parameter denotes a Gamma distribution for the difference between the parameter and 1.

Table 3. Results from posterior simulation of model M1

	prior distrib.		**post. distr. linear m.**				**posterior distribution, quadr. m.**			
parameter	**mean**	**sd**	**mean**	**sd**	**lower q**	**upper q**	**mean**	**sd**	**lower q**	**upper q**
β	0.900	0.090	0.994	0.001	0.992	0.996	0.994	0.002	0.991	0.998
γ	1.997	0.709	3.286	0.885	1.820	5.271	4.242	1.087	2.354	6.601
h	0.699	0.139	0.473	0.063	0.347	0.592	0.431	0.065	0.304	0.559
ϕ	4.002	0.999	5.001	1.128	3.082	7.481	4.356	0.861	2.912	6.255
θ	7.973	2.645	6.914	2.260	3.338	12.120	6.373	2.090	3.201	11.022
ζ	0.601	0.148	0.400	0.071	0.265	0.543	0.447	0.087	0.286	0.621
ι	0.667	0.178	0.076	0.038	0.022	0.167	0.073	0.036	0.019	0.158
ψ_π	1.999	0.182	1.947	0.173	1.643	2.318	1.917	0.171	1.616	2.288
ψ_y	0.050	0.035	0.044	0.031	0.006	0.124	0.058	0.042	0.007	0.163
ρ_i	0.800	0.099	0.893	0.015	0.861	0.920	0.850	0.050	0.747	0.914
ρ_τ	0.500	0.151	0.503	0.154	0.211	0.801	0.703	0.222	0.266	0.940
ρ_a	0.901	0.090	0.998	0.002	0.993	1.000	0.999	0.001	0.996	1.000
ρ_π	0.900	0.091	0.990	0.007	0.973	0.999	0.943	0.037	0.858	0.988
$100 \times \sigma_\tau$	3.986	1.257	4.154	1.344	1.949	7.206	5.859	2.245	2.171	10.031
$100 \times \sigma_a$	0.333	0.150	1.352	0.166	1.068	1.714	1.494	0.182	1.180	1.892
$100 \times \sigma_\pi$	0.125	0.056	0.134	0.019	0.101	0.175	0.174	0.051	0.106	0.297
$100 \times \sigma_i$	0.075	0.033	0.193	0.013	0.169	0.220	0.180	0.019	0.145	0.216
τ	0.400	0.283	0.448	0.323	0.054	1.267	0.836	0.520	0.093	1.981
π	1.005	0.003	1.005	0.003	1.001	1.013	1.011	0.004	1.004	1.018

Table 4. Results from posterior simulation of model M2

	prior distrib.		**post. distr. linear m.**				**post. distr. quadratic m.**			
par.	**mean**	**sd**	**mean**	**sd**	**lower**	**upper**	**mean**	**sd**	**lower**	**upper**
β	0.900	0.090	0.996	0.001	0.993	0.999	0.994	0.002	0.991	0.999
γ	1.996	0.702	3.793	1.233	1.750	6.532	2.378	0.923	1.246	4.302
h	0.700	0.137	0.499	0.066	0.369	0.628	0.501	0.064	0.358	0.625
ϕ	5.000	2.006	2.429	0.731	1.395	4.220	3.397	0.937	2.012	5.203
θ	7.979	2.646	7.327	2.369	3.615	12.777	6.803	1.760	3.834	10.664
ζ	0.601	0.147	0.465	0.090	0.280	0.634	0.300	0.070	0.175	0.436
ψ_π	1.497	0.352	1.451	0.284	1.095	2.169	1.397	0.235	1.134	1.971
ψ_y	0.050	0.035	0.053	0.038	0.007	0.149	0.045	0.028	0.011	0.108
ρ_i	0.700	0.138	0.878	0.020	0.838	0.915	0.876	0.022	0.836	0.922
ρ_τ	0.500	0.151	0.153	0.068	0.050	0.311	0.305	0.143	0.123	0.604
ρ_a	0.900	0.091	0.999	0.001	0.996	1.000	0.997	0.002	0.993	0.999
$100 \times \sigma_\tau$	4.006	3.999	29.949	11.119	11.313	54.859	16.526	8.119	3.197	32.065
$100 \times \sigma_a$	0.332	0.468	2.102	0.541	1.234	3.330	1.322	0.239	0.895	1.820
$100 \times \sigma_\pi$	0.075	0.105	0.196	0.058	0.092	0.309	0.260	0.035	0.184	0.331
$100 \times \sigma_i$	0.124	0.177	0.187	0.013	0.165	0.214	0.188	0.016	0.160	0.222
τ	0.402	0.286	1.049	0.403	0.422	1.989	0.929	0.502	0.245	1.954

Table 5. Marginal likelihood comparisons

	AT1				AT2			
	linear		quadratic		linear		quadratic	
truncation	**ML**	**emp. cov.**	**ML**	**emp. cov.**	**ML**	**emp. cov.**	**ML**	**emp. cov.**
0.1	1714.06	0.13	1715.21	0.16	1687.45	0.16	1677.30	0.08
0.2	1714.07	0.25	1714.78	0.28	1687.46	0.28	1677.12	0.19
0.3	1714.07	0.36	1714.85	0.38	1687.45	0.39	1676.65	0.31
0.4	1714.08	0.45	1714.82	0.47	1687.45	0.48	1676.68	0.42
0.5	1714.09	0.54	1714.87	0.56	1687.46	0.57	1676.50	0.53
0.6	1714.08	0.63	1714.93	0.64	1687.49	0.65	1676.33	0.63
0.7	1714.10	0.71	1715.00	0.72	1687.50	0.73	1676.32	0.72
0.8	1714.11	0.79	1715.04	0.79	1687.52	0.80	1676.32	0.81
0.9	1714.10	0.88	1715.06	0.87	1687.55	0.88	1676.35	0.91

Note: based on the modified Gelfand and Dey approach (see Geweke, 1999): integral is approximated by fitting truncated (using theoretical coverage) multivariate Gaussian to posterior distribution. For computations to be reliable empirical coverage should match theoretical coverage. AT1 estimated using 1970:1-2004:4; AT2 estimated using 1970:2-2004:4.

Table 6. Model M1, forecast error variance decomposition (FEVD) at different horizons; posterior means reported

variable	horizon	tech. shock	target shock	tax shock	mon. pol. sh.
inflation	1	2.54	76.24	7.53	8.52
interest	1	1.20	7.41	3.28	83.95
output	1	88.21	5.35	1.84	2.43
inflation	4	1.06	83.22	4.67	6.45
interest	4	1.62	34.91	9.05	38.85
output	4	92.28	2.92	2.02	1.11
inflation	12	0.70	88.07	3.38	4.23
interest	12	0.96	64.45	7.64	13.88
output	12	96.39	1.13	1.38	0.35
inflation	40	0.62	90.57	2.61	3.16
interest	40	0.91	75.90	4.29	8.80
output	40	98.84	0.36	0.45	0.11

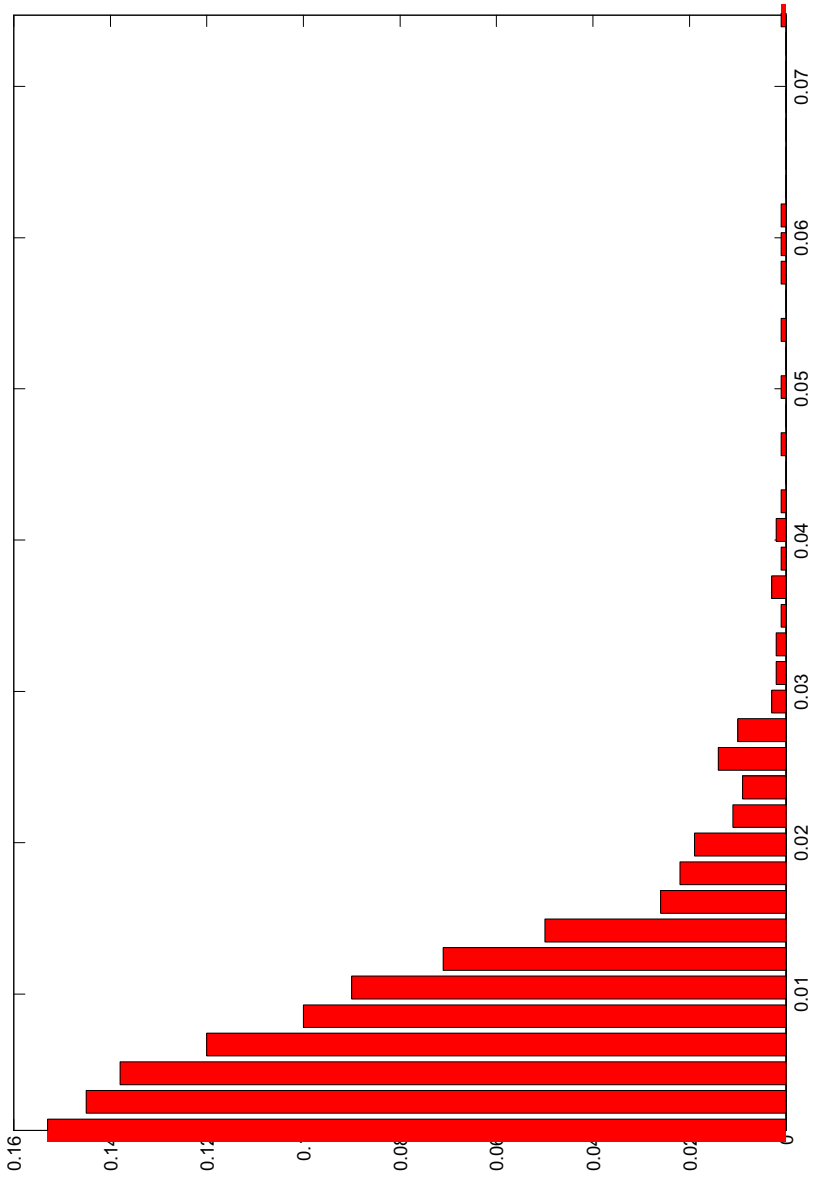

Fig. 1. Distribution of absolute percentage approximation errors using conditional particle filter, N=2000 versus N=50000, for different parameter values taken from their joint posterior distribution

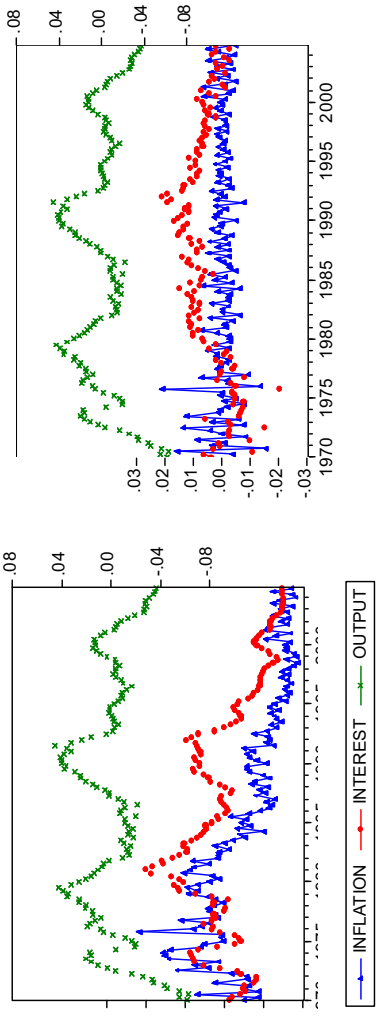

Fig. 2. Model datasets for M1 and M2

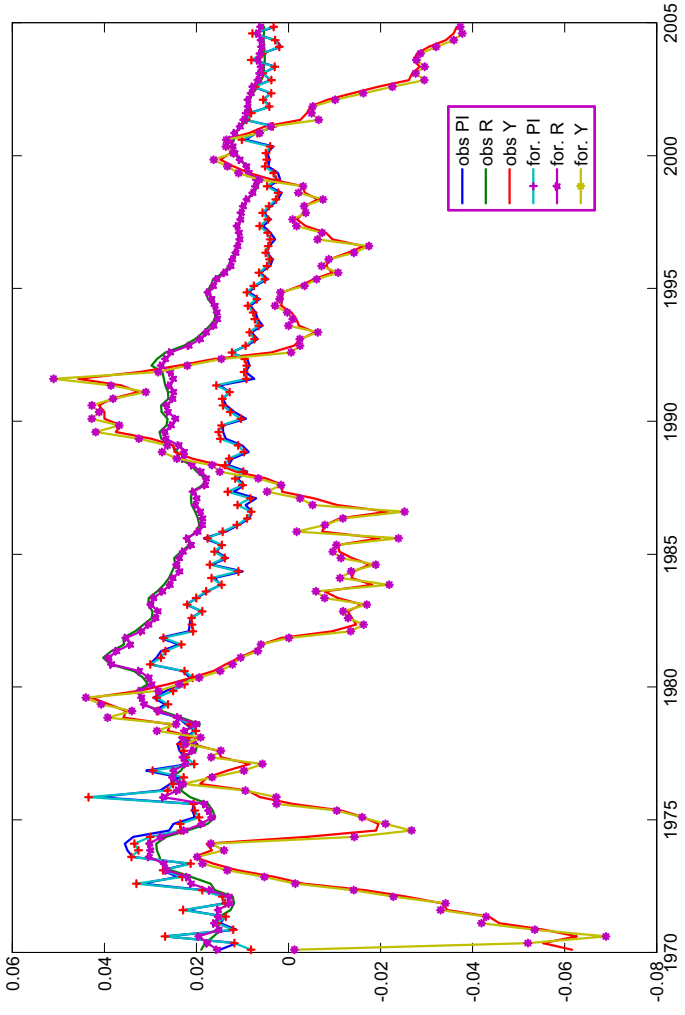

Fig. 3. Actual series and one step ahead forecasts

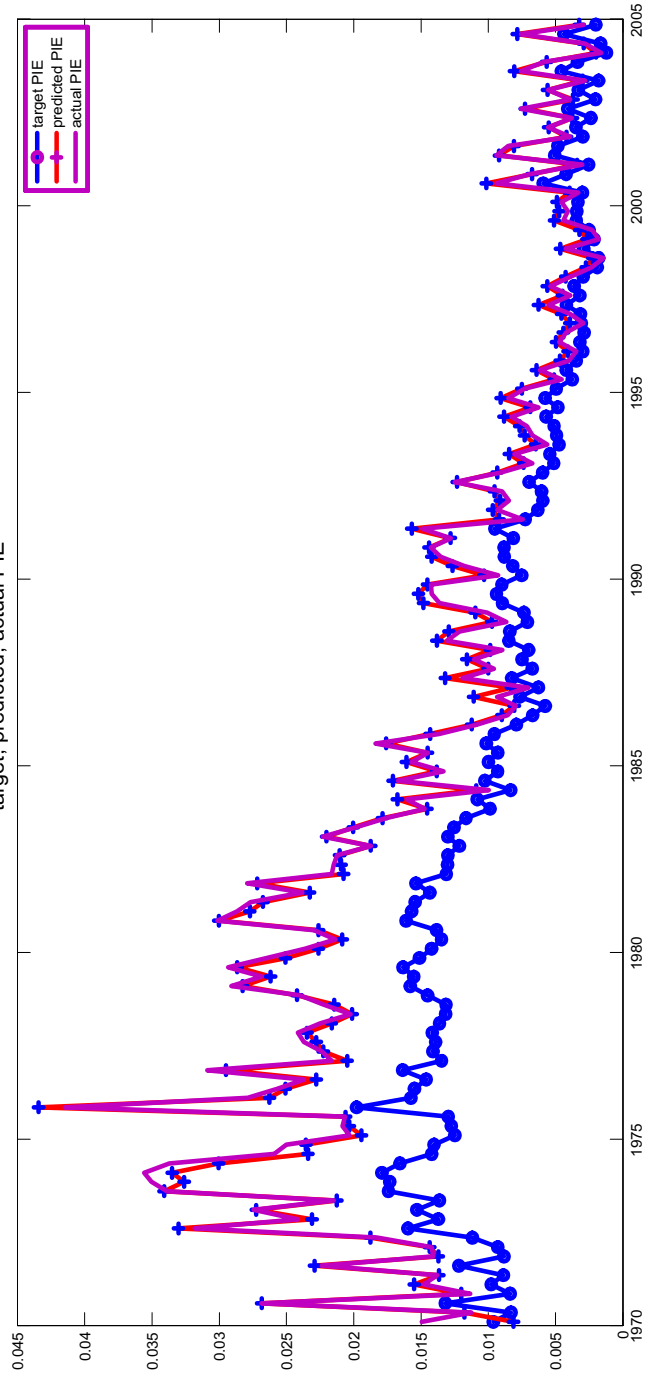

Fig. 4. Posterior simulation of model M1: actual inflation, target inflation and 1-step ahead forecast (posterior means)

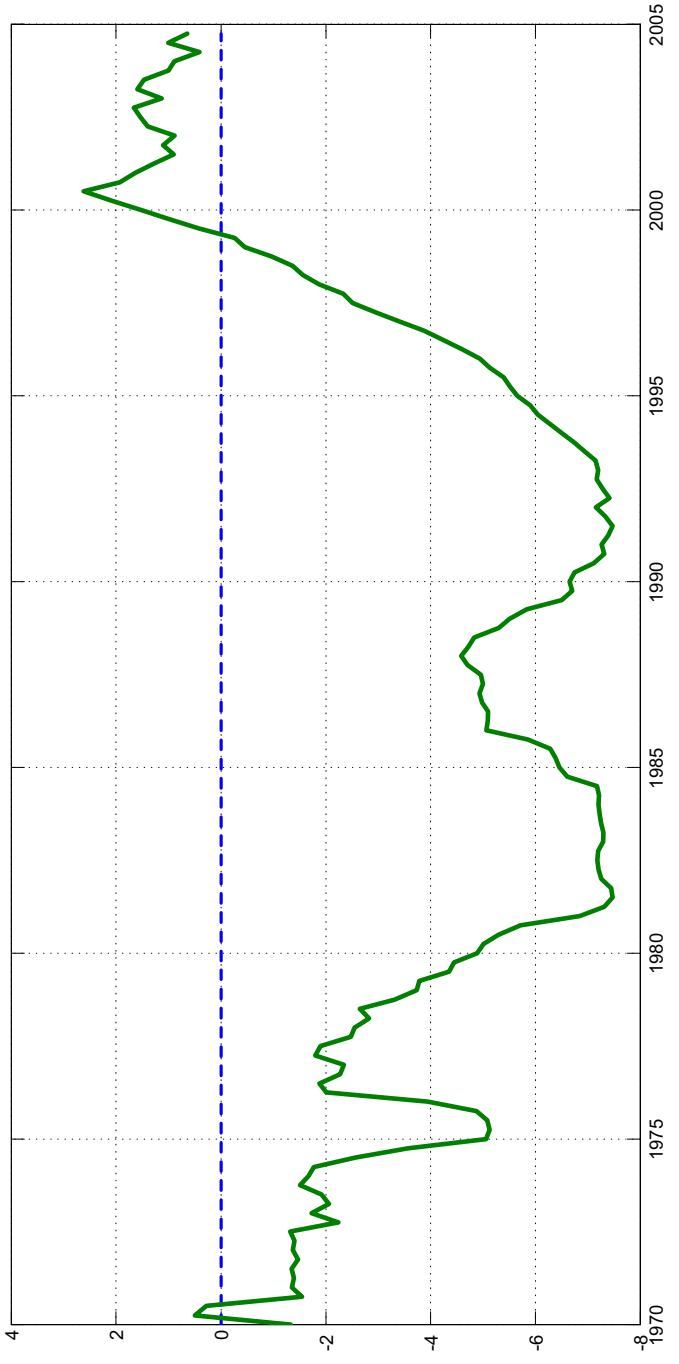

Fig. 5. Decomposition of log conditional predictive pdf, linear-quadratic, model M1

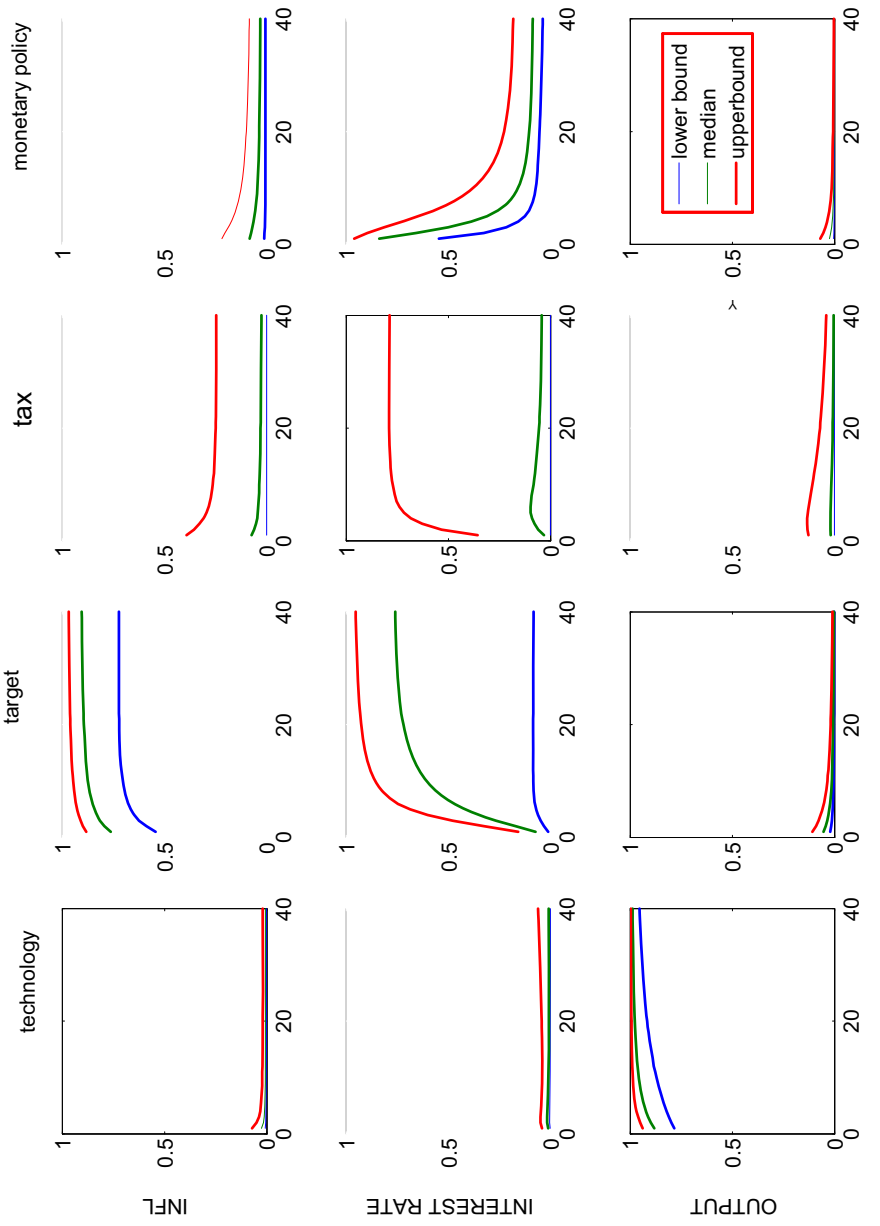

Fig. 6. FEVD for Model M1

Fig. 7. Impulse responses for model M1: technology and target shocks

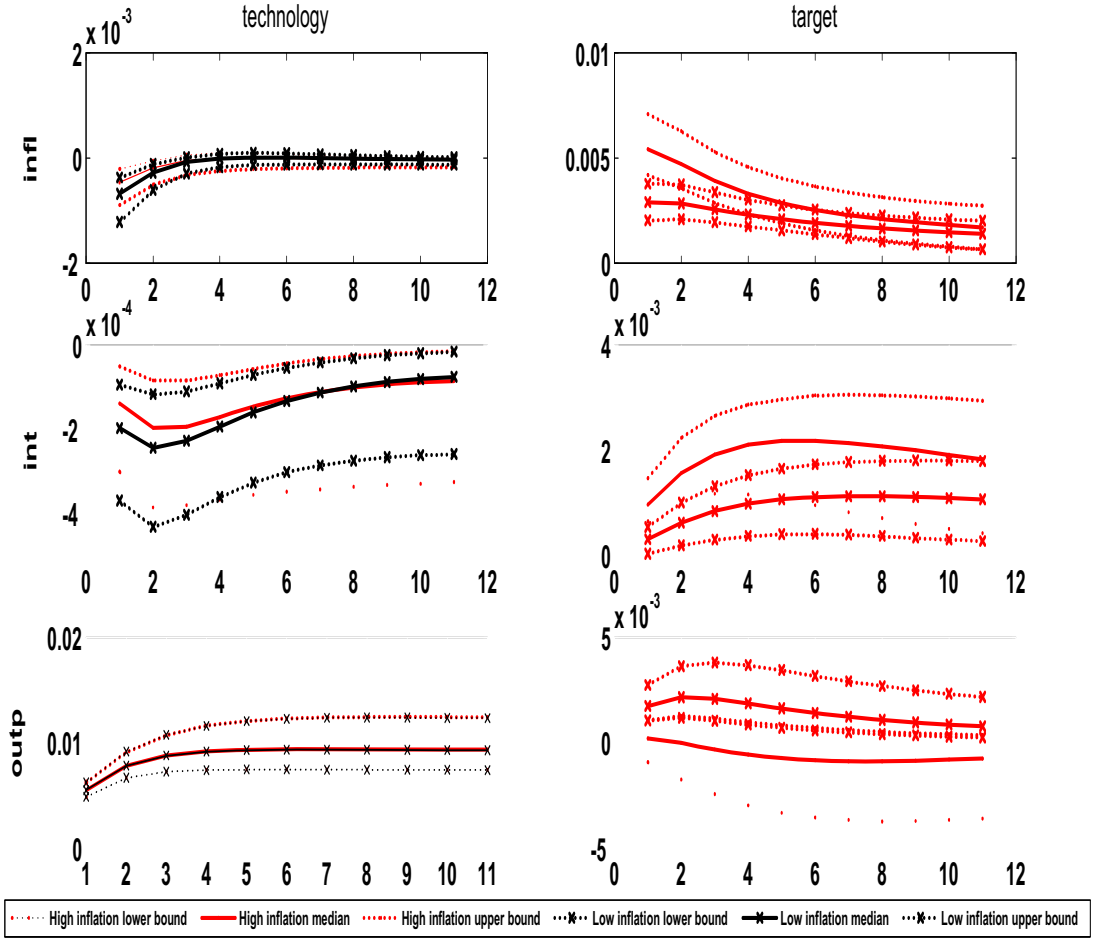

Fig. 8. Impulse responses for model M1: tax and monetary policy shocks

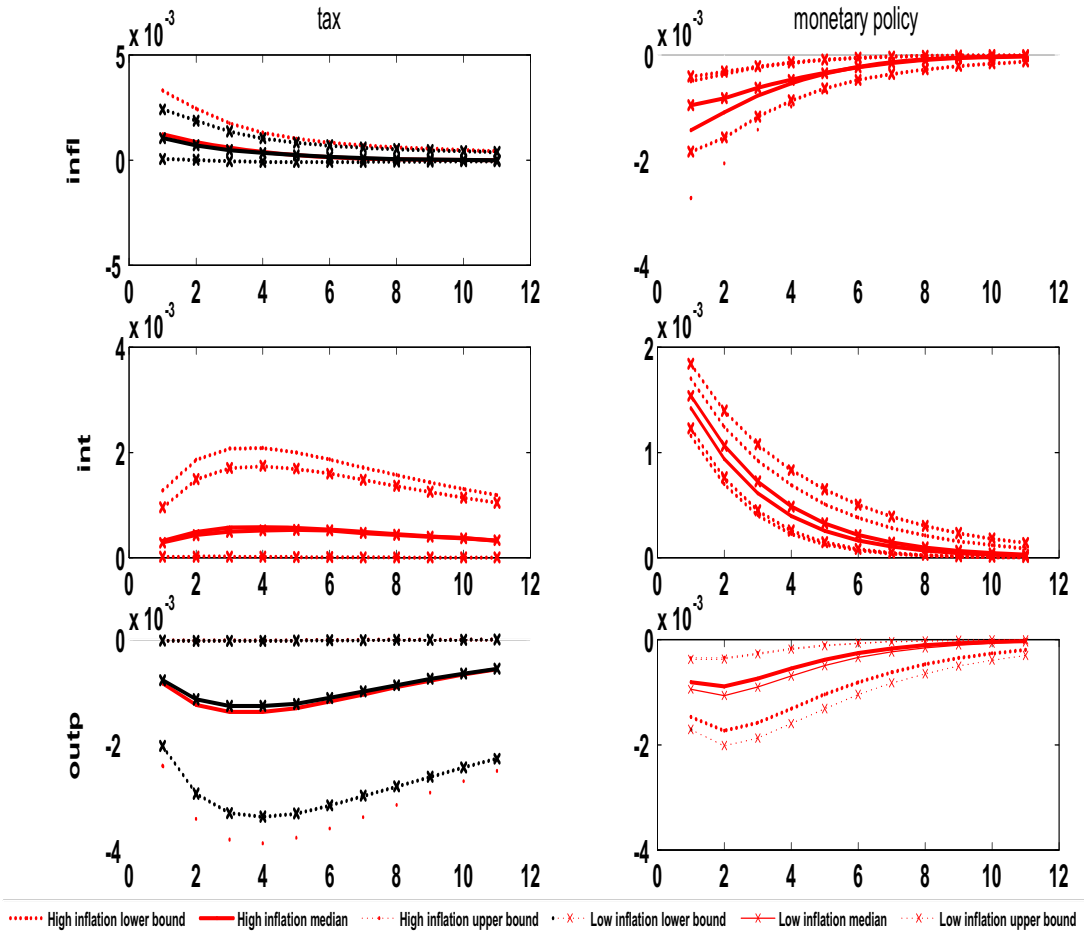

Fig. 9.