EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1341 / MAY 2011

EXACT LIKELIHOOD COMPUTATION FOR NONLINEAR DSGE MODELS WITH HETEROSKEDASTIC INNOVATIONS

by Gianni Amisano
and Oreste Tristani

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1341 / MAY 2011

EXACT LIKELIHOOD COMPUTATION FOR NONLINEAR DSGE MODELS WITH HETEROSKEDASTIC INNOVATIONS

by Gianni Amisano [1] and Oreste Tristani [2]

In 2011 all ECB publications feature a motif taken from the €100 banknote.

1 DG-Research, European Central Bank, Kaiserstrasse 29, D-60311, Frankfurt, Germany;and Department of Economics University of Brescia; e-mail: gianni.amisano@ecb.europa.eu; amisano@eco.unibs.it

2 DG-Research, European Central Bank, Kaiserstrasse 29, D-60311, Frankfurt, Germany; e-mail: oreste.tristani@ecb.europa.eu

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract

Phenomena such as the Great Moderation have increased the attention of macroeconomists towards models where shock processes are not (log-)normal. This paper studies a class of discrete-time rational expectations models where the variance of exogenous innovations is subject to stochastic regime shifts. We first show that, up to a second-order approximation using perturbation methods, regime switching in the variances has an impact only on the intercept coefficients of the decision rules. We then demonstrate how to derive the exact model likelihood for the second-order approximation of the solution when there are as many shocks as observable variables. We illustrate the applicability of the proposed solution and estimation methods in the case of a small DSGE model.

Key words: DSGE models, second-order approximation, regime switching, time-varying volatility.

JEL classification: E0, C63

Non-technical summary

The estimation of linear DSGE (Dynamic Stochastic General Equilibrium) models is commonplace in macroeconomics. By and large, applications tend to rely on the assumption that shocks have constant variance over time. However, there are reasons to also explore the role of nonlinearities and/or time-varying variances in macro-models. Nonlinearities are important to understand asset prices or crisis-type phenomena. Time-varying variances are important to account for stylised facts like the “Great moderation” of the eighties and nineties.

The main obstacle to the estimation of nonlinear DSGE models has been computational. The likelihood of nonlinear models, and hence their dynamic statistical properties, are not known exactly and must be approximated by methods with very slow convergence properties. This computational difficulty increases further when the assumption of constant volatility is relaxed.

The main contribution of this paper is to develop a methodology to compute exactly the likelihood of quadratic (and possibly more highly nonlinear) DSGE models, including models in which the variance of the shocks can change randomly across different regimes. If two technical conditions are satisfied, the methodology is generally applicable.

We illustrate our proposed methodology in the case of a small DSGE model with nominal rigidities, where shocks are potentially characterised by high or low variance regimes. The model is estimated on US data over the 1966Q1-2009Q1 sample. Consistently with the aforementioned evidence, the estimates recover the Great moderation and attribute it to a fall in the variance of technology shocks. The results are also consistent with an end of the Great moderation in the early years of the new millennium.

1 Introduction

The estimation of the first-order approximation of the solution of nonlinear DSGE (dynamic stochastic general equilibrium) models is now commonplace in macroeconomics. By and large, applications tend to rely on the assumption that the state vector of the system is hit by i.i.d. innovations.

The estimation of higher order approximations of the solution of DSGE models has proven more difficult. The problem is that the shape of the likelihood function is unknown and must itself be approximated through numerical methods. While some general methods, namely applications of the particle filter, have been proposed in this context (see e.g. Fernandez-Villaverde and Rubio-Ramirez, 2007), these methods require a considerable effort from the computational viewpoint. The approximation only converges in probability to the true likelihood when the number of particles used for each likelihood evaluation goes to infinity, and there are no general results ensuring that approximation errors are negligible when only a finite, albeit large number of particles is used.

The computational burden increases further when the assumption of i.i.d. innovations is relaxed. While such relaxation is not common in macroeconomics, a number of authors have forcefully argued for the presence of heteroskedasticity in macroeconomic data. For example, the literature on the so-called Great Moderation (see e.g. McConnell and Perez-Quiros, 2000) finds evidence for regime-switching in the variance of macroeconomic variables. Evidence for regime-switching is also found in Sims and Zha (2006), while Cogley and Sargent (2005), Primiceri (2005) and Fernandez-Villaverde and Rubio-Ramirez (2007), amongst others, argue in favour of stochastic volatility. A simpler procedure to compute the likelihood of nonlinear DSGE models with heteroskedastic shocks would clearly be appealing.

In this paper, we propose such a procedure for the case of nonlinear DSGE models in which the variances of innovations to the state vector is subject to stochastic regime switches. Our procedure ultimately amounts to inverting the observation equation for the unobservable state variables of the model. As in the case of linearised models with i.i.d. innovations, the likelihood function deriving from the state space representation can be computed exactly. The state space representation, however, derives from an approximate solution of the model (notably a second-order approximation). This feature makes our procedure appealing compared to existing alternatives, such as the particle filter.

Compared to the particle filter, however, our likelihood computation method is not generally applicable, but can only be used under two specific conditions. The first one is that the number of variables used in estimation must be equal to the number of stochastic innovations (including both structural innovations and measurement errors). We believe that this is a soft restriction, since additional measurement errors can always be introduced when additional variables are included in the econometrician's information set. This restriction does however preclude the use of our methodology in cases where there are more shocks than variables. The second condition which needs to be met for the applicability of our method is that the economic system cannot include unobservable non-stochastic state variables. This implies, for example, that a general macroeconomic model can only be estimated using our proposed method if capital can be treated as an observable variable.

To analyse the case of models with heteroskedastic shocks, we first need to demonstrate how this class of models can be solved using fast methods. In general, the solution of DSGE models with regime-switching coefficients requires computationally expensive methods (e.g. Coleman, 1991, Andolfatto and Gomme, 2003, Davig, Leeper and Chung, 2004, use an Euler equation iteration technique). In the case we focus on, however, regime switching only affects the variance of structural shocks. Our case can therefore be tackled using standard perturbation methods (see Judd, 1998, and the references therein, Schmitt-Grohé and Uribe, 2001; Kim, Kim, Schaumburg and Sims, 2003; Gomme and Klein, 2006; Lombardo and Sutherland, 2007). We show that, up to a second-order approximation, the coefficients on the linear and quadratic terms in the state vector of the decision rules are independent of the volatility of the exogenous shocks. The only impact of regime switching is on the constant terms of the decision rules, which become regime-dependent.

We finally illustrate our proposed method with an application to a small DSGE model with nominal rigidities, where technology and government spending shocks are potentially characterised by high or low variance regimes. We estimate the model on US data over the 1966Q1-2009Q1 sample and find evidence consistent with the assumption that the Great Moderation is related to a fall in the variance of technology shocks. There is instead only weak evidence of heteroskedasticity in government spending shocks.

2 A general model with heteroskedastic conditional variances

We are interested in a general nonlinear model of the form

$$\mathrm{E}_t \left[f \left(\mathbf{y}_{t+1}, \mathbf{y}_t, \mathbf{x}_{1t+1}, \mathbf{x}_{1t}, \mathbf{x}_{2t+1}, \mathbf{x}_{2t}; \mathbf{s}_t \right) \right] = 0 \tag{1}$$

where E_t is the expectation operator conditional on information available at time t, $\mathbf{y}_t$ represent a vector of non-predetermined variables, $\mathbf{x}_{1t}$ is a vector of endogenous predetermined variables, $\mathbf{x}_{2t}$ are exogenous variables with continuous support, and $\mathbf{s}_t$ is a vector including indicators of discrete regimes. The vectors have length n_y, n_{x1}, n_{x2} and n_s, respectively. Note that

$$\mathbf{x}_{2t+1} = \mathbf{A}\mathbf{x}_{2t} + \sigma\mathbf{\Sigma}_{s_t}\varepsilon_{t+1} \tag{2}$$
$$\xi(\mathbf{s}_{t+1}) = \mathbf{B}\xi(\mathbf{s}_t) + \nu_{t+1} \tag{3}$$

for known functions $\mathbf{A}$ and $\mathbf{\Sigma}_{s_t}$, mapping $R^{n_{x2}}$ into $R^{n_{x2}}$, and $\mathbf{B}$, mapping R^{n_s} into R^{n_s}. The function $\mathbf{A}$ is such that all eigenvalues of its first derivative evaluated at the non-stochastic steady state lie within the unit circle. The innovation vector ε_{t+1} is independently and identically distributed, with zero mean and unit variance. The vector ν_{t+1} has zero mean and heteroskedastic variance. We assume ε_{t+1} and ν_{t+1} to be mutually uncorrelated. The scalar σ is the perturbation parameter.

The key distinguishing feature of the model in equations (1)-(3) is the presence of the discrete regimes $\mathbf{s}_t$, which characterise the conditional variance of the shocks ε_{t+1}. This formulation generalises the standard model with Gaussian innovations typically studied in the macroeconomic literature. The generalisation allows one to analyse the nature of some stylised facts in a micro-founded setting, including for example the structural sources of the Great Moderation. A formulation with regime-switching conditional variances also enables one to study in an arguably more satisfactory fashion the implications of micro-founded models for asset prices (see e.g. Amisano and Tristani, 2011).

Equations (1)-(3) do not, however, represent the most general possible formulation of a model with regime shifts. More specifically, we do not allow regime switching to influence any other structural parameters than the variances of the shocks. This assumption allows us to compute the solution through a straightforward and fast extension of standard perturbation methods. More involved solution methods are required to appropriately capture the effects of regime switching in, for example, the parameters of the monetary policy rule in a linearised macro-model (see e.g. Davig and Leeper, 2007; Farmer, Waggoner and Zha, 2010). The development of fast and efficient solution methods for general nonlinear models with regime-switching parameters is currently an active area of research.

In order to write an n_s-state Markov chain $\mathbf{s}_t$ as in equation (3), we rely on Hamilton (1994) which shows that $\xi(\mathbf{s}_t)$ must be a vector whose i-th element is equal to 1 if $s_t = i$ and zero otherwise, $\mathbf{B}$ is the transition matrix of the Markov chain, and $\nu_{t+1} \equiv \mathbf{s}_{t+1} - \mathrm{E}(\mathbf{s}_{t+1}|\mathbf{s}_t)$.

Going back to the general model (1)-(3), we can define the vectors

$$\mathbf{x}'_t \equiv [\mathbf{x}'_{1t}, \mathbf{x}'_{2t}], \mathbf{u}'_t \equiv [\mathbf{0}', \varepsilon'_t]$$

and the matrix

$$\widetilde{\mathbf{\Sigma}}_{s_t} \equiv \begin{bmatrix} \underset{n_{x1}\times n_{x1}}{\mathbf{0}} & \underset{n_{x1}\times n_{x2}}{\mathbf{0}} \\ \underset{n_{x2}\times n_{x1}}{\mathbf{0}} & \mathbf{\Sigma}_{s_t} \end{bmatrix}$$

so the model in equation (1) can be rewritten as

$$\mathrm{E}_t \left[f \left(\mathbf{y}_{t+1}, \mathbf{y}_t, \mathbf{x}_{t+1}, \mathbf{x}_t; \mathbf{s}_t \right) \right] = \mathbf{0}$$

and the solution is of the form

$$\mathbf{y}_t = g\left(\mathbf{x}_t, \sigma; \mathbf{s}_t\right) \tag{4}$$

$$\mathbf{x}_{t+1} = h\left(\mathbf{x}_t, \sigma; \mathbf{s}_t\right) + \sigma \widetilde{\mathbf{\Sigma}}_{s_t} \mathbf{u}_{t+1} \tag{5}$$

3 Approximating the solution

We seek a second-order approximation to the functions $g\left(\mathbf{x}_t, \sigma; s_t\right)$ and $h\left(\mathbf{x}_t, \sigma; s_t\right)$ around the non-stochastic steady state $\mathbf{x}_t = \overline{\mathbf{x}}_{s_t}$ and $\sigma = 0$. We leave the dependence on s_t outside the Taylor expansion. In other words, we seek for an expansion such that the coefficients of the second-order approximate solution are potentially functions of s_t.

To write the approximation, we follow Gomme and Klein (2006) and use the following representation (from Magnus and Neudecker, 1999) of the second-order Taylor expansion of a twice-differentiable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

$$f\left(\mathbf{x}\right) \approx f\left(\overline{\mathbf{x}}; s_t\right) + \left[\mathrm{D}f\left(\overline{\mathbf{x}}; s_t\right)\right]\left(\mathbf{x} - \overline{\mathbf{x}}_{s_t}\right) + \frac{1}{2}\left(\mathbf{I}_m \otimes \left(\mathbf{x} - \overline{\mathbf{x}}_{s_t}\right)'\right)\left[\mathrm{H}f\left(\overline{\mathbf{x}}; s_t\right)\right]\left(\mathbf{x} - \overline{\mathbf{x}}_{s_t}\right)$$

where $\mathrm{D}f\left(\overline{\mathbf{x}}; s_t\right)$ and $\mathrm{H}f\left(\overline{\mathbf{x}}; s_t\right)$ are the gradient and Hessian matrices structured as follows

$$\mathrm{D}f\left(\overline{\mathbf{x}}; s_t\right) \equiv \begin{bmatrix} \frac{\partial f_1(\overline{\mathbf{x}}; s_t)}{\partial x_1} & \frac{\partial f_1(\overline{\mathbf{x}}; s_t)}{\partial x_2} & \cdots & \frac{\partial f_n(\overline{\mathbf{x}}; s_t)}{\partial x_n} \\ \frac{\partial f_2(\overline{\mathbf{x}}; s_t)}{\partial x_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_m(\overline{\mathbf{x}}; s_t)}{\partial x_1} & \cdots & \cdots & \frac{\partial f_m(\overline{\mathbf{x}}; s_t)}{\partial x_n} \end{bmatrix}$$

and

$$\mathrm{H}f\left(\overline{\mathbf{x}}; s_t\right) \equiv \mathrm{D}\ \mathrm{vec}\left[\left(\mathrm{D}f\left(\overline{\mathbf{x}}; s_t\right)\right)'\right]$$

Using the solution (4)-(5), we can rewrite model (1) as the function

$$F(\mathbf{x}_t,\sigma;s_t) = \mathrm{E}_t\left[f\left\{\begin{array}{c} h(\mathbf{x}_t,\sigma;s_t)+\sigma\widetilde{\mathbf{\Sigma}}_{s_t}\mathbf{u}_{t+1}, \\ g\left[h(\mathbf{x}_t,\sigma;s_t)+\sigma\widetilde{\mathbf{\Sigma}}_{s_t}\mathbf{u}_{t+1},\sigma;\mathbf{B}\mathbf{s}_t+\nu_{t+1}\right], \\ \mathbf{x}_t, g(\mathbf{x}_t,\sigma;s_t);s_t \end{array}\right\}\right]$$
$$= \mathbf{0}$$

We analyse the second-order approximation to the functions h and g which can be represented as

$$g(\mathbf{x}_t,\sigma;s_t) = g(\overline{\mathbf{x}};0;s_t)+\mathbf{F}_{s_t}(\mathbf{x}_t-\overline{\mathbf{x}}_{s_t})+\frac{1}{2}\left(\mathbf{I}_{n_y}\otimes(\mathbf{x}_t-\overline{\mathbf{x}}_{s_t})'\right)\mathbf{E}_{s_t}(\mathbf{x}_t-\overline{\mathbf{x}}_{s_t})+\mathbf{k}_{y,s_t}\sigma^2$$

and

$$h(\mathbf{x}_t,\sigma;s_t) = h(\overline{\mathbf{x}};0;s_t)+\mathbf{P}_{s_t}(\mathbf{x}_t-\overline{\mathbf{x}}_{s_t})+\frac{1}{2}\left(\mathbf{I}_{n_x}\otimes(\mathbf{x}_t-\overline{\mathbf{x}}_{s_t})'\right)\mathbf{G}_{s_t}(\mathbf{x}_t-\overline{\mathbf{x}}_{s_t})+\mathbf{k}_{x,s_t}\sigma^2$$

for potentially state-dependent vectors and matrices $\mathbf{F}_{s_t}$, $\mathbf{E}_{s_t}$, $\mathbf{P}_{s_t}$, $\mathbf{G}_{s_t}$, $\mathbf{k}_{y,s_t}$, $\mathbf{k}_{x,s_t}$.

3.1 Steady state

By definition, $\overline{\mathbf{y}}(s_t) = g(\overline{\mathbf{x}},0;s_t)$ and $\overline{\mathbf{x}}_{s_t} = h(\overline{\mathbf{x}},0;s_t)$. In general, the steady state of a model with regime switches would be a function of the discrete regimes s_t. Given our assumption that discrete regimes only have an impact on (1) through the variance of the innovations, however, the steady state which arises when $\sigma = 0$ is not regime-dependent. Thus $\overline{\mathbf{x}}_{s_t} = \overline{\mathbf{x}}$ and $\overline{\mathbf{y}}(s_t) = \overline{\mathbf{y}}$. We can therefore simplify the form of the second -order approximations as

$$g(\widehat{\mathbf{x}}_t,\sigma;s_t) = \overline{\mathbf{y}}+\mathbf{F}_{s_t}\widehat{\mathbf{x}}_t+\frac{1}{2}\left(\mathbf{I}_{n_y}\otimes\widehat{\mathbf{x}}_t'\right)\mathbf{E}_{s_t}\widehat{\mathbf{x}}_t+\mathbf{k}_{y,s_t}\sigma^2$$

and

$$h(\widehat{\mathbf{x}}_t,\sigma;s_t) = \overline{\mathbf{x}}+\mathbf{P}_{s_t}\widehat{\mathbf{x}}_t+\frac{1}{2}(\mathbf{I}_{n_x}\otimes\widehat{\mathbf{x}}_t')\mathbf{G}_{s_t}\widehat{\mathbf{x}}_t+\mathbf{k}_{x,s_t}\sigma^2$$

where $\widehat{\mathbf{x}}_t \equiv \mathbf{x}_t-\overline{\mathbf{x}}$.

3.2 First-order approximation

The assumed form of the solution implies that

$$F(\mathbf{x}_t, \sigma; s_t) = \mathrm{E}_t \left[f \left\{ \begin{array}{c} P_{s_t}\widehat{\mathbf{x}}_t + \sigma \widetilde{\mathbf{\Sigma}}_{s_t}\mathbf{u}_{t+1}, \\ \mathbf{F}_{s_t}\left[\mathbf{P}_{s_t}\widehat{\mathbf{x}}_t + \sigma \widetilde{\mathbf{\Sigma}}_{s_t}\mathbf{u}_{t+1}, \sigma; \mathbf{B}\mathbf{s}_t + \nu_{t+1}\right], \\ \mathbf{x}_t, \mathbf{F}_{s_t}\widehat{\mathbf{x}}_t; s_t \end{array} \right\} \right] = \mathbf{0}$$

In order to identify the coefficients $\mathbf{F}_{s_t}$ and $\mathbf{P}_{s_t}$, we exploit the property that the solution must be such that $\mathrm{D}F_1(\overline{\mathbf{x}}, 0; s_t) = 0$, where a subscript i indicates differentiation with respect to vector i in a given function. Using the definition of function $F(x, \sigma; s_t)$, we obtain

$$\mathrm{D}F_1(\overline{\mathbf{x}}, 0; s_t) = \mathbf{f}_1\mathbf{P}_{s_t} + \mathbf{f}_2\mathbf{F}_{s_t}\mathbf{P}_{s_t} + \mathbf{f}_3 + \mathbf{f}_4\mathbf{F}_{s_t} = [\mathbf{0}]$$

which is a (potentially regime dependent) system of quadratic equations in the elements of $\mathbf{F}_{s_t}$ and $\mathbf{P}_{s_t}$.

Note that this equation implies that

$$\mathbf{f}_1\mathbf{P}_{s_t}\mathrm{E}_t\widehat{\mathbf{x}}_{t+1} + \mathbf{f}_2\mathbf{F}_{s_t}\mathbf{P}_{s_t}\mathrm{E}_t\widehat{\mathbf{y}}_{t+1} + \mathbf{f}_3\widehat{\mathbf{x}}_t + \mathbf{f}_4\mathbf{F}_{s_t}\widehat{\mathbf{y}}_t = [\mathbf{0}]$$

or

$$\begin{bmatrix} \mathbf{f}_1\mathbf{P}_{s_t} & \mathbf{f}_2\mathbf{F}_{s_t}\mathbf{P}_{s_t} \end{bmatrix} \begin{bmatrix} \mathrm{E}_t\widehat{\mathbf{x}}_{t+1} \\ \mathrm{E}_t\widehat{\mathbf{y}}_{t+1} \end{bmatrix} + \begin{bmatrix} \mathbf{f}_3 & \mathbf{f}_4\mathbf{F}_{s_t} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{x}}_t \\ \widehat{\mathbf{y}}_t \end{bmatrix} = [\mathbf{0}]$$

Using

$$\begin{aligned} \mathrm{E}_t\widehat{\mathbf{x}}_{t+1} &= \begin{bmatrix} \mathrm{E}_t\widehat{\mathbf{x}}_{1t+1} \\ \mathrm{E}_t\widehat{\mathbf{x}}_{2t+1} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{x}}_{1t+1} \\ \widehat{\mathbf{x}}_{2t+1} \end{bmatrix} - \sigma\widetilde{\mathbf{\Sigma}}_{s_t} \begin{bmatrix} \mathbf{0} \\ \varepsilon_{t+1} \end{bmatrix} \\ &= \widehat{\mathbf{x}}_{t+1} - \sigma\widetilde{\mathbf{\Sigma}}_{s_t}\mathbf{u}_t \end{aligned}$$

we can further rewrite this as

$$\begin{bmatrix} \mathbf{f}_1\mathbf{P}_{s_t} & \mathbf{f}_2\mathbf{F}_{s_t}\mathbf{P}_{s_t} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{x}}_{t+1} \\ \mathrm{E}_t\widehat{\mathbf{y}}_{t+1} \end{bmatrix} = \begin{bmatrix} -\mathbf{f}_3 & -\mathbf{f}_4\mathbf{F}_{s_t} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{x}}_t \\ \widehat{\mathbf{y}}_t \end{bmatrix} + \sigma \begin{bmatrix} \mathbf{f}_1\mathbf{P}_{s_t} & \mathbf{f}_2\mathbf{F}_{s_t}\mathbf{P}_{s_t} \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \end{bmatrix}$$

or

$$\mathbf{A}_{s_t} \begin{bmatrix} \widehat{\mathbf{x}}_{t+1} \\ \mathrm{E}_t\widehat{\mathbf{y}}_{t+1} \end{bmatrix} = \mathbf{B}_{s_t} \begin{bmatrix} \widehat{\mathbf{x}}_t \\ \widehat{\mathbf{y}}_t \end{bmatrix} + \sigma\mathbf{C}_{s_t} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \end{bmatrix}$$

By assumption, $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ and $\mathbf{f}_4$ are not regime dependent in system (1). Hence, the solution such that the $\mathbf{A}$ and $\mathbf{B}$ matrices are regime independent is a solution of the system. This solution can be obtained using standard algorithms (see Blanchard and Kahn, 1980, Sims, 2001, Klein, 2000, Söderlind, 1999). The conditions for the local uniqueness of the solution are also unchanged.

Our first-order approximation should also include conditions for which $\mathrm{D}F_2(\overline{z}, 0) = [\mathbf{0}]$. As in Schmitt-Grohé and Uribe (2001), the resulting equations would be linear and homogeneous in the coefficients attached to σ, which implies that those coefficients must be zero.

Hence, to a first-order approximation, the solution is not affected by the variance of the shocks, and specifically the variance of the Markov-switching states.

3.3 Second-order approximation

The assumed solutions imply that (exploiting the finding that $\mathbf{P}_{s_t} = \mathbf{P}$ and $\mathbf{F}_{s_t} = \mathbf{F}$)

$$
\begin{aligned}
&F(\mathbf{x}_t, \sigma; s_t) \\
&= \mathrm{E}_t \left[f \left\{ \begin{array}{c}
\mathbf{P}\widehat{\mathbf{x}}_t + \frac{1}{2}(\mathbf{I}_{n_x} \otimes \widehat{\mathbf{x}}_t')\,\mathbf{G}_{s_t}\widehat{\mathbf{x}}_t + \mathbf{k}_{x,s_t}\sigma^2 + \sigma\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}, \\
\mathbf{F}\left[\mathbf{P}\widehat{\mathbf{x}}_t + \frac{1}{2}(\mathbf{I}_{n_x} \otimes \widehat{\mathbf{x}}_t')\,\mathbf{G}_{s_t}\widehat{\mathbf{x}}_t + \mathbf{k}_{x,s_t}\sigma^2 + \sigma\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}, \sigma; \mathbf{B}\mathbf{s}_t + \nu_{t+1}\right] \\
+\frac{1}{2}\left(\mathbf{I}_{n_y} \otimes \left[\mathbf{P}\widehat{\mathbf{x}}_t + \frac{1}{2}(\mathbf{I}_{n_x} \otimes \widehat{\mathbf{x}}_t')\,\mathbf{G}_{s_t}\widehat{\mathbf{x}}_t + \mathbf{k}_{x,s_t}\sigma^2 + \sigma\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}, \sigma; \mathbf{B}\mathbf{s}_t + \nu_{t+1}\right]'\right) \cdot \\
\cdot\mathbf{E}_{s_t}\left[\mathbf{P}\widehat{\mathbf{x}}_t + \frac{1}{2}(\mathbf{I}_{n_x} \otimes \widehat{\mathbf{x}}_t')\,\mathbf{G}_{s_t}\widehat{\mathbf{x}}_t + \mathbf{k}_{x,s_t}\sigma^2 + \sigma\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}, \sigma; \mathbf{B}\mathbf{s}_t + \nu_{t+1}\right] + \mathbf{k}_{y,s_t}\sigma^2, \\
\mathbf{x}_t, F\widehat{\mathbf{x}}_t + \frac{1}{2}\left(I_{n_y} \otimes \widehat{\mathbf{x}}_t'\right) E_{s_t}\widehat{\mathbf{x}}_t + \mathbf{k}_{y,s_t}\sigma^2; s_t
\end{array} \right\} \right] = \\
&= [\mathbf{0}]
\end{aligned}
$$

Now evaluate $\mathrm{H}_{11}F(\overline{\mathbf{x}}, 0; s_t) = [\mathbf{0}]$. We obtain

$$
\begin{aligned}
\mathrm{H}_{11}F(\overline{\mathbf{x}}, \sigma; s_t) = {} & (\mathbf{I}_m \otimes \mathbf{P})'\,\mathbf{f}_{11}\mathbf{P} + 2(\mathbf{I}_m \otimes \mathbf{P})'\,\mathbf{f}_{12}\mathrm{D}_1 g(h(\overline{\mathbf{x}}, \sigma; s_t)) \\
& + 2(\mathbf{I}_m \otimes \mathbf{P})'\,\mathbf{f}_{13} + 2(\mathbf{I}_m \otimes \mathbf{P})'\,\mathbf{f}_{14}\mathbf{F} \\
& + (\mathbf{I}_m \otimes \mathrm{D}_1 g(h(\overline{\mathbf{x}}, \sigma; s_t)))'\,\mathbf{f}_{22}\mathrm{D}_1 g(y) + 2(\mathbf{I}_m \otimes \mathrm{D}_1 g(h(\overline{\mathbf{x}}, \sigma; s_t)))'\,\mathbf{f}_{23} \\
& + 2(\mathbf{I}_m \otimes \mathrm{D}_1 g(h(\overline{\mathbf{x}}, \sigma; s_t)))'\,\mathbf{f}_{24}\mathbf{F} + \mathbf{f}_{33} + 2\mathbf{f}_{34}\mathbf{F} + (\mathbf{I}_m \otimes \mathbf{F})'\,\mathbf{f}_{44}\mathbf{F} \\
& + (\mathbf{f}_1 \otimes \mathbf{I}_{n_x})(\mathbf{G}_{s_t} + \sigma\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}) + (\mathbf{f}_2 \otimes \mathbf{I}_{n_x})\,\mathrm{H}_{11}g(h(\overline{\mathbf{x}}, \sigma; s_t)) + (\mathbf{f}_4 \otimes \mathbf{I}_{n_x})\,\mathbf{E}_{s_t}
\end{aligned}
$$

Now note that

$$
\mathrm{H}_{11}g(h(\overline{\mathbf{x}}, \sigma; s_t)) = \left(\mathbf{I}_{n_y} \otimes \mathbf{P}'\right)\mathbf{E}_{s_t}\mathbf{P} + (\mathbf{F} \otimes \mathbf{I}_{n_x})\,\mathbf{G}_{s_t}
$$

and

$$\mathrm{D}_1 g\left(h\left(\overline{\mathbf{x}},\sigma;s_t\right)\right)=\mathbf{FP}$$

Once these expressions are substituted into $\mathrm{H}_{11}F\left(\overline{\mathbf{x}},\sigma;s_t\right)$ above and the latter is evaluated at $\sigma=0$, we find

$$\begin{aligned}
&\left(\mathbf{f}_1\otimes\mathbf{I}_{n_x}\right)\mathbf{G}_{s_t}+\left(\mathbf{f}_2\otimes\mathbf{I}_{n_x}\right)\left(\left(\mathbf{I}_{n_y}\otimes\mathbf{P}'\right)\mathbf{E}_{s_t}\mathbf{P}+\left(\mathbf{F}\otimes\mathbf{I}_{n_x}\right)\mathbf{G}_{s_t}\right)+\left(\mathbf{f}_4\otimes\mathbf{I}_{n_x}\right)\mathbf{E}_{s_t}+\\
&+\left(\mathbf{I}_m\otimes\mathbf{P}'\right)\mathbf{f}_{11}\mathbf{P}+\left(\mathbf{I}_m\otimes\mathbf{P}'\mathbf{F}'\right)\mathbf{f}_{22}\mathbf{FP}+\mathbf{f}_{33}+\left(\mathbf{I}_m\otimes\mathbf{F}'\right)\mathbf{f}_{44}\mathbf{F}+\\
&+2\left(\mathbf{I}_m\otimes\mathbf{P}'\right)\mathbf{f}_{12}\mathbf{FP}+2\left(\mathbf{I}_m\otimes\mathbf{P}'\right)\mathbf{f}_{13}+2\left(\mathbf{I}_m\otimes\mathbf{P}'\right)\mathbf{f}_{14}\mathbf{F}+\\
&+2\left(\mathbf{I}_m\otimes\mathbf{P}'\mathbf{F}'\right)\mathbf{f}_{23}+2\left(\mathbf{I}_m\otimes\mathbf{P}'\mathbf{F}'\right)\mathbf{f}_{24}\mathbf{F}+2\mathbf{f}_{34}\mathbf{F}\\
&=[\mathbf{0}]
\end{aligned}$$

This is a linear equation which can be solved for $\mathbf{E}_{s_t}$ and $\mathbf{G}_{s_t}$. Note that, as in the case of the first-order approximation, all coefficients in the Hessians $\mathbf{f}_{ij}$ are constant by assumption. Hence, a solution such that the $\mathbf{E}$ and $\mathbf{G}$ matrices are regime independent is a solution of the system. In other words, $\mathbf{E}$ and $\mathbf{G}$ will be identical to the case with homoskedastic shocks.

Now consider the second derivative with respect to σ, namely $\mathrm{H}_{22}\mathbf{F}$. We obtain

$$\begin{aligned}
\mathrm{H}_{22}F=&\left(\mathbf{I}_m\otimes\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}\right)'\mathbf{f}_{11}\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}+2\left(\mathbf{I}_m\otimes\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}\right)'\mathbf{f}_{12}\mathrm{D}_2 g\left(h\left(\overline{\mathbf{x}},\sigma;s_t\right)\right)\\
&+\left(\mathbf{I}_{n_y}\otimes\mathrm{D}_2 g\left(h\left(\overline{\mathbf{x}},\sigma;s_t\right)\right)\right)'\mathbf{f}_{22}\mathrm{D}_2 g\left(h\left(\overline{\mathbf{x}},\sigma;s_t\right)\right)\\
&+\mathbf{f}_1\mathbf{k}_z+\mathbf{f}_2\mathrm{H}_{22}g\left(h\left(\overline{\mathbf{x}},\sigma;s_t\right)\right)+\mathbf{f}_4\mathbf{k}_y
\end{aligned}$$

where

$$\mathrm{D}_2 g\left(h\left(\overline{\mathbf{x}},\sigma;s_t\right)\right)=\mathbf{F}\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}$$

and

$$\mathrm{H}_{22}g\left(h\left(\overline{\mathbf{x}},\sigma;s_t\right)\right)=\left(\mathbf{I}_{n_y}\otimes\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}\right)'\mathbf{E}\left(\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}\right)+\mathbf{k}_y+\mathbf{F}\mathbf{k}_z$$

Substitute the expressions for $\mathrm{D}_2 g$ and $\mathrm{H}_{22}g$ into $\mathrm{H}_{22}F\left(\overline{\mathbf{x}},\sigma;s_t\right)$ and evaluate the result at $\sigma=0$ to find

$$\begin{aligned}
\mathrm{H}_{22}F=&\,\mathbf{f}_1\mathbf{k}_z+\mathbf{f}_2\left(\mathbf{k}_y+\mathbf{F}\mathbf{k}_z+\mathrm{E}_t\left(\mathbf{I}_{n_y}\otimes\mathbf{u}'_{t+1}\boldsymbol{\Sigma}'_{s_t}\right)\mathbf{E}\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}\right)+\mathbf{f}_4\mathbf{k}_y\\
&+\mathrm{E}_t\left(\mathbf{I}_m\otimes\boldsymbol{\Sigma}_{s_t}\tilde{\mathbf{u}}_{t+1}\right)'\mathbf{f}_{11}\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}+2\mathrm{E}_t\left(\mathbf{I}_m\otimes\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}\right)'\mathbf{f}_{12}\mathbf{F}\boldsymbol{\Sigma}_{s_t}\tilde{\mathbf{u}}_{t+1}\\
&+\mathrm{E}_t\left(\mathbf{I}_{n_y}\otimes\mathbf{u}'_{t+1}\boldsymbol{\Sigma}'_{s_t}\mathbf{F}'\right)\mathbf{f}_{22}\mathbf{F}\boldsymbol{\Sigma}_{s_t}\mathbf{u}_{t+1}
\end{aligned}$$

Taking expectations, the elements of vectors $\mathbf{k}_y$ and $\mathbf{k}_z$ are the solution of the following linear equations

$$\begin{aligned}
&\mathbf{f}_1\mathbf{k}_z+\mathbf{f}_2\mathbf{k}_y+\mathbf{f}_2\mathbf{F}\mathbf{k}_z+\mathbf{f}_2\mathrm{trm}\left(\mathbf{I}_{n_y}\otimes\left(\boldsymbol{\Sigma}'_{s_t}\boldsymbol{\Sigma}_{s_t}\right)\right)\mathbf{E}+\mathbf{f}_4\mathbf{k}_y\\
&+\mathrm{trm}\left(\mathbf{I}_m\otimes\left(\boldsymbol{\Sigma}'_{s_t}\boldsymbol{\Sigma}_{s_t}\right)\right)\mathbf{f}_{11}+2\mathrm{trm}\left(\mathbf{I}_m\otimes\left(\boldsymbol{\Sigma}'_{s_t}\boldsymbol{\Sigma}_{s_t}\right)\right)\mathbf{f}_{12}\mathbf{F}\\
&+\mathrm{trm}\left(\mathbf{I}_{n_y}\otimes\left(\boldsymbol{\Sigma}'_{s_t}\boldsymbol{\Sigma}_{s_t}\mathbf{F}'\right)\right)\mathbf{f}_{22}\mathbf{F}=[\mathbf{0}]
\end{aligned}\tag{6}$$

where $\boldsymbol{\Sigma}'_{s_t}\boldsymbol{\Sigma}_{s_t}$ is the conditional variance covariance matrix of vector $\mathbf{x}_t$ and where, as in Gomme and Klein (2006), we define the matrix trace (trm) of an $(nm \times n)$ matrix

$$\begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_m \end{bmatrix}$$

as the $m \times 1$ vector

$$\begin{bmatrix} \mathrm{tr}\left(\mathbf{M}_1\right) \\ \mathrm{tr}\left(\mathbf{M}_2\right) \\ \vdots \\ \mathrm{tr}\left(\mathbf{M}_m\right) \end{bmatrix}$$

Due to the presence of the regime-switching terms, the variance-covariance matrix $\boldsymbol{\Sigma}'_{s_t}\boldsymbol{\Sigma}_{s_t}$ is regime dependent. Equation (6) can therefore be satisfied only if the coefficients $\mathbf{k}_{y,s_t}$, $\mathbf{k}_{x,s_t}$ assume different values depending on the realisation of the regime-switching states. If there are n_s possible regimes, n_s systems of the form (6) have to be solved, yielding n_s pairs $(\mathbf{k}_{y,s_t}, \mathbf{k}_{z,s_t})$.

Finally, along the lines above, one can confirm that the coefficients of terms in $\sigma\widehat{\mathbf{x}}_t$ are zero solving the equation $\mathrm{H}_{12}F = [\mathbf{0}]$.

To summarise, the second -order approximation of the policy functions $g\left(\widehat{\mathbf{x}}_t, \sigma; s_t\right)$ and $h\left(\widehat{\mathbf{x}}_t, \sigma; s_t\right)$ can be written as

$$g\left(\widehat{\mathbf{x}}_t, \sigma; s_t\right) = \overline{\mathbf{y}} + \mathbf{F}\widehat{\mathbf{x}}_t + \frac{1}{2}\left(\mathbf{I}_{n_y} \otimes \widehat{\mathbf{x}}'_t\right)\mathbf{E}\widehat{\mathbf{x}}_t + \mathbf{k}_{y,s_t}\sigma^2 \quad (7)$$

$$h\left(\widehat{\mathbf{x}}_t, \sigma; s_t\right) = \overline{\mathbf{x}} + \mathbf{P}\widehat{\mathbf{x}}_t + \frac{1}{2}\left(\mathbf{I}_{n_x} \otimes \widehat{\mathbf{x}}'_t\right)\mathbf{G}\widehat{\mathbf{x}}_t + \mathbf{k}_{x,s_t}\sigma^2 \quad (8)$$

Compared to the case with homoskedastic shocks, the only difference in the case of regime-switching is in the $\mathbf{k}_{x,s_t}$ and $\mathbf{k}_{y,s_t}$ vectors. A change in regime only affects the policy rule through a precautionary savings or risk premium component. To a first -order approximation, the policy rules are exactly as in the case with homoskedastic shocks.

4 Estimation

In this section we demonstrate how to get the likelihood from the reduced form obtained by solving, up to second-order terms, the DSGE model with Markov switching variances. In order to better explain the way in which we conduct likelihood inference, let us simplify a bit the notation and rewrite (7) and (8) as

$$\mathbf{y}_{t+1} = \mathbf{c}_j + \mathbf{C}_1\mathbf{x}_{t+1} + \mathbf{C}_2\overline{vech}(\mathbf{x}_{t+1}\mathbf{x}'_{t+1}) + \mathbf{D}\mathbf{v}_{t+1} \tag{9}$$

$$\mathbf{x}_{t+1} = \mathbf{a}_i + \mathbf{A}_1\mathbf{x}_t + \mathbf{A}_2\overline{vech}(\mathbf{x}_t\mathbf{x}'_t) + \mathbf{B}_i\mathbf{w}_{t+1} \tag{10}$$

$$\mathbf{s}_t \backsim \text{Markov switching} \tag{11}$$

where $\mathbf{x}_{t+1}$ indicates directly deviations of the continuous variables with respect to their steady state values. In addition, the vector $\mathbf{y}_t^o$ includes all observable variables, and $\mathbf{v}_{t+1}$ and $\mathbf{w}_{t+1}$ are measurement and structural shocks, respectively. In this representation, the regime switching variables affect the system by changing the intercepts $\mathbf{a}_i$ and $\mathbf{c}_j$, and the loadings of the structural innovations $\mathbf{B}_i$ (we indicate here with i the value of the discrete state variables at t and with j the value of the discrete state variables at $t+1$).

If a linear approximation were used, we would be left with a linear state space model with Markov switching affecting some of the parameters (see Kim, 1994; Kim and Nelson, 1999; and Schorfheide, 2005, for a DSGE application). The likelihood cannot be obtained by recursive methods and it is approximated using a discrete mixture approach.

Things are easier when the number of continuous shocks (measurement and structural) is equal to the number of observables. In such a case the continuous latent variables can be obtained via inversion and the system can be written as a Markov Switching VAR. The likelihood can be obtained using Hamilton's filter, i.e. integrating out the discrete latent variables. See Hamilton (1994).

In the quadratic case, the likelihood cannot in general be obtained in closed form. One possible approach to compute the likelihood is to rely on sequential Monte Carlo techniques (for an application of these techniques in a DSGE setting, see e.g. Amisano and Tristani, 2010). These methods, however, are computationally expensive in a case, such as the one of our model, in which both non-linearities and non-Gaussianity of the shocks characterise the economy. It is in fact well known that the particle filter can be quite inefficient, especially in the presence of abnormal observations (see Pitt and Shephard, 1999, 2001), such as those implicated by shocks being drawn from mixture of distributions with different variances. In these circumstances, the number of particles to be used in order to provide a basis for accurate likelihood evaluations might be so high to render estimation unfeasible.

We thus adopt a simple extension of the filter employed in the linear case when the number of observed variables is equal to the sum of measurement and structural shocks. It is worth emphasising that our preferred method might also be used as a benchmark capable of producing exact likelihood evaluations, in order to assess the precision of likelihood evaluations obtained by using sequential Monte Carlo methods like the particle filter.

In order to compute the likelihood function, the main problem is the quadratic term in $\mathbf{x}_t$ in the observation equation (9). In particular, given that we assume that $\mathbf{x}_{2t}$, the vector of predetermined variables, contains only lagged endogenous variables, the problem in computing the likelihood is generated by the fact that $\mathbf{x}_{2t}$ enters the reduced form state space representation in a nonlinear way.

We are interested in a particular case which renders likelihood computations quite easy. Let us define

$$\mathbf{z}_{t+1} = \begin{bmatrix} \mathbf{x}_{2t+1} \\ \mathbf{v}_{t+1} \end{bmatrix}$$

the $(n_{x_2} + n_{me}) \times 1$ vector containing all continuous latent variables of the models which are not predetermined. When $n_z = n_{x_2} + n_{me} = n_y$, i.e. when there are as many observables as continuous latent variables in the system, the mapping

$$\mathbf{z}_{t+1}^{(j,:)} = \zeta(\mathbf{y}_{t+1}|s_{t+1} = j, \mathbf{x}_{1t+1}) = \zeta_j(\mathbf{y}_{t+1}) \tag{12}$$

has $K = 2^{n_{x_2}}$, possibly complex, solutions. In fact, we can see that the reduced form state space representation describes a quadratic system of $(n_y - n_{x_2})$ equations for $\mathbf{x}_{2t+1}$ and n_{me} linear equations in $\mathbf{v}_{t+1}$. We limit our attention to real valued solutions, given that the domain of $\mathbf{z}_{t+1}$ is real, and we call $K_{jt+1} \leq K$ the number of real roots in (12). $\mathbf{z}_{t+1}^{(j,k)}$ is the k^{th} solution of (12). We use an arbitrary ordering of the solutions to uniquely identify them

$$\mathbf{z}_{1t+1}^{(j,1)} \leq \mathbf{z}_{1t+1}^{(j,2)} \leq ... \leq \mathbf{z}_{1t+1}^{(j,K_{jt+1})}$$

and define as $d_{k,t+1}$ the event associated to $\mathbf{z}_{t+1} = \mathbf{z}_{t+1}^{(jk)}$.

Finding the solutions of this mapping entails filtering $\mathbf{z}_{t+1}$ out. When the number of continuous exogenous state variables $n_{x_2} = 1$ or 2, it is very easy to find all roots. With higher dimensional problems, other methods can be used. As an example, we can use the the polynomial homotopy continuation (PHC) method (and its Matlab interface [1]) by Verschelde (1999). More details on the methodology can be found in Morgan (1987) and Judd (1998).

[1] See http://www.math.uic.edu/~jan/. The Matlab interface is documented in Guan and Verschelde (2008).

With the solutions (12) in hand, it is possible to obtain the likelihood. Clearly, before observing $\mathbf{y}_{t+1}$, which ties down the set of admissible solutions, each of the roots is equally likely, i.e. we can write

$$p(d_{j,t+1}|s_{t+1} = j, s_t = i, d_{it}, \underline{\mathbf{y}}_t) = \frac{1}{K_{j,t+1}}$$

Conditionally on the roots obtained in the previous, time t, step, however, the roots computed at $t+1$ have different likelihood.

In order to explain the intuition for the algorithm, we illustrate it in the following subsections for the case of a simplified model in reduced form, i.e. a univariate model with $n_y = n_{x_2} = 1, n_{x1} = n_{me} = 0$. For such a model, we first discuss the case in which shocks are homoskedastic and all the coefficients of the reduced form are constant, then move on to the case in which the variance of the shock can switch amongst different regimes. Finally, we generalise the derivations to the multivariate case with regime-switches.

4.1 Univariate, homoskedastic example

We start from the following quadratic model with Gaussian shocks

$$\begin{aligned} y_{t+1} &= c_0 + c_1 z_{t+1} + c_2 z_{t+1}^2 \\ z_{t+1} &= a_1 z_t + b w_{t+1} \end{aligned}$$

in which y_{t+1}, z_{t+1} and w_{t+1} are univariate real processes, w_{t+1} is i.i.d. with known pdf, and c_0, c_1, c_2, a_1, b_1 are real valued coefficients and only the time series $\underline{\mathbf{y}}_t = \{y_\tau, \tau = 1, 2, ..., t\}$ is observable. In this case the solutions in terms of z_t are easily found as

$$z_t = \frac{-c_1 \pm \sqrt{c_1^2 - 4c_2(c_0 - y_t)}}{2c_2}, t = 1, 2, ..., T \tag{13}$$

and at each point in time there are either $K_t = 2$ or $K_t = 0$ real roots (the event of having two coinciding roots has zero probability). Using Bayes' theorem we can update the probabilities of each root as

$$\begin{aligned} p(d_{t+1} = k|\underline{\mathbf{y}}_{t+1}) &= \frac{p(d_{t+1} = k|\underline{\mathbf{y}}_t) \times p(y_{t+1}|d_{t+1} = k, \underline{\mathbf{y}}_t)}{p(y_{t+1}|\underline{\mathbf{y}}_t)} \\ &\propto p(y^o_{t+1}|d_{t+1} = k, \underline{\mathbf{y}}_t) = \\ &= \sum_{h=1}^{K_t} p(z^{(k)}_{t+1}|z^{(h)}_t) \times \left\| \frac{\partial z^{(k)}_{t+1}}{\partial y_{t+1}} \right\| \times p(d_t = h|\underline{\mathbf{y}}_t) \\ &\propto \sum_{h=1}^{K_t} p(z^{(k)}_{t+1}|z^{(h)}_t) \times p(d_t = h|\underline{\mathbf{y}}^o_t), \end{aligned} \tag{14}$$

We can move from line 1 to line 2 of equation (14) because before observing y_{t+1} all roots are equiprobable; the third line arises by writing the probability as a function of the roots at time t and by using the variable transformation rule for probabilities. Line 4 is obtained by noting that

$$\left\| \frac{\partial z^{(k)}_{t+1}}{\partial y_{t+1}} \right\| = \left| c_1^2 - 4c_2(c_0 - y_t) \right|^{-1/2}, k = 1, 2. \tag{15}$$

i.e. the Jacobian of the transformation is constant across both roots. Equation (14) offers a way to dynamically update the probabilities associated with individual roots at each point in time in a filtering fashion. Note that the update of the probabilities is sequential insofar the process for z_t is persistent, i.e. $a_1 \neq 0$. When $a_1 = 0$, then $p(z^{(k)}_{t+1}|z^{(h)}_t) = p(z^{(k)}_{t+1})$ and expression (14) simplifies as

$$p(d_{t+1} = k|\underline{\mathbf{y}}_{t+1}) \propto p(z^{(k)}_{t+1}) \tag{16}$$

Recursion (14) gives also the key to obtain the likelihood of each observation as

$$p(y_{t+1}|\underline{\mathbf{y}}_t) = \left\| \frac{\partial z_{t+1}}{\partial y_{t+1}} \right\| \times \sum_{k=1}^{K_{t+1}} \sum_{h=1}^{K_t} p(z^{(k)}_{t+1}|z^{(h)}_t) \times p(d_t = h|\underline{\mathbf{y}}_t). \tag{17}$$

The recursion is initialised with $p(z_1 = z_1^{(h)}|y_1^o) = \frac{1}{K_1}$.

Note that when at least one observation is associated with no real roots, i.e. when $c_1^2 - 4c_2(c_0 - y_t) < 0$, then the conditional likelihood of that observation is set to zero and so is the likelihood over the entire sample. This amounts to assuming that a given point in the parameter space is not capable of delivering the observed data with a positive probability.

4.2 Univariate example with Markov switching

We now generalise the model to the case in which the variance of the shocks change with the prevailing regime. As illustrated in section 3, this implies that the constant in the policy functions will also be subject to regime switches. The simplified, reduced-form model is therefore

$$y_{t+1} = c_{0,s_{t+1}} + c_1 z_{t+1} + c_2 z_{t+1}^2 \tag{18}$$
$$z_{t+1} = a_{s_t} + a_1 z_t + b_{s_t} w_{t+1} \tag{19}$$

where $s_{t+1} = j$ denotes the value assumed by a discrete Markov switching state variable at time $t+1$:

$$p(s_{t+1} = j|s_{t-1} = i, \underline{\mathbf{s}}_{t-2}, \underline{\mathbf{y}}_{t-1}) = p(s_{t+1} = j|s_{t-1} = i) = p_{ij}, i = 1, 2, .., m, j = 1, 2, ..., \tag{20}$$

Here the problem of multiplicity of roots is exacerbated by the fact that there are potentially 2 roots for each discrete state s_t

$$z_t^{(i,:)} = \frac{-c_1 \pm \sqrt{c_1^2 - 4c_2(c_{0,i} - y_t)}}{2c_2}, t = 1, 2, ..., T$$

In the case of a two state process ($m = 2$), we have up to 4 potential roots for each observation. Hence, we have to take into account the unobservable nature of the Markov switching process and recursion (14) has to be modified as follows:

$$\begin{aligned} p(z_{t+1} = z_t^{(jk)}, s_{t+1} = j|\underline{\mathbf{y}}_{t+1}) &\propto \sum_{i=1}^{m}\sum_{h=1}^{K_{it}} p(z_{t+1}^{(jk)}|z_t^{(ih)}) \times \left\| \frac{\partial z_{t+1}^{(jk)}}{\partial y_{t+1}} \right\| \times \\ \times p_{ij} \times p(z_t = z_t^{(ih)}, s_t = i|\underline{\mathbf{y}}_t) \\ &\propto \sum_{i=1}^{m}\sum_{h=1}^{K_{it}} p(z_{t+1}^{(jk)}|z_t^{(ih)}) \times p_{ij} \times p(z_t = z_t^{(ih)}, s_t = i|\underline{\mathbf{y}}_t) \end{aligned} \tag{21}$$

and the conditional likelihood of each observation is obtained as

$$p(y_{t+1}|\underline{\mathbf{y}}_t) = \sum_{j=1}^{m} \left\| \frac{\partial z_{t+1}^{(j,:)}}{\partial y_{t+1}} \right\| \times \sum_{k=1}^{K_{jt+1}} \sum_{i=1}^{m}\sum_{h=1}^{K_{it}} p(z_{t+1}^{(jk)}|z_t^{(ih)}) \times p_{ij} \times p(z_t = z_t^{(ih)}, s_t = i|\underline{\mathbf{y}}_t) \tag{22}$$

The recursion is initialised using

$$p(z_1 = z_1^{(i,h)}, s_1 = i) = \pi_i \times \frac{1}{K_1}, i = 1, 2.., m \tag{23}$$

where π_i are the ergodic probabilities associated to the Markov switching process.

Note that here, in order to have a zero value for the likelihood it is necessary that, for at least one observation t, the number of real roots be zero for all states, i.e.

$$K_{i,t} = 0, i = 1, 2, .., m. \tag{24}$$

When we have $K_{i,t} = 0$ only for some state i, then the probability of $s_t = i$ in the recursion is zero.

Note also that the probabilities obtained by the recursion (21) can be marginalised with respect to the roots, in order to obtain filtered probabilities of the discrete states

$$p(s_{t+1} = j|\underline{\mathbf{y}}_{t+1}) = \sum_{j=1}^{m} p(z_{t+1} = z_t^{(jk)}, s_{t+1} = j|\underline{\mathbf{y}}_{t+1}) \tag{25}$$

These probabilities can then be used to obtain smoothed probabilities or smoothed simulations of the discrete states. Smoothed and filtered distributions of the continuous latent variable can also be obtained very easily. This is another advantage of using our procedure with respect to using a sequential Monte Carlo approach, where to obtain smoothed distributions of the latent variables can be computationally very involved. In this regard, see Fernandez-Villaverde and Rubio-Ramirez (2007).

4.2.1 A numerical example

In this section, we consider a simple data generation mechanism of the kind (18)-(19) with numerical values of the parameters set as in Table (1).

We use these parameter values to generate $T = 200$ data points. The generated series are represented in the upper left panel of Figure (1). In order to empirically analyse the properties of our proposed inferential procedure, we use it to estimate the parameter values from the generated series. The results are reassuring: the resulting recursive probabilities give rise to sensible results. In fact, the filtered probabilities of the discrete state s_t obtained applying equation (25), which can be seen in the upper left panel of Figure (1), are in very close accordance with the true values of the s_t process (the sample correlation is 99.7%). In the lower panel of Figure (1), we report the probabilities assigned by our procedure to the true root of the map from y_t to z_t. With few exceptions including the initial observations, the model assigns high probabilities to the true values which have generated the observed series, even if substantial uncertainty remains for many observations. This reflects the sporadic relevance of the roots multiplicity structure that we have in the quadratic model under study.

4.3 The quadratic MS-DSGE model

Turning now to the MS-quadratic DSGE model (9) and (10), inference is only mildly more complicated with respect to the univariate MS case by the fact that the model is multivariate and that it involves some predetermined, yet observable state variables $\mathbf{x}_{1t}$.

These are just marginal complications that can be easily accounted for by writing the model as follows

$$\mathbf{y}_{t+1} = \mathbf{c}_j + \mathbf{C}_1\mathbf{x}_{t+1} + \mathbf{C}_2\overline{vech}(\mathbf{x}_{t+1}\mathbf{x}'_{t+1}) + \mathbf{D}\mathbf{v}_{t+1} \tag{26}$$

$$\mathbf{x}_{1t+1} = \mathbf{R}_y(\mathbf{y}_t - \overline{\mathbf{y}}) \tag{27}$$

$$\mathbf{z}_{t+1} = \widetilde{\mathbf{a}}_i + \widetilde{\mathbf{A}}_i\mathbf{z}_t + \widetilde{\mathbf{B}}_i\widetilde{\mathbf{w}}_{t+1} \tag{28}$$

$$\mathbf{z}_{t+1} = \begin{bmatrix} \mathbf{z}_{2t+1} \\ \mathbf{v}_{t+1} \end{bmatrix}, \widetilde{\mathbf{a}}_i = \begin{bmatrix} \mathbf{a}_i \\ \mathbf{0} \end{bmatrix}, \widetilde{\mathbf{A}}_i = \begin{bmatrix} \mathbf{A}_{1,22} & [\mathbf{0}] \\ [\mathbf{0}] & [\mathbf{0}] \end{bmatrix} \tag{29}$$

$$\widetilde{\mathbf{w}}_{t+1} = \begin{bmatrix} \mathbf{w}_{t+1} \\ \mathbf{v}_{t+1} \end{bmatrix} \backsim N\left(\mathbf{0}, \mathbf{I}_{n_y}\right), \widetilde{\mathbf{B}}_i = \begin{bmatrix} \mathbf{B}_{2,i} & [\mathbf{0}] \\ [\mathbf{0}] & \mathbf{D} \end{bmatrix} \tag{30}$$

and we know that

$$p(\mathbf{z}_{t+1}^{(j,k)}|\mathbf{z}_t^{(ih)}) = N(\mathbf{z}_{t+1}^{(jk)}, \widetilde{\mathbf{a}}_i + \widetilde{\mathbf{A}}_i\mathbf{z}_t^{(i,h)}, \mathbf{B}_i\mathbf{B}_i') \tag{31}$$

Define K_{jt+1} the number of real solutions in $\mathbf{z}_{t+1}$ of equation (26), conditional on $s_{t+1} = j$, identified imposing the conventional ordering

$$z_{1t+1}^{(j,1)} \leq z_{1t+1}^{(j,2)} ... \leq z_{1t}^{(j,K_{jt+1})}$$

We can therefore resort to the multivariate extension of recursion (21) to update the probabilities of the roots and of the discrete states

$$p(d_{t+1} = k, s_{t+1} = j|\underline{\mathbf{y}}_{t+1}) \propto \sum_{h=1}^{K}\sum_{i=1}^{m^*} p(\mathbf{z}_{t+1}^{(jk)}|\mathbf{z}_t^{(ih)}) \times \left\| \frac{\partial \mathbf{z}_{t+1}^{(jk)}}{\partial \mathbf{y}_{t+1}'} \right\| \times p_{ij} \times p(d_t = h, s_t = i|\underline{\mathbf{y}}_t) \tag{32}$$

with

$$\left\|\frac{\partial \mathbf{z}_{t+1}^{(k)}}{\partial \mathbf{y}_{t+1}^{'}}\right\|_{ij} = \left\|\begin{bmatrix} \mathbf{C}_1 & \mathbf{D} \end{bmatrix} + \mathbf{C}_2\mathbf{T}\left[\left(\mathbf{I}_{n_y} \otimes \mathbf{x}_{t+1}\right) + \left(\mathbf{x}_{t+1} \otimes \mathbf{I}_{n_y}\right)\right]\mathbf{U}_2\right\|^{-1}$$

$$\mathbf{U}_2 = \frac{\partial \mathbf{x}_{t+1}}{\partial \mathbf{z}_{t+1}^{'}} = \begin{bmatrix} \underset{(n_{x_1} \times n_{x_2})}{[\mathbf{0}]} & \underset{(n_{x_1} \times n_{me})}{[\mathbf{0}]} \\ \mathbf{I}_{n_{x2}} & \underset{(n_{x2} \times n_{me})}{[\mathbf{0}]} \end{bmatrix}$$

and $\mathbf{T}$ is defined as

$$\mathbf{T} \times vec(\mathbf{v}) = \overline{vech}(\mathbf{v})$$

The conditional likelihood of each observation is

$$p(\mathbf{y}_{t+1}|\underline{\mathbf{y}}_t) = \sum_{k=1}^{K}\sum_{j=1}^{m^*}\sum_{h=1}^{K}\sum_{i=1}^{m^*} p(\mathbf{z}_{t+1}^{jk}|\mathbf{z}_t^{ih}) \times \left\|\frac{\partial \mathbf{z}_{t+1}^{jk}}{\partial \mathbf{y}_{t+1}^{'}}\right\| \times p_{ij} \times p(d_t = h, s_t = i|\underline{\mathbf{y}}_t) \quad (33)$$

Equation (33) describes a filtering recursion over a discrete set of realisations which can be marginalised to obtain marginal filtered probabilities of discrete states ($\mathbf{s}_{t+1}$) and continuous latent variables ($\mathbf{z}_{t+1}$).

An interesting question is how we handle complex roots. If some roots at time $t+1$ for some given discrete state j are complex, these will be assigned zero probability, since the support of the latent variables is the real line. If at time $t+1$ for some j all roots are complex, then the algorithm will work anyway and will assign probability zero to state j conditioned at time $t+1$. If at time $t+1$ all roots are complex for all j, then the likelihood of that observation is set to zero.

From the computational point of view, the Jacobians $\left\|\frac{\partial \mathbf{z}_{t+1}^{(jk)}}{\partial \mathbf{y}_{t+1}^{'}}\right\|$ are readily computable analytically. In addition, it is important to bear in mind that the only computationally expensive part of the algorithm is to obtain the roots $\mathbf{z}_{jt+1}$. The problem can be fully parallelised, since at each $t+1$ and j the roots computations are functions only of $\mathbf{y}_{t+1}$ and of the coefficients of the reduced form measurement equations.

As already mentioned in the introduction, the main limitation of the inference procedure described so far is that it cannot be applied when latent variables enter the vector $\mathbf{x}_{1t}$ of non-stochastic, predetermined variables. In this case, the solution for variables $\mathbf{z}_t$ at each point in time would also be conditional on the value of the latent variables in $\mathbf{x}_{1t}$ – see equation (12). In turn, filtered values of the latent variables in $\mathbf{x}_{2t+1}$ would have to solve quadratic equations conditional on the values assumed by variables $\mathbf{z}_t$. This interdependence between $\mathbf{z}_t$ and $\mathbf{x}_{2t+1}$ generates time dependence in the solutions of the quadratic equations. As a result, the number of solutions for $\mathbf{z}_t$ to compute at each point in time would increase exponentially at a factor 2^{n_L}, where n_L denotes the number of latent, non-stochastic, predetermined variables.

Provided that there are no unobservable predetermined state variables in the system (and that $n_{x_2} + n_{me} = n_y$), however, there is no conceptual problem in extending the algorithm to contexts in which the DSGE model is solved to third order.

5 An application to a simple model

In order to highlight the marginal contribution of heteroskedasticity, we rely on a standard model in the spirit of Woodford (2003). The central feature is the assumption of nominal rigidities and the presence of discrete shifts in conditional volatility of the shocks. The model is kept deliberately simple in order to have a very low number of continuous unobserved non-predetermined state variables. In Amisano and Tristani (2011), a quadratic model with heteroskedastic shocks is used in a context where the vector of observable variables includes term structure data.

5.1 Households

We assume that each household i provides $N(i)$ hours of differentiated labour services to firms in exchange for a labour income $w_t(i) N_t(i)$. Each household owns an equal share of all firms j and receives profits $\int_0^1 \Pi_t(j) \mathrm{d}j$.

As in Erceg, Henderson and Levin (2000), an employment agency combines households' labour hours in the same proportions as firms would choose. The agency's demand for each household's labour is therefore equal to the sum of firms' demands. The labour index L_t has the Dixit-Stiglitz form

$$L_t = \left[\int_0^1 N_t(i)^{\frac{\theta_w - 1}{\theta_w}} \mathrm{d}i \right]^{\frac{\theta_w}{\theta_w - 1}}$$

where $\theta_w > 1$. At time t, the employment minimizes the cost of producing a given amount of the aggregate labour index, taking each household's wage rate $w_t(i)$ as given and then sells units of the labour index to the production sector at the aggregate wage index $w_t = \left[\int_0^1 w(i)^{1-\theta_w} \mathrm{d}i \right]^{\frac{1}{1-\theta_w}}$. The employment agency's demand for the labour hours of household i is given by

$$N_t(i) = L_t \left(\frac{w_t(i)}{w_t} \right)^{-\theta_w} \tag{34}$$

Each household i maximizes its intertemporal utility with respect to consumption, the wage rate and holdings of contingent claims, subject to its labour demand function (34) and the budget constraint

$$P_t C_t(i) + \mathrm{E}_t Q_{t,t+1} W_{t+1}(i) \leq W_t(i) + w_t(i) N_t(i) + \int_0^1 \Xi_t(j) \, \mathrm{d}j \tag{35}$$

where C_t is a consumption index satisfying

$$C_t = \left(\int_0^1 C_t(z)^{\frac{\theta-1}{\theta}} \, \mathrm{d}z \right)^{\frac{\theta}{\theta-1}} \tag{36}$$

the price level P_t is defined as the minimal cost of buying one unit of C_t, hence equal to

$$P_t = \left(\int_0^1 p(z)^{1-\theta} \, \mathrm{d}z \right)^{\frac{1}{1-\theta}}. \tag{37}$$

W_t denotes the beginning-of-period value of a complete portfolio of state contingent assets, $Q_{t,t+1}$ is their price, $w_t(i)$ is the nominal wage rate and $\Xi_t(j)$ are the profits received from investment in firm j.

Equation (35) states that each household can only consume or hold assets for amounts that must be less than or equal to its salary, the profits received from holding equity in all the existing firms and the revenues from holding a portfolio of state-contingent assets.

Households maximise the discounted sum of the period utility

$$u_t = \left[\left(C_t(i) - hC_{t-1}(i) \right) \left(\overline{N} - N_t^{\phi}(i) \right) \right]^{1-\gamma} \tag{38}$$

subject to the budget constraint (35)

$$P_t C_t(i) + \mathrm{E}_t Q_{t,t+1} W_{t+1}(i) \leq W_t(i) + w_t(i) N_t(i) + \int_0^1 \Xi_t(j) \, \mathrm{d}j$$

and

$$N_t(i) = L_t \left(\frac{w_t(i)}{w_t} \right)^{-\theta_w}$$

where the choice variables are $w_t(i)$ and $C_t(i)$.

The first -order conditions for this problem can be written as

$$\widetilde{w}_t = \phi \mu_w N_t^{\phi-1} \frac{(C_t - hC_{t-1})^{1-\gamma} \left(\overline{N} - N_t^{\phi}\right)^{-\gamma}}{\widetilde{\Lambda}_t} \tag{39}$$

$$Q_{t,t+1} = \beta \frac{\widetilde{\Lambda}_{t+1}}{\widetilde{\Lambda}_t} \frac{1}{\pi_{t+1}} \tag{40}$$

$$\widetilde{\Lambda}_t = (C_t - hC_{t-1})^{-\gamma} \left(\overline{N} - N_t^{\phi}\right)^{1-\gamma} - \beta h \mathrm{E}_t \left(C_{t+1} - hC_t\right)^{-\gamma} \left(\overline{N} - N_{t+1}^{\phi}\right)^{1-\gamma} \tag{41}$$

where $\widetilde{w}_t \equiv w_t / P_t$ and $\mu_w \equiv \theta_w / (\theta_w - 1)$.

The gross interest rate, I_t, equals the conditional expectation of the stochastic discount factor, i.e.

$$I_t^{-1} = \mathrm{E}_t Q_{t,t+1} \tag{42}$$

Note that we will focus on a symmetric equilibrium in which nominal wage rates are all allowed to change optimally at each point in time, so that individual nominal wages will equal the average w_t.

5.2 Firms

We assume a continuum of monopolistically competitive firms (indexed on the unit interval by j), each of which produces a differentiated good. Demand arises from households' consumption and from government purchases G_t, which is an aggregate of differentiated goods of the same form as households' consumption. It follows that total demand for the output of firm i takes the form $Y_t(j) = \left(\frac{P_t(i)}{P_t}\right)^{-\theta} Y_t$. Y_t is an index of aggregate demand which satisfies $Y_t = C_t + G_t$.

Firms have the production function

$$Y_t(j) = A_t L_t^{\alpha}(j)$$

where L_t is the labour index L_t defined above.

Once aggregate demand is realised, the firm demands the labour necessary to satisfy it

$$L_t(j) = \left(\frac{Y_t(j)}{A_t}\right)^{\frac{1}{\alpha}}$$

The total nominal cost function for the firm will therefore be given by

$$TC_t(j) = w_t \left(\frac{Y_t(j)}{A_t}\right)^{\frac{1}{\alpha}}$$

where w_t is the wage index defined above. As a result, real marginal costs will be

$$mc_t(j) = \frac{1}{\alpha}\frac{w_t}{P_t}\frac{1}{A_t}\left(\frac{Y_t(j)}{A_t}\right)^{\frac{1-\alpha}{\alpha}}$$

where nominal costs are deflated using the aggregate price level (not the individual firm's price).

As in Rotemberg (1982), we assume the firms face quadratic costs in adjusting their prices. This assumption is also adopted, for example, by Schmitt-Grohé and Uribe (2004) and Ireland (1997). It is well-known to yield first-order inflation dynamics equivalent to those arising from the assumption of Calvo pricing.[2] From our viewpoint, it has the advantage of greater computational simplicity, as it allows us to avoid having to include an additional state variable in the model, i.e. the cross-sectional dispersion of prices across firms.

The specific assumption we adopt is that firm j faces a quadratic cost when changing its prices in period t, compared to period $t-1$. Consistently with what is typically done in the Calvo literature, we modify the original Rotemberg (1982) formulation to allow for indexation of prices in part to lagged inflation, in part to the inflation objective

$$\frac{\zeta}{2}\left(\frac{P_t^j}{P_{t-1}^j} - (\Pi^*)^{1-\iota}\,\Pi_{t-1}^{\iota}\right)^2 Y_t$$

So, firms maximise their real profits

$$\max_{P_t^j} \mathrm{E}_t \sum_{s=t}^{\infty} Q_{t,s}\left[\frac{P_s^j Y_s^j(P_s^j)}{P_s} - \frac{TC_s(Y_s^j(P_s^j))}{P_s} - \frac{\zeta}{2}\left(\frac{P_s^j}{P_{s-1}^j} - (\Pi^*)^{1-\iota}\,\Pi_{s-1}^{\iota}\right)^2 Y_s\right]$$

subject to

$$Y_t(j) = \left(\frac{P_t(j)}{P_t}\right)^{-\theta} Y_t$$

and to

$$Y_t(j) = A_t L_t^{\alpha}(j)$$

[2] The two pricing models, however, have in general different welfare implications – see Lombardo and Vestin (2008).

Focusing on a symmetric equilibrium in which all firms adjust their price at the same time, the first -order condition for price setting can be written as

$$(\theta - 1) Y_t + \zeta \left(\Pi_t - (\Pi^*)^{1-\iota} \Pi_{t-1}^{\iota} \right) Y_t \Pi_t = \frac{\theta}{\alpha} \widetilde{w}_t \left(\frac{Y_t}{A_t} \right)^{\frac{1}{\alpha}} + \\ + \mathrm{E}_t Q_{t,t+1} \zeta \left(\Pi_{t+1} - (\Pi^*)^{1-\iota} \Pi_t^{\iota} \right) Y_{t+1} \Pi_{t+1}$$

5.3 Monetary policy

We close the model with the simple Taylor-type policy rule

$$I_t = \left(\frac{\Pi^*}{\beta} \right)^{1-\rho_I} \left(\frac{\Pi_t}{\Pi^*} \right)^{\psi_\Pi} \left(\frac{Y_t}{Y} \right)^{\psi_Y} I_{t-1}^{\rho_I} \tag{43}$$

where Y_t is aggregate output and Π^* is a constant inflation target.

5.4 Market clearing

Market clearing in the goods market requires

$$Y_t = C_t + G_t$$

In the labour market, labour demand will have to equal labour supply. In addition, the total demand for hours worked in the economy must equal the sum of the hours worked by all individuals. Taking into account that at any point in time the nominal wage rate is identical across all labour markets because all wages are allowed to change optimally, individual wages will equal the average w_t. As a result, all households will chose to supply the same amount of labour and labour market equilibrium will require that

$$L_t = \left(\frac{Y_t}{A_t} \right)^{\frac{1}{\alpha}}$$

5.5 Exogenous shocks

In macroeconomic applications, exogenous shocks are almost always assumed to be (log) normal, partly because models are typically log-linearised and researchers are mainly interested in characterising conditional means. However, Hamilton (2008) argues that a correct modelling of conditional variances is always necessary, for example because inference on conditional means can be inappropriately influenced by outliers and high-variance episodes. The need for an appropriate treatment of heteroskedasticity becomes even more compelling when models are solved nonlinearly, because conditional variances have a direct impact on conditional means.

In this paper, we assume that variances are subject to stochastic regime switches for all shocks other than the inflation target. More specifically, we assume a deterministic trend in technology growth, so that

$$
\begin{aligned}
A_t &= Z_t B_t \\
B_t &= B_{t-1}\Xi \\
Z_t &= Z_{t-1}^{\rho_z} e^{\varepsilon_t^z}, \qquad \varepsilon_{t+1}^z \backsim N\left(0, \sigma_{z,s_{z,t}}\right)
\end{aligned}
$$

where Ξ is the productivity trend and Z_t is a standard "level" technology shock. We specify the exogenous government spending process in deviation from the trend, so that

$$
\frac{G_t}{B_t} = \left(\frac{gY}{B}\right)^{1-\rho_g} \left(\frac{G_{t-1}}{B_{t-1}}\right)^{\rho_g} e^{\varepsilon_t^g} \qquad \varepsilon_{t+1}^G \backsim N\left(0, \sigma_{G,s_{G,t}}\right)
$$

where the long run level g is specified in percent of output, so that $g \equiv G/Y$.

Both technology and government spending shocks have regime-switching variances, namely

$$
\begin{aligned}
\sigma_{z,s_{z,t}} &= \sigma_{z,L} s_{z,t} + \sigma_{z,H}\left(1 - s_{z,t}\right) \\
{}_{G,s_{G,t}} &= \sigma_{G,L} s_{G,t} + \sigma_{G,H}\left(1 - s_{G,t}\right)
\end{aligned}
$$

and the variables $s_{z,t}$, $s_{G,t}$ can assume the discrete values 0 and 1. For each variable $s_{j,t}$ $(j = z, G)$, the probabilities of remaining in state 0 and 1 are constant and equal to $p_{j,0}$ and $p_{j,1}$, respectively.

We assume regime switches in these particular variances for the following reasons. The literature on the Great Moderation (see e.g. McDonnell and Perez-Quiros, 2000) has emphasised the reduction in the volatility of real aggregate variables starting in the second half of the 1980s. We conjecture that this phenomenon could be captured by a reduction in the volatility of technology shocks in our structural setting. The literature has also often found a relationship between regimes and the business cycle. In our model, this relationship could be accounted for by regime switches of the volatility of demand (government spending) shocks.

5.6 Data and prior distributions

We estimate the model on quarterly US data over the sample period from 1966Q1 to 2009Q1. Our estimation sample starts in 1966, because this is often argued to be the date after which a Taylor rule provides a reasonable characterisation of Federal Reserve policy.[3]

The data included in the information set are real per-capita consumption growth, the consumption deflator and the 3-month nominal interest rate. Measurement error characterises the nominal interest rate.

The data are described in figure (2). Note that, beside the differencing for log consumption, the data are not subjected to any prior transformation.

We include in the information set total real personal consumption per-capita and the consumption deflator (from the FRED database of the St. Louis Fed). In addition, we use the 3-month nominal interest rate (from the Federal Reserve Board).

Prior distributions for our model parameters are presented in Table (2).

Concerning regime switching processes, we assume beta priors for transition probabilities. The distributions imply that persistences in each state are symmetric and have high means. In the prior, we assume that the standard deviations of the structural shocks are identical in the various states.

The priors for the standard deviation and persistence of shocks, as well as for the long run growth rate of technology and for the long run inflation target, are centred on values which allow us to roughly match unconditional data moments in the first 10 years of the sample, given the other parameter.

For the policy rule, we use relatively loose priors centred around the classic Taylor (1993) parameters for the responses to inflation and output, but we also allow for a substantial degree of interest rate smoothing. Finally, for the other parameters we use priors broadly in line with other macro studies.

Note that we impose priors which are completely symmetric across states. Therefore, we sample from the posterior and ex post impose the constraint that state 1 for each of the discrete unobservable state variables is that with

[3] According to Fuhrer (1996), "since 1966, understanding the behaviour of the short rate has been equivalent to understanding the behaviour of the Fed, which has since that time essentially set the federal Funds rate at a target level, in response to movements in inflation and real activity". Goodfriend (1991) argues that even under the period of official reserves targeting, the Federal Reserve had in mind an implicit target for the Funds rate.

the lowest variance. This is the way in which we deal with the so-called label switching problem of Markov switching models. For a discussion of this problem and its bearing on posterior simulation, see Geweke and Amisano (2011).

5.7 Results

Bayesian estimation is performed by using a single block random walk Metropolis-Hastings algorithm and a multivariate Gaussian distribution as candidate density. We initialise the algorithm by finding the mode of the log posterior via simulated annealing and computing the covariance matrix of the candidate distribution by using a numerical Hessian of the posterior distribution at its mode.

We run several chains and we report results here based on 250,000 draws. The results from the posterior distribution are summarised in Table (3). A few notable features are apparent from this table.

Looking at the transition probabilities of posterior distribution, we note that their marginal distributions are centered on mean values which are not very different from their prior counterparts (with the exception of the probability of staying in the regime of low variability for the G process). The state of high volatility seems to be much more persistent than that of low volatility for the G process; to a lesser extent, this occurs also for the Z process.

Note also that the posterior distributions of the state specific standard errors are quite polarised for the Z process: the standard deviation of Z in the high volatility state is twice as high as the low volatility standard deviation (140 vs. 70 bps in quarterly terms), whereas this ratio is just 1.5 for the G shock (340 vs 230 bps).

Looking at the policy rule parameters, we see that the interest smoothing parameter has a posterior mean which is higher than its prior mean (.80 versus .70), while the inflation response coefficient posterior mean is lower than its prior mean (.23 vs. .50).

It is interesting to consider one step ahead forecasting errors as a way to gauge model fit. These are illustrated in Figure (3). Looking at one step ahead forecasting errors, the fit seems to be reasonably good. One step ahead forecasts track actual variables quite well, even in the last part of the sample which is affected by the recent financial crisis. Moreover, forecast errors do not display clear signs of serial correlation.

Figure (4) displays the filtered values for the continuous latent variables. Looking at the filtered values for the measurement error on interest rate, we see that it shows abnormal volatility in the early 80s, possibly in relation with the so called "monetarist experiment". We do not model this feature of the data in this paper, but in Amisano and Tristani (2011) we capture it through the assumption of heteroskedasticity in policy shocks.

Another interesting feature is the pattern displayed by the technology process and the government spending process, as documented in the first and second panels of figure (4). Keeping in mind that these variables are expressed in deviations from a deterministic trend which has a posterior mean growth rate of 20 quarterly bps, we notice from the upper left panel that technology has oscillated around the deterministic growth trend with sharp dips during the mid 1970s, early 1980s and early 1990s recessions., while the government implied by the model, followed the deterministic trend until the early 1980s, and then sharply fell until the onset of the current recession.

Turning to the analysis of the discrete states indexing volatility regimes (Figure (5)), we notice some interesting features.

First of all, given the posterior distribution of transition probabilities, we expect G to spend much more time in the high volatility state. The same for the Z process, even if the difference between posterior means of within state persistence probabilities is smaller than in the G case. This is clearly confirmed by looking at filtered and smoothed probabilities. In the case of the G process, it seems that the low volatility state is never visited in the sample.

Overall, the interpretation for the variances of the productivity growth and government spending processes is different from our conjecture. The technology shock does show also a very prolonged spell of low volatility from the mid 1980s to the early 2000s, which is consistent with the Great Moderation phenomenon. The model also suggests that the Great Moderation came to an end after 2004, when the technology shock is estimated to return to a high volatility regime. However, the government spending shock does not display a cyclical pattern. This is instead the case for the technology shock, whose variance tends to increase during various recessions over the mid-1970s and early 1980s. This also appears to be true for the recession associated with the recent financial crisis.

6 Conclusions

This paper shows that the second-order approximate solution of DSGE models with Markov switching variances is characterised by coefficients on the linear and quadratic terms in the state vector of the decision rules that are independent of the volatility of the exogenous shocks. Up to second order, only the constant term of the decision rules is affected by the introduction of regime switching. We devise a procedure to compute the likelihood in situations in which the number of shocks in the empirical model match the number of observable variables (and there are no unobservable predetermined variables). In such an environment we can compute the likelihood exactly via recursive methods without resorting to approximations or to simulation filtering techniques.

The results of an application of our solution and estimation methods to US data are consistent with the hypothesis that changes in the conditional variance of technology shocks are responsible for both the Great Moderation and the cyclical features of macroeconomic volatility.

Acknowledgement

The opinions expressed are personal and should not be attributed to the European Central Bank. We thank Michael Juillard, two anonymous referees and Sungbae An, our discussant at the JEDC Conference on "Frontiers in Structural Macroeconomic Modelling", Tokyo, January 23rd and 24th 2010, where the first draft of this paper was presented, for useful comments and suggestions. We also thank Francesco Bianchi, Szabolcs Deak, Giovanni Lombardo and participants in the "Estimation of DSGE models" session of the Econometric Society World Conference in Shanghai, August 2010.

References

[1] Amisano, G., Tristani O, 2010. Euro area inflation persistence in an estimated nonlinear DSGE model. Journal of Economic Dynamics and Control, 34, 1837-1858.

[2] Amisano, G., Tristani O, 2011. A DSGE model of the term structure with regime shifts, forthcoming, European Central Bank.working paper series.

[3] Andolfatto, D, Gomme P., 2004. Monetary policy regimes and beliefs. International Economic Review 44, 1-30.

[4] Aruoba, B., Fernandez-Villaverde, J., Rubio-Ramirez, J.F., 2006. Comparing Solution Methods for Dynamic Equilibrium Economies. Journal of Economic Dynamics and Control 30, 2477-508.

[5] Benigno, P., Woodford M., 2006. Linear-quadratic approximation of optimal policy problems, mimeo.

[6] Blanchard, O. Kahn, C., 1980. The solution of linear difference models under rational expectations, Econometrica 48, 1305-1311.

[7] Cecchetti, S.G., Lam, P., Mark, N.C., 1993. The equity premium and the risk-free rate, Journal of Monetary Economics 31, 21-45.

[8] Cogley, T., Sargent T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII U.S., Review of Economic Dynamics 8, 262-302.

[9] Coleman, W.J. , 1991. Equilibrium in a production economy with an income tax, Econometrica 59, 1091-104.

[10] Davig, T. Leeper, E.M., 2007. Generalising the Taylor principle, American Economic Review 97, 607-635

[11] Davig, T., Leeper, E.M., Chung, H., 2004. Monetary and fiscal policy switching, NBER WP No. 10362.

[12] Farmer, R.E.A., Waggoner, D.F., Zha, T., 2010. Generalizing the Taylor principle: a comment, American Economic Review 100, 608-617

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

[14] Geweke, J., Amisano G., 2011, Hierarchical Markov mixture models and their applications in finance, Journal of Applied Econometrics 26, 1-29.

[15] Gomme, P. Klein, P., 2006. Second-order approximation of dynamic models without the use of tensors, University of Western Ontario, *mimeo*, November.

[16] Guan, Y., Verschelde, J., 2008. PHClab: A MATLAB/Octave interface to PHCpack, in Stillman, M.E., Takayama, N., Verschelde, J. (eds.), IMA Volume 148: Software for Algebraic Geometry, 15-32, Springer-Verlag.

[17] Hamilton, J.D. , 1994. Time series analysis, , Princeton University Press, Princeton).

[18] Hamilton, J.D. , 2008. Regime switching models, in Durlauf, S., Blume, L. (eds.), New Palgrave Dictionary of Economics, second edition, Palgrave, Mc Millan Ltd.

[19] Judd, K. , 1998. Numerical methods in economics, Cambridge (MA): MIT Press.

[20] Kim, J., Kim, S., Schaumburg E., Sims, C.A., 2003. Calculating and using second -order accurate solutions of discrete time dynamic equilibrium models. *mimeo*, August.

[21] 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.

[22] 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.

[23] Magnus, I., Neudecker, H., 1999. Matrix differential calculus with applications in statistics and econometrics, New York: Wiley.

[24] McConnell, M., Perez-Quiros, G., 2000. Output fluctuations in the United States: what has changed since the early 1980s?" American Economic Review 90, 1464-76.

[25] Morgan, A., 1987. Solving polynomial systems using continuation for engineering and scientific problems. Englewood Cliffs, N.J.: Prentice-Hall.

[26] Pitt, M., Shephard, N., 1999. Filtering by simulation: auxiliary particle filters, Journal of the American Statistical Association 94, 590-599.

[27] Pitt, M., Shephard, N., 2001. Auxiliary variable based particle filters, in Doucet, A., de Freitas, N., Gordon N. (eds.), Sequential Monte Carlo methods in practice, Springer, New York.

[28] Primiceri, G. , 2005. Time varying structural vector autoregressions and monetary policy, Review of Economic Studies 72, 821–852.

[29] 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.

[30] Schorfheide, F., 2005. Learning and monetary policy shifts, Review of Economic Dynamics 8, 392-419.

[31] Sims, C.A., Zha T., 2006. Were there regime switches in U.S. monetary policy?, American Economic Review 96, 54-81.

[32] Sims, C.A. , 2001. Solving linear rational expectations models, Journal of Computational Economics 20, 1-20.

[33] Söderlind, P., 1999. Solution and estimation of RE macromodels with optimal policy, European Economic Review 43, 813-823.

[34] Verschelde, J. , 1999: Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation. ACM Trans. Math. Softw. 25(2), 251-276. Software available at http://www.math.uic.edu/~jan.)

Fig. 1. Univariate example

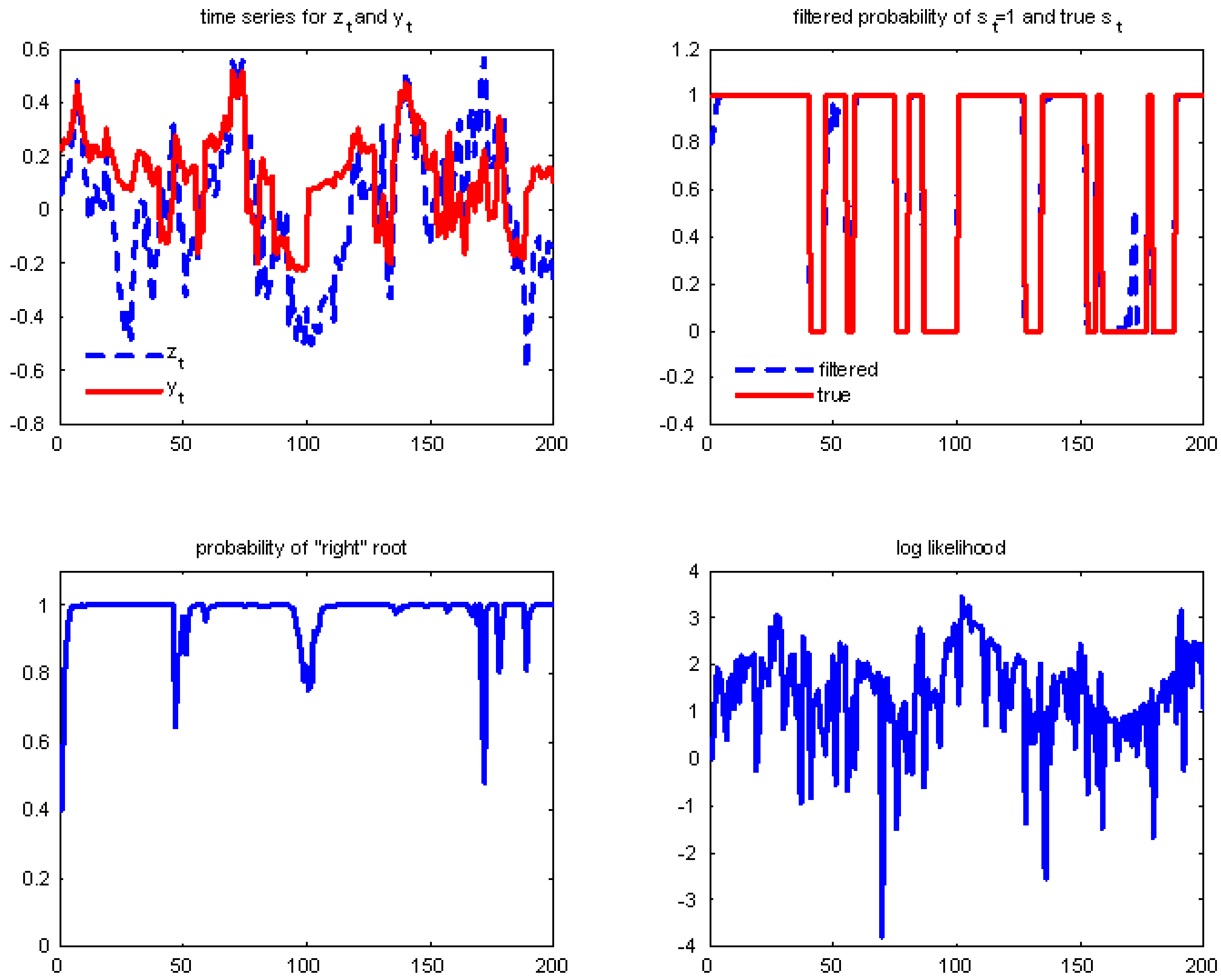
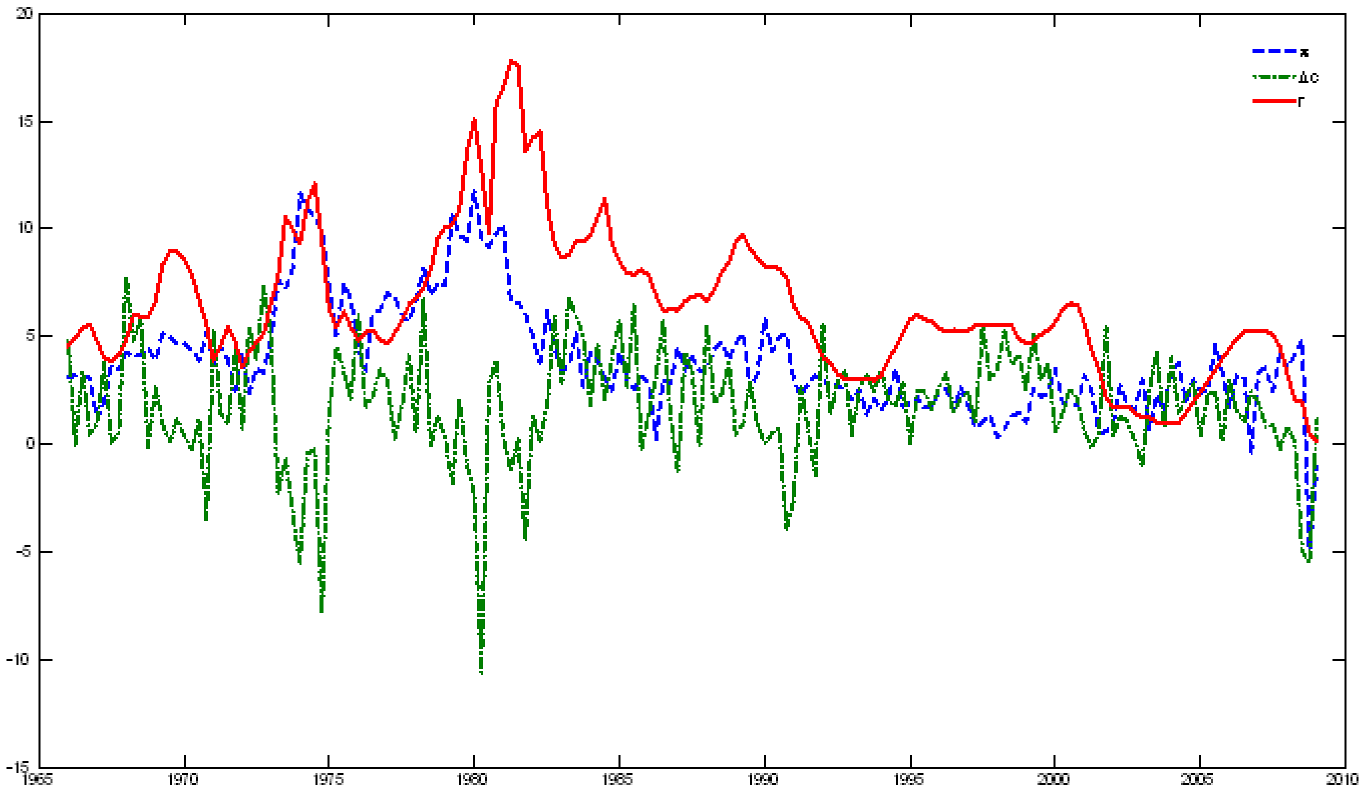
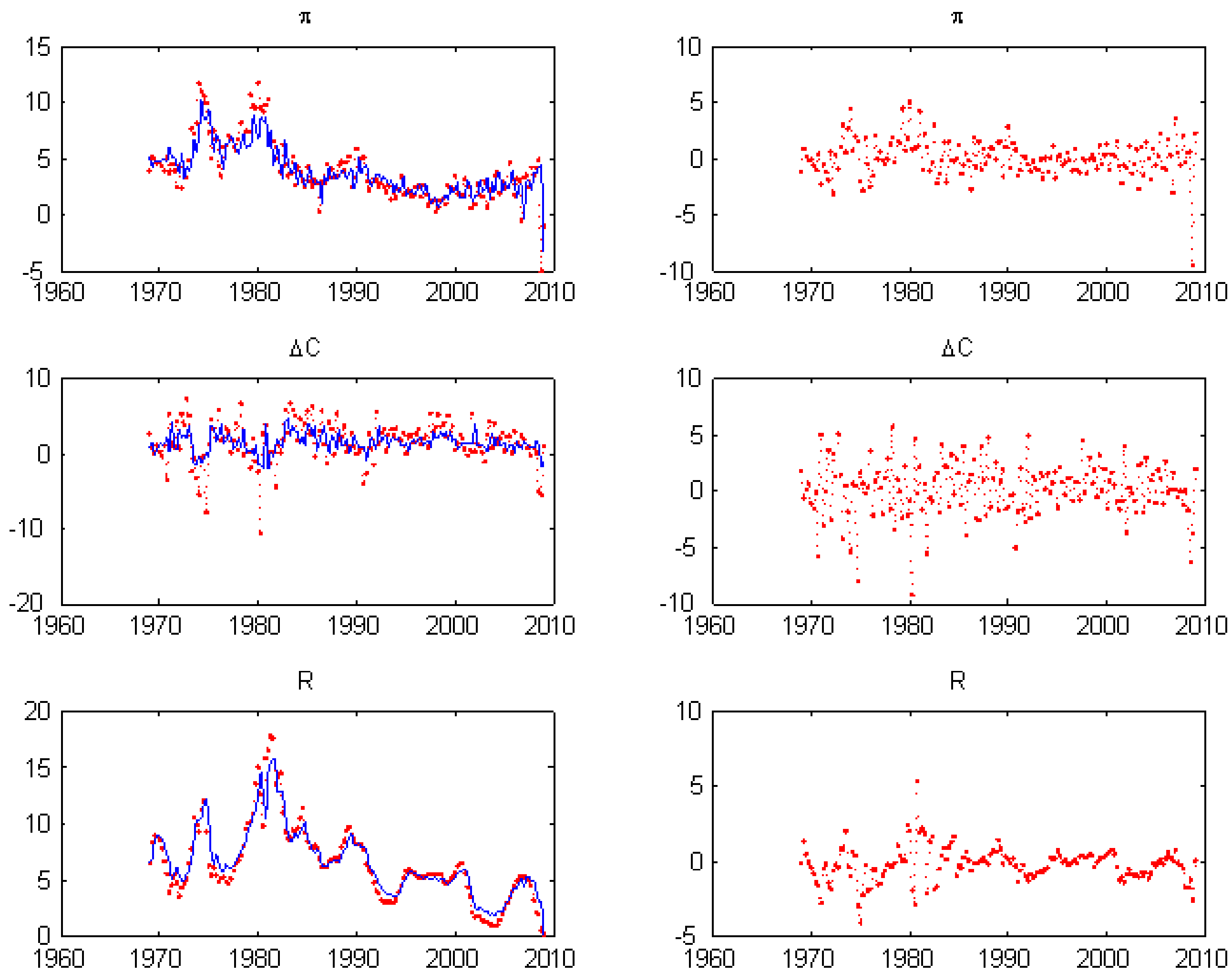


Fig. 2. Series used for model estimation



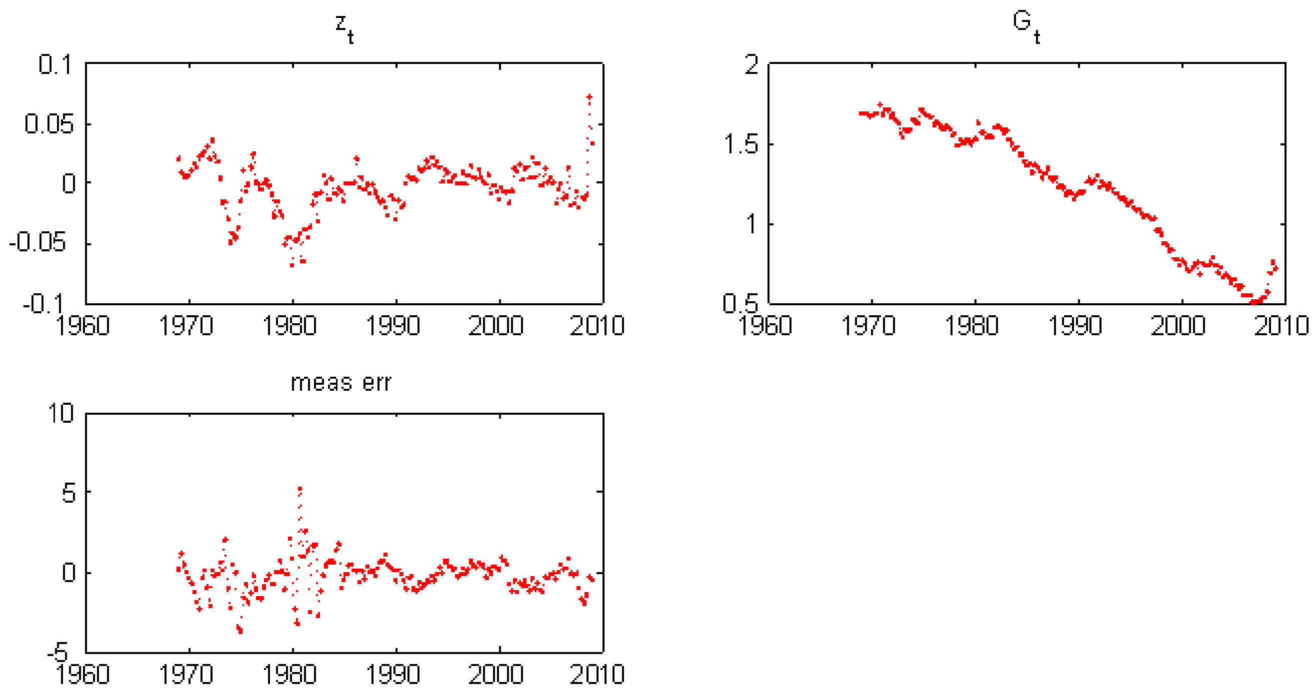
Note: in this figure series are represented in annual percentage terms.

Fig. 3. Model fit: actual series and one step ahead predictions



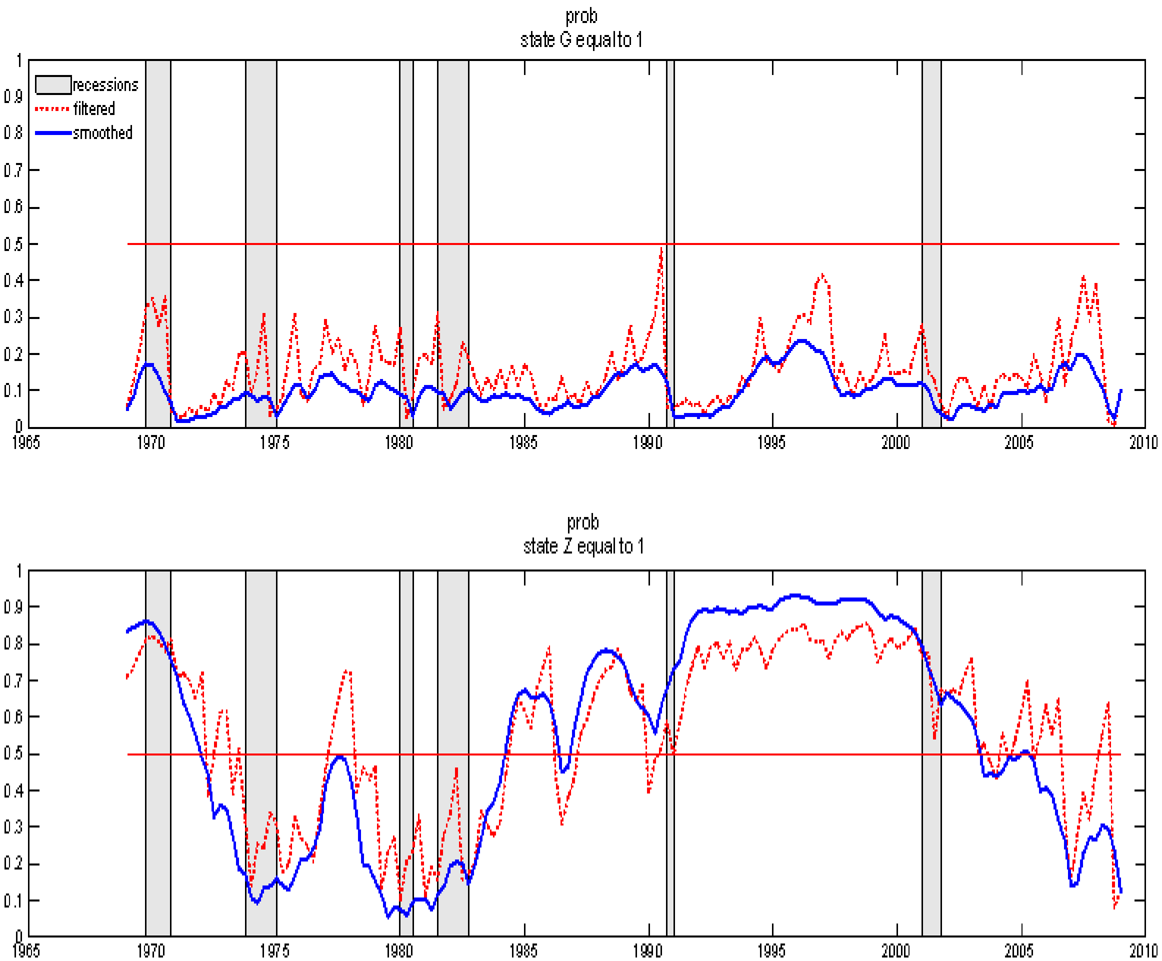
Legend: The left-hand side panels contain actual (red) and one step ahead forecasts of observed variables. The right-hand side panels contain one step ahead forecast errors. The reported values are posterior means of these quantities using the posterior distribution of estimates.

Fig. 4. Continuous latent variables, filtered variables



Legend: Filtered values are computed by averaging over draws from posterior distribution of the parameters.

Fig. 5. Filtered and smoothed values for discrete state variables



Legend: Filtered and smoothed probabilities are computed by averaging over draws from joint posterior distribution of the parameters and discrete states.

Table 1. parameter values in the simple model

$a_{0,1}$	$a_{0,2}$	a_1	b_1	b_2	$c_{0,1}$	$c_{0,2}$	c_1	c_2	$\mathbf{P}$
.0	$-.01$	0.9	.01	.2	.2	$-.1$	0.4	0.3	$\begin{bmatrix} .95 & .05 \\ .2 & .8 \end{bmatrix}$

Table 2
Prior specification

parameter name	role	pr. type	mean	sd	low. q.	up. q.
$p_{G,11}$	prob G staying in reg. 1	Beta	0.8994	0.0660	0.7358	0.9872
$p_{G,22}$	prob G staying in reg. 0	Beta	0.8998	0.0654	0.7397	0.9869
$p_{z,11}$	prob z staying in reg. 1	Beta	0.9005	0.0653	0.7414	0.9872
$p_{z,22}$	prob z staying in reg. 0	Beta	0.9005	0.0650	0.7399	0.9862
$\sigma_{G,1}$	std dev G state 1	inv. Ga.	0.0066	0.0016	0.0042	0.0103
$\sigma_{G,2}$	std dev G state 0	inv. Ga.	0.0095	0.0032	0.0056	0.0177
$\sigma_{z,1}$	std dev z state 1	inv. Ga.	0.0019	0.0004	0.0013	0.0028
$\sigma_{z,2}$	std dev z state 0	inv. Ga.	0.0026	0.0007	0.0017	0.0042
ρ_G	persistence G	Beta	0.9001	0.0300	0.8352	0.9503
Ξ	deterministic growth rate	shifted Ga.	1.0040	0.0028	1.0005	1.0113
ρ_z	persistence z	Beta	0.2028	0.1225	0.0286	0.4892
$\overline{\pi}$	inflation target	shifted Ga.	1.0070	0.0026	1.0029	1.0129
ψ_π	Taylor r. inflation par	Gamma	0.4952	0.3506	0.0583	1.3708
ψ_y	Taylor r. output par	Gamma	0.0308	0.0217	0.0038	0.0859
ρ_I	Taylor r. int.rate par	Norm.	0.6963	0.3006	0.1074	1.2840
ι	Infl. indexation	Beta	0.6040	0.1986	0.1984	0.9320
ϕ	labour elasticity	Gamma	2.0142	1.4147	0.2445	5.5442
σ	relative risk aversion	shifted Ga.	1.9938	0.9990	1.0274	4.7546
ζ	price adj. cost	Norm.	16.9960	2.0089	13.0531	20.9463
h	habit parameter	Beta	0.7047	0.1074	0.4753	0.8914
θ	el. subst across goods	shifted Ga.	7.9667	2.6746	3.7435	14.1693
β	discount factor	Beta	0.9939	0.0042	0.9835	0.9992
ξ_0	intercept det trend	Norm.	-0.9928	0.9857	-2.9430	0.9281
$\sigma_{me,I}$	meas std error on int rate	inv. Ga.	0.0001	0.0001	0.0001	0.0003

Legend: "sd" denotes the standard deviation; "low q" and "up q" denote the 5th and 95th percentiles of the distribution. Note that shifted gamma distribution for z means that z-1 has Gamma distribution.

Table 3
Results from posterior simulation of model

	po. m.	po. sd.	po. low. q.	po. up. q.	pr. m.	pr. sd.	pr. low. q.	pr. up. q.
$p_{G,11}$	0.8034	0.0741	0.6635	0.9390	0.8994	0.0660	0.7358	0.9872
$p_{G,22}$	0.8875	0.0468	0.7652	0.9467	0.8998	0.0654	0.7397	0.9869
$p_{z,11}$	0.9065	0.0457	0.8074	0.9720	0.9005	0.0653	0.7414	0.9872
$p_{z,22}$	0.9230	0.0481	0.8195	0.9834	0.9005	0.0650	0.7399	0.9862
$\sigma_{G,1}$	0.0231	0.0046	0.0149	0.0330	0.0066	0.0016	0.0042	0.0103
$\sigma_{G,2}$	0.0340	0.0038	0.0272	0.0415	0.0095	0.0032	0.0056	0.0177
$\sigma_{z,1}$	0.0071	0.0014	0.0050	0.0103	0.0019	0.0004	0.0013	0.0028
$\sigma_{z,2}$	0.0138	0.0024	0.0101	0.0195	0.0026	0.0007	0.0017	0.0042
ρ_G	0.9974	0.0008	0.9958	0.9986	0.9001	0.0300	0.8352	0.9503
Ξ	1.0020	0.0004	1.0013	1.0027	1.0040	0.0028	1.0005	1.0113
ρ_z	0.7868	0.0442	0.6918	0.8630	0.2028	0.1225	0.0286	0.4892
$\overline{\pi}$	1.0065	0.0017	1.0038	1.0097	1.0070	0.0026	1.0029	1.0129
ψ_π	0.2276	0.0088	0.2092	0.2428	0.4952	0.3506	0.0583	1.3708
ψ_y	0.0035	0.0012	0.0016	0.0063	0.0308	0.0217	0.0038	0.0859
ρ_I	0.7979	0.0081	0.7834	0.8150	0.6963	0.3006	0.1074	1.2840
ι	0.1241	0.0654	0.0530	0.3216	0.6040	0.1986	0.1984	0.9320
ϕ	2.1883	0.3812	1.5075	3.0168	2.0142	1.4147	0.2445	5.5442
σ	2.1090	0.3682	1.6070	3.0490	1.9938	0.9990	1.0274	4.7546
ζ	18.9902	1.8762	15.2896	22.7451	16.9960	2.0089	13.0531	20.9463
h	0.4944	0.0261	0.4433	0.5439	0.7047	0.1074	0.4753	0.8914
θ	1.7306	0.1766	1.4140	2.1205	7.9667	2.6746	3.7435	14.1693
β	0.9994	0.0004	0.9985	0.9998	0.9939	0.0042	0.9835	0.9992
ξ_0	0.1252	0.0761	-0.0171	0.2805	-0.9928	0.9857	-2.9430	0.9281
$\sigma_{me,I}$	0.0024	0.0000	0.0024	0.0025	0.0001	0.0001	0.0001	0.0003

Results based on 250,000 draws from random walk MH algorithm initialised at posterior mode, saved after 25,000 (burn-in) draws were discarded to get rid of initial conditions. Acceptance rate = .55. One in ten draws were then used to compute posterior moments of functions of interest.