WORKING PAPERS

Paul Pichler

Forecasting with estimated dynamic stochastic general equilibrium models: The role of nonlinearities

March 2007

Working Paper No: 0702

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

All our working papers are available at: http://mailbox.univie.ac.at/papers.econ

Forecasting with estimated dynamic stochastic general equilibrium models: The role of nonlinearities

Paul Pichler*

March 2, 2007

Abstract

In this paper we study the effects of nonlinearities on the forecasting performance of a dynamic stochastic general equilibrium model. We compute first and second-order approximations to a New Keynesian monetary model, and use artificial data to estimate the model's structural parameters based on its linear and quadratic solution. We find that, although our model in not far from being linear, the forecasting performance improves by capturing the second-order terms in the solution. Our findings suggest that accounting for nonlinearities will improve the predictive abilities of DSGE models in many applications.

Keywords: forecasting, particle filter, Kalman filter, DSGE, New Keynesian model

JEL classification: C68, E47, E52

*Department of Economics, University of Vienna, Tel.+43-1-4277-37409, e-mail: paul.pichler@univie.ac.at. I thank Jesús Fernández-Villaverde, Gerhard Sorger and Peter Hackl for many helpful comments and suggestions. I further gratefully acknowledge financial support from the Austrian Science Fund (FWF) under Project Number P17886.

1 Introduction

Forecasting macroeconomic variables is essential for economic decision making. Firms and households, for example, usually base their investment decisions on forecasts of future interest rates. Central banks, on the other hand, typically use forecasts of future growth and inflation rates to decide upon changes in monetary aggregates. Similarly, fiscal policy choices in general require predictions of future economic conditions, such as unemployment rates, growth rates, and tax revenues.

Unfortunately, forecasting macroeconomic variables is not an easy exercise. Actual economies are dynamic, nonlinear, highly dimensional entities, macroeconomic time series are highly aggregated, samples are short, data are measured with error, and some important macroeconomic series are not measured at all.[1] Obviously it is hard to construct models that capture the essential features of the data generating process, and consequently, forecasting macroeconomic time series is a challenging task.

Given its importance and complexity, it is not surprising that economic forecasting is a very active field of research. Various competing methods have been proposed so far, among which are very different approaches such as large-scale macroeconometric models, univariate time series models, and vector autoregressions, to name just a few.[2]

Only very recently, macroeconomic dynamic stochastic general equilibrium (DSGE) models have been suggested as forecasting tools. As was demonstrated in recent work by Ireland (2004a), Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2003), among others, modern DSGE models are sufficiently rich to capture the dynamics of actual data, and therefore ready to be used for forecasting. Two good reasons suggest their particular value in this field.[3] First, since DSGE models typically have less parameters than non-structural econometric models, they can in principle be more precisely estimated. This is likely to result in a relatively better forecasting performance, especially when samples are short. The second reason is that DSGE models are less subject to the Lucas critique. Their parameters usually describe preferences and technology, and are thus *deep* in the sense that they do not vary with policy. Therefore, DSGE models allow to evaluate and forecast the impacts of changes in policy, which makes them particularly attractive for policymakers such as central banks.

[1] See Hendry (1995).

[2] See Clements and Hendry (1998, 1999) for textbook expositions.

[3] See Diebold (1998) and Del Negro and Schorfheide (2003) for a more detailed discussion.

Focusing on the class of New Keynesian models[4], several recent contributions have demonstrated that DSGE models perform well in forecasting. Results from Smets and Wouters (2004), Adolfson, Linde, and Villani (2005), and Dib, Gammoudi, and Moran (2006), show that forecasts generated directly from a New Keynesian model are no worse than forecasts from unrestricted vector autoregressions. The good performance is demonstrated to hold for models estimated in both classical and Bayesian environments, and for macroeconomic data sets of different countries. A second role for DSGE models in forecasting has been pointed out by Del Negro and Schorfheide (2004) and Del Negro, Schorfheide, Smets, and Wouters (2004). They demonstrate the particular good forecasting properties of DSGE-VARs, i.e., Bayesian vector autoregressions where prior distributions are generated from DSGE models.

A limitation of the existing literature on DSGE forecasting, however, is that it restricts attention to linearized economies. The model's structural parameters are estimated based on the *approximate* likelihood function implied by the linearized model, and forecasts are generated based on the dynamics implied by linearized decision rules. Recent work by Fernandez-Villaverde and Rubio-Ramirez (2005, forthcoming) and Fernandez-Villaverde, Rubio-Ramirez, and Santos (2006) points out that this practice could be misleading. They argue that even in scenarios where linearization is accurate enough in terms of the policy functions, it is likely to be not accurate enough in terms of the likelihood function, as second-order errors in the policy functions have first-order effects on the likelihood function. Fernandez-Villaverde and Rubio-Ramirez (2005) show that even in the almost-linear neoclassical growth model, linearization leads to biased estimates of the model's parameters, and seriously distorts the dynamic properties of the model. Consequently, they suggest to move to at least second-order approximations when taking DSGE models to the data. Indeed, recent papers by An (2005), An and Schorfheide (forthcoming), Fernandez-Villaverde and Rubio-Ramirez (forthcoming), and Amisano and Tristani (2006) demonstrate that the fit of DSGE models can be improved by accounting for nonlinearities.

Unfortunately, however, capturing nonlinearities is not a straightforward exercise. When we use a second-order accurate solution, the model translates into a quadratic state space system. As a consequence, we can no longer use the Kalman filter to construct the likelihood function, since it relies on the linearity of the system. Instead we need to rely on Sequential Monte Carlo

[4]Broadly speaking, New Keynesian (NK) models are DSGE environments where imperfect competition and nominal rigidities allow monetary policy to have real effects. See Woodford (2003) for more information.

methods, for example *particle filters*, to approximate the likelihood function. Since these filters use simulation methods, this slows down the estimation substantially and introduces sampling error, which may negatively affect the predictive abilities of the model.

Obviously the following question arises. Is it worth moving from linear to nonlinear approximations when using DSGE models as forecasting tools? In this paper we analyze this question within the framework of a simulation exercise. We generate several artificial data sets, and compare the forecasting performances of the log-linearized New Keynesian model to the quadratic model, and to unrestricted VAR(1) and VAR(2) models. We first analyze the case where the data generating process is the *nonlinear* DSGE model, i.e., the data are generated from the New Keynesian model solved with a third-order Galerkin projection method.[5] Finally, we repeat this exercise using artificial time series that come from a VAR(2) model.

Our main findings are the following. First, the New Keynesian model performs well in forecasting, even under misspecification. This result affirms the conclusions of the papers that have demonstrated the good performance of DSGE models in forecasting actual data. Secondly, we find that capturing second-order terms in the policy functions improves the forecasting performance of the New Keynesian model. This is not only true when the New Keynesian model resembles the data generating process, but holds also when the data are generated from a VAR(2) model.

The remainder of this paper is organized as follows. Section two presents our version of the New Keynesian model and defines its equilibrium. Section three briefly illustrates the first and second-order accurate perturbation solutions. Section four recasts the linear and quadratic model in state-space form. Section five describes how to use Kalman and particle filtering to conduct Maximum likelihood inference. Section six outlines our forecasting exercise. Section seven presents its results. Section eight summarizes and concludes. Finally, we provide the derivation of the model's equilibrium conditions, the model's non-stochastic steady state, and details on the construction of artificial data in three appendices.

[5]We are aware of the fact that the third-order projection method is computationally very demanding and delivers data which are not much different from the data we could obtain from a second-order perturbation solution. However, we follow this approach because then, neither the first nor the second-order accurate solution coincides with the actual data generating process. Instead, both methods are approximations to a *more nonlinear* DGP. This feature, we believe, is interesting when studying the effects of approximation methods.

2 The Model

This section develops the New Keynesian DSGE model that we will use later to forecast macroeconomic aggregates. We choose to work with this model for basically two reasons. First, because New Keynesian models are among the most popular tools in modern dynamic macroeconomics,[6] such that their properties are very well understood. Secondly, because these models are typically not far from being linear. If nonlinearities turn out to play a positive role in forecasting with New Keynesian models, they will be of value for most forecasting applications of DSGE models.

2.1 The economic environment

Our model builds on the framework developed by Ireland (1997), and is closely related to models studied by Kim (2000), Ireland (2004a, 2004b), and Dib, Gammoudi, and Moran (2006), among others. Time is discrete and goes on forever, i.e. $t \in \{0, 1, 2, ...\}$. The economy is populated by a representative household, a representative finished goods-producing firm, a continuum of intermediate goods-producing firms indexed by $j \in [0, 1]$, a monetary authority and a government. The following sections discuss these agents in turn.

2.1.1 The representative household

The household enters period t with M_{t-1} units of money, B_{t-1} units of bonds and k_t units of capital. Within the period, it receives factor payments from supplying labor $h_t(j)$ and capital $k_t(j)$ to each intermediate goods-producing firm $j \in [0, 1]$, whereby it takes the nominal factor prices W_t and Q_t as given. The total amounts of labor and capital supplied by the household are denoted by $h_t = \int_0^1 h_t(j)dj$ and $k_t = \int_0^1 k_t(j)dj$. Labor income is taxed by the government at the rate τ_t. Finally, the household receives dividend payments from the intermediate goods-producing firms, $D_t = \int_0^1 D_t(j)dj$, and a nominal transfer T_t from the government. If negative, T_t can be interpreted as a lump-sum tax.
The household's expenditures are composed as follows. First, the household purchases the final good which is used for both consumption c_t and investment x_t. In order to transform x_t units of of the final good into x_t units of productive capital, the household has to pay a transformation cost, which is

[6] See Woodford (2003).

given by

$$CAC_t = \frac{\phi_k}{2}\left(\frac{k_{t+1}}{k_t} - 1\right)^2 k_t.$$

This specification reflects the idea that it is easier to install new capital gradually at a slow rate. The parameter ϕ_k measures the size of adjustment costs relative to the capital stock. Furthermore, the household purchases new bonds at a nominal price of $1/i_t$, where i_t denotes the gross nominal interest rate between periods t and $t+1$. The remainder of funds is carried over into the next period in the form of money, M_t. Letting P_t denote the price level at time t, the household's budget constraint is given by

$$\frac{M_{t-1} + B_{t-1} + (1-\tau_t)W_t h_t + Q_t k_t + D_t + T_t}{P_t} \geqslant c_t + x_t + \frac{\phi_k}{2}\left(\frac{k_{t+1}}{k_t} - 1\right)^2 k_t + \frac{B_t/i_t + M_t}{P_t}. \quad (1)$$

Capital depreciates at the constant rate δ. The capital stock thus evolves according to

$$k_{t+1} = (1-\delta)k_t + x_t. \quad (2)$$

The household's spending decisions are made to maximize expected utility derived from lifetime consumption, money holdings and leisure. Utility in future periods is discounted by a constant factor β. The household's choice is subject to its budget constraint, a no-borrowing constraint on capital, i.e. $k_t \geqslant 0 \forall t \in \{0, 1, 2, ...\}$, and the law of motion for the capital stock. Formally, the household solves

$$\max_{c_t, M_t/P_t, h_t, B_t/P_t, k_{t+1}} \sum_{t=0}^{\infty} \beta^t u(c_t, M_t/P_t, h_t)$$

where

$$u(c_t, \frac{M_t}{P_t}, h_t) = a_t \frac{\gamma}{\gamma - 1} \log\left(c_t^{\frac{\gamma-1}{\gamma}} + b_t^{\frac{1}{\gamma}}\left(\frac{M_t}{P_t}\right)^{\frac{\gamma-1}{\gamma}}\right) + \eta \log(1 - h_t),$$

subject to (1) and (2). The parameter γ corresponds to the interest elasticity of money demand, η values leisure in the utility function, a_t and b_t are taste shocks. As shown by Ireland (1997), a_t resembles a shock to the IS curve in traditional Keynesian analyses, whereas b_t can be interpreted as a

money demand shock. The logarithms of the shocks follow the stationary autoregressive processes

$$\begin{aligned}\log(a_{t+1}) &= (1-\rho_a)\log(a) + \rho_a \log(a_t) + \epsilon_{a,t},\\ \log(b_{t+1}) &= (1-\rho_b)\log(b) + \rho_b \log(b_t) + \epsilon_{b,t},\end{aligned}$$

with $\epsilon_{a,t} \sim N(0, \sigma_a^2)$ and $\epsilon_{b,t} \sim N(0, \sigma_b^2)$.

2.1.2 The representative finished-goods-producing firm

The representative finished-goods-producing firm is assumed to produce y_t units of a single output good on a perfectly competitive market. Intermediate goods, $y_t(j)$, $j \in [0,1]$, serve as the only inputs in the production. The firm's objective is to maximize profits, whereby it takes the price of its own output good, as well as the prices of all intermediate goods as given. Formally, it solves the problem

$$\max_{y_t, y_t(j)} P_t y_t - \int_0^1 y_t(j) P_t(j) dj \tag{3}$$

subject to a constant returns to scale technology,

$$y_t = \left[\int_0^1 y_t(j)^{(\theta-1)/\theta} dj\right]^{\theta/(\theta-1)}, \tag{4}$$

and for given prices P_t and $P_t(j)$, $j \in (0,1)$. Due to perfect competition on the final goods market, the firm earns zero profits in equilibrium. It is easy to check that, as a consequence, the equilibrium price for the output good is given by

$$P_t = \left[\int_0^1 P_t(j)^{1-\theta} dj\right]^{\frac{1}{1-\theta}},$$

and the demand for intermediate good j equals

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

2.1.3 The representative intermediate-goods-producing firm

The intermediate goods-producing firm j hires $h_t(j)$ units of labor and $k_t(j)$ units of productive capital from the household to produce $y_t(j)$ units of the

intermediate good j. The production technology is described by a Cobb-Douglas production function with labor augmenting technological change, i.e.,

$$y_t(j) = k_t(j)^{\alpha}[z_t h_t(j)]^{1-\alpha}.$$

The parameter α gives capital's share in output, and z_t is a technology shock which follows the autoregressive process

$$\log(z_{t+1}) = (1-\rho_z)\log(z) + \rho_z \log(z_t) + \epsilon_{z,t},$$

where $\epsilon_{z,t} \sim N(0, \sigma_z^2)$. Intermediate goods are produced on a monopolistically competitive market, such that each firm can set its nominal price. After having set its price, the firm satisfies the demand from the final goods-producing firm.

We assume that price adjustment is costly: in terms of the finished good, real price adjustment costs are given by

$$PAC_t^j = \frac{\phi_p}{2}\left[\frac{P_t(j)/P_{t-1}(j)}{\pi} - 1\right]^2 y_t$$

where ϕ_p is the adjustment cost parameter, and π denotes the steady state rate of inflation. This specification of price adjustment costs goes back to Rotemberg (1982), and has been used in a DSGE framework by Ireland (1997) and Kim (2000), among others. As emphasized by Rotemberg (1982) and Ireland (1997), it captures the negative effects of price changes on the relationship between customers and firms, which increase in magnitude with the size of the price change and with the quantity purchased.
Due to price adjustment costs the firm faces a dynamic optimization problem. We follow Ireland (1997) and assume that the firm seeks to maximize its total market value,

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{P_t} D_t(j).$$

In the above expression, λ_t denotes the Lagrangian multiplier associated with the household's budget constraint. Consequently, $\beta^t \frac{\lambda_t}{P_t}$ can be interpreted as the marginal utility value of one unit of profits in period t to the representative household. Formally, the firm's dynamic problem is given by

$$\max_{h_t(j), k_t(j), P_t(j)} E_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{P_t} D_t(j)$$

subject to the constraints

$$\begin{aligned} D_t(j) &= P_t(j)y_t(j) - W_t h_t(j) - Q_t k_t(j) - \frac{\phi_p}{2}\left[\frac{P_t(j)/P_{t-1}(j)}{\pi} - 1\right]^2 y_t P_t, \\ y_t(j) &= \frac{P_t(j)^{-\theta}}{P_t} y_t, \text{ and} \\ y_t(j) &= k_t(j)^{\alpha}[z_t h_t(j)]^{1-\alpha}. \end{aligned}$$

2.1.4 The monetary authority

The central bank conducts monetary policy by adjusting the nominal interest rate i_t. It follows a modified Taylor rule, in which it smoothes interest rates and reacts to deviations in output and inflation from their target values. Formally, the policy rule is given by

$$\log \frac{i_t}{i} = \rho_i \log \frac{i_{t-1}}{i} + \rho_y \log \frac{y_t}{y} + \rho_\pi \log \frac{\pi_t}{\pi} + \epsilon_{i,t}, \tag{5}$$

where i, y, and π denote the target (or steady-state) values of the respective variables. The central bank can choose the level of one of these target variables, as well as the parameters ρ_i, ρ_y and ρ_π. In the following we assume that the central bank sets its inflation target, π, and then implements its policy rule by adjusting the nominal money stock, such that (5) holds and the money market clears. The expression $\epsilon_{i,t} \sim N(0, \sigma_i^2)$ denotes a monetary policy shock: we assume that the central bank can influence the nominal interest rate only indirectly by setting the bank rate, such that the market interest rate is given by the central bank's target rate plus an error term which reflects movements in financial markets that cannot be influenced by the central bank. Finally, we assume that the revenue from money creation is transferred entirely to the fiscal authority.

2.1.5 The fiscal authority

The fiscal authority receives newly created money from the central bank , issues new debt, and taxes labor income. It spends its revenue to finance public goods, to provide lump-sum transfers to the representative household, and to repay debt. Formally, the government's budget constraint is given by

$$P_t g_t + T_t + B_{t-1} = \tau_t W_t h_t + (M_t - M_{t-1}) + B_t/i_t.$$

To keep the analysis as simple as possible, we assume that government spending is exogenous and constant over time, i.e. $g_t = g$. Furthermore, we assume

the tax rate is exogenously given and follows the AR(1) process

$$\log \tau_{t+1} = (1 - \rho_\tau) \log \tau + \rho_\tau \log \tau_t + \epsilon_{\tau,t},$$

where $\epsilon_{\tau t} \sim N(0, \sigma_\tau^2)$.[7] The government faces a No-Ponzi constraint, which implies that the present value of government expenditures must equal the present value of government revenues plus the initial amount of public debt. As the representative household in our model lives forever and expectations are rational, our model exhibits Ricardian equivalence in the following sense: given the tax rate on labor income, it does not matter in equilibrium whether the government finances its spending by lump-sum taxes or by debt. We thus abstract from government borrowing, i.e. we set $B_t = 0$ for all $t \in \{0, 1, 2, \ldots\}$. The government's budget constraint is then - without loss of generality - given by

$$P_t g + T_t = \tau_t W_t h_t + (M_t - M_{t-1}) \tag{6}$$

Given the exogenous process for the tax rate and the money transfers from the central bank, the government's lump-sum transfers T_t are determined residually so that the budget constraint is satisfied in every period.

2.2 The symmetric competitive equilibrium

We study the model's implications by analyzing its symmetric competitive equilibrium, in which all intermediate goods-producing firms can be represented by an aggregate firm. The symmetric competitive equilibrium in sequence formulation is defined as follows:

Definition A symmetric competitive equilibrium is a set of initial values, $\{k_0, M_{-1}, P_{-1}\}$, a price system $\{P_t, W_t, Q_t, i_t\}_{t=0}^{\infty}$, a sequence of allocations, $\{c_t, M_t, h_t, k_t, y_t, D_t, T_t\}_{t=0}^{\infty}$, a tax system $\{\tau_t\}_{t=0}^{\infty}$, and a sequence of exogenous shocks $\{a_t, b_t, z_t, \epsilon_{i,t}\}_{t=0}^{\infty}$ such that:

1. Given initial values $\{k_0, M_{-1}, P_{-1}\}$, shocks $\{a_t, b_t\}_{t=0}^{\infty}$, prices $\{P_t, W_t, Q_t, i_t\}_{t=0}^{\infty}$, profits $\{D_t\}_{t=0}^{\infty}$, lump sum transfers $\{T_t\}_{t=0}^{\infty}$, and taxes $\{\tau_t\}_{t=0}^{\infty}$, the sequences $\{c_t, h_t, M_t, k_{t+1}\}_{t=0}^{\infty}$ solve the household's optimization problem.

2. Given prices $\{P_t\}_{t=0}^{\infty}$ and $\{P_t(j)\}_{t=0}^{\infty}$ for all $j \in [0, 1]$, the quantities $\{y_t\}_{t=0}^{\infty}$ and $\{y_t(j)\}_{t=0}^{\infty}$, $j \in [0, 1]$, solve the finished-goods-producing firm's optimization problem.

[7] The simple design of the fiscal authority in our model is obviously controversial. However, since we use only simulated data in our analysis and do not want to judge the model against real data, we believe it is justified in our application.

3. Given the initial price $P_{-1}(j)$, sequences $\{P_t, W_t, Q_t, y_t, z_t\}_{t=0}^{\infty}$, and the final demand for good j, $\{y_t(j)\}_{t=0}^{\infty}$, the sequences $\{h_t(j), k_t(j), P_t(j)\}_{t=0}^{\infty}$ solve firm j's optimization problem; this holds for every $j \in [0,1]$.

4. The sequences $\{i_t, y_t, P_t, \epsilon_{it}\}_{t=0}^{\infty}$ satisfy the monetary authority's policy rule at every time $t \in \{0,1,2,\ldots\}$.

5. Given initial money holdings M_{-1}, the sequences $\{P_t, T_t, \tau_t, W_t, h_t, M_t\}_{t=0}^{\infty}$ satisfy the government's budget constraint at every $t \in \{0,1,2,\ldots\}$.

6. The sequences $\{a_t, b_t, z_t, \tau_t\}_{t=0}^{\infty}$ obey their respective log-linear laws of motion.

7. For every firm $j \in [0,1]$ and every $t \in \{-1,0,1,2,\ldots\}$, it holds that $y_t(j) = y_t$, $P_t(j) = P_t$, $h_t(j) = h_t$, $k_t(j) = k_t$, and $D_t(j) = D_t$.

The equilibrium defined above is given by a collection of infinite sequences of decision variables, and the model's dynamics are therefore obviously difficult to compute and to characterize from this definition. It is more convenient to work with a recursive equilibrium in which agents choose time invariant policy functions that map the state of the economy into decision outcomes in every period $t \in \{0,1,2,...\}$. Solving for the model's equilibrium then translates into finding a set of functions rather than finding infinite sequences of decision variables.

Let us begin the formal description of the recursive competitive equilibrium by characterizing the model's equilibrium conditions. Let $m_t = M_t/P_t$ denote real money holdings, $w_t = W_t/P_t$ and $q_t = Q_t/P_t$ real factor prices, and $\pi_t = P_t/P_{t-1}$ denote inflation. Furthermore, let λ_t denote the co-state variable associated with the household's optimization problem, and let $\omega_t(j)$ be the co-state variable associated with the problem of intermediate-goods-producing firm j. With this notation at hand, the model's equilibrium conditions are given by:

$$0 = c_t^{\frac{1}{\gamma}} \lambda_t [c_t^{\frac{\gamma-1}{\gamma}} + b_t^{\frac{1}{\gamma}} m_t^{\frac{\gamma-1}{\gamma}}] - a_t \tag{7}$$

$$0 = \lambda_t w_t (1-h_t)(1-\tau_t) - \eta \tag{8}$$

$$0 = \beta i_t E_t(\lambda_{t+1}/\pi_{t+1}) - \lambda_t \tag{9}$$

$$0 = m_t \left(1 - \frac{1}{i_t}\right)^{\gamma} - b_t c_t \tag{10}$$

$$\begin{aligned} 0 = \beta E_t \Bigg\{ & \lambda_{t+1}(q_{t+1} + 1 - \delta) - \frac{\phi_k}{2} \lambda_{t+1} \left(\frac{k_{t+2}}{k_{t+1}} - 1\right)^2 \\ & + \phi_k \lambda_{t+1} \left(\frac{k_{t+2}}{k_{t+1}} - 1\right) \left(\frac{k_{t+2}}{k_{t+1}}\right) \Bigg\} - \phi_k \lambda_t \left(\frac{k_{t+1}}{k_t} - 1\right) - \lambda_t \end{aligned} \tag{11}$$

$$
\begin{aligned}
0 &= k_t^{\alpha}[z_t h_t]^{1-\alpha} - y_t && (12)\\
0 &= y_t - w_t h_t - q_t k_t - \frac{\phi_p}{2}\left(\frac{\pi_t}{\pi} - 1\right)^2 y_t - d_t && (13)\\
0 &= \omega_t(1-\alpha)y_t - \lambda_t w_t h_t && (14)\\
0 &= \omega_t \alpha y_t - \lambda_t q_t k_t && (15)\\
0 &= (1-\theta)\lambda_t + \theta\omega_t - \lambda_t\phi_p\Big(\frac{\pi_t}{\pi} - 1\Big)\frac{\pi_t}{\pi} \\
&\qquad +\beta\phi_p E_t\left\{\lambda_{t+1}\Big(\frac{\pi_{t+1}}{\pi} - 1\Big)\frac{\pi_{t+1}}{\pi}\frac{y_{t+1}}{y_t}\right\} && (16)\\
0 &= \rho_i \log\frac{i_{t-1}}{i} + \rho_y \log\frac{y_t}{y} + \rho_\pi \log\frac{\pi_t}{\pi} + \epsilon_{i,t} - \log\frac{i_t}{i} && (17)\\
0 &= c_t + x_t + g + \frac{\phi_k}{2}\left(\frac{x_t}{k_t} - \delta\right)^2 k_t + \frac{\phi_p}{2}\left(\frac{\pi_t}{\pi} - 1\right)^2 y_t - y_t && (18)\\
0 &= k_{t+1} - (1-\delta)k_t - x_t && (19)\\
0 &= (1-\rho_a)\log(a) + \rho_a \log(a_t) + \epsilon_{a,t+1} - \log(a_{t+1}) && (20)\\
0 &= (1-\rho_b)\log(b) + \rho_b \log(b_t) + \epsilon_{b,t+1} - \log(b_{t+1}) && (21)\\
0 &= (1-\rho_z)\log(z) + \rho_z \log(z_t) + \epsilon_{z,t+1} - \log(z_{t+1}) && (22)\\
0 &= (1-\rho_\tau)\log(\tau) + \rho_\tau \log(\tau_t) + \epsilon_{\tau,t+1} - \log(\tau_{t+1}) && (23)
\end{aligned}
$$

A formal derivation is provided in Appendix A. The intuition behind these conditions is the following. Equations (7) -(11) describe optimal household behavior. More precisely, (7) and (8) equate the marginal rate of substitution between labor and consumption to the real after-tax wage, (9) describes the household's indifference between consumption and bond holdings, (10) describes optimal money holdings, and equation (11) states that, in equilibrium, the marginal utility cost of one unit of additional investment at time t equals the discounted expected marginal utility value of its return in period $t+1$. Equations (12) -(16) come from the production side of the model. (12) gives the aggregate production function, (13) characterizes the intermediate firm's budget constraint, (14) and (15) compute the marginal products of labor and capital to their respective factor prices, and (16) descibes the price-setting behavior of firms.[8] Finally, (17) describes the monetary policy

[8]Note that a first-order approximation of (16) leads to the popular New Keynesian Phillips curve equation,

$$\hat{\pi}_t = \frac{\theta - 1}{\phi_p}\hat{\varphi}_t + \beta E_t \hat{\pi}_{t+1},$$

where hatted variables denote percentage deviations from the steady state. $\hat{\varphi}_t = \hat{\lambda}_t\hat{\omega}_t$ describes real marginal costs of production, as can be derived from equations (14) and (15).

rule, (18) denotes the aggregate resource constraint in the economy, (19) defines investment and (20)-(23) characterize the evolution of the exogenous state variables.

More generally, the equilibrium conditions can be summarized as

$$E_t R(f_{t+1}, f_t, , s_{t+1}, s_t, \varepsilon_t; \Theta) = 0. \tag{24}$$

The non-linear function R represents the equilibrium conditions. Its arguments are the state vector $s_t = (k_t, i_{t-1}, a_t, b_t, z_t, \tau_t)'$, the vector of decision variables $f_t = (y_t, c_t, w_t, h_t, q_t, d_t, k_{t+1}, \pi_t, \lambda_t, \omega_t, m_t)'$[9], the vector $\varepsilon_t = (\epsilon_{a,t}, \epsilon_{b,t}, \epsilon_{z,t}, \epsilon_{\tau,t}, \epsilon_{i,t})'$ which collects the exogenous disturbances, and $\Theta = (\beta, \alpha, \delta, \theta, \eta, \phi_k, \phi_p, \gamma, g, \pi, a, b, z, \tau, \rho_a, \rho_b, \rho_z, \rho_\tau, \sigma_a, \sigma_b, \sigma_z, \sigma_\tau, \sigma_r, \rho_r, \rho_y, \rho_\pi)'$, which summarizes the model's structural parameters. Let us denote the variance matrix of ε_t by Σ_ε.
Together with transversality and No-Ponzi conditions, the conditions summarized in (24) are sufficient for equilibrium.[10] We can thus define the symmetric competitive equilibrium of our model recursively as follows:

Definition A recursive symmetric competitive equilibrium is a pair of policy functions, $\{\Phi, \Psi\}$, such that - for every initial state s_0 and exogenous process $\{\varepsilon_t\}_{t=0}^{\infty}$ - the sequences $\{f_t\}_{t=0}^{\infty}$ and $\{s_{t+1}\}_{t=0}^{\infty}$ generated recursively by $f_t = \Phi(s_t)$ and $s_{t+1} = \Psi(s_t, \varepsilon_t)$ satisfy the system of functional equations (24), as well as transversality and No-Ponzi conditions.

3 The linear and quadratic perturbation solutions

From the equilibrium definition stated above it is clear that, to compute the model's equilibrium, we need to solve a rational expectations system of nonlinear functional difference equations. Analytical solutions of such systems are hardly ever feasible, and like in most applications of DSGE models, this is true also for our model. Consequently, we need to resort to numerical methods and work with approximate solutions.
The by far most popular approach to solve DSGE models numerically is by using perturbation methods, in particular, the (log-)linearization of the model around its non-stochastic steady state. Only recently, second-order perturbation solutions have become more popular, as they have been shown

[9] We include k_{t+1} in the vector of control variables to account for the occurrence of k_{t+2} in the equilibrium conditions.

[10] See Stokey and Lucas (1989) for further details.

to deliver substantially better approximations in many applications. The remainder of this section briefly presents the two approaches.

3.1 First-order perturbation

Linearizing a model around its steady state is popular in economics since it is conceptually simple, straightforward to implement on a computer, and fast. In the empirical analysis of DSGE models, an additional advantage is that the linearized model represents a linear state space system. In that case, one can easily conduct likelihood-based inference, since - given the numerical solution of the model - the Kalman filter allows to construct the implied likelihood function analytically.
However, linearization is known to often give only a very poor description of the theoretical model. In particular, models solved by log-linearization are not suited to study economies in which risk has important effects, since they display the certainty equivalence property.[11]
Log-linearization implies laws of motion for our model's variables given by

$$\hat{s}_{t+1} = \tilde{\Psi}_1(\Theta)\hat{s}_t + \Gamma\varepsilon_t, \text{ and} \tag{25}$$

$$\hat{f}_t = \tilde{\Phi}_1(\Theta)\hat{s}_t. \tag{26}$$

The vectors $\hat{s}_t = \log(s_t/s)$ and $\hat{f}_t = \log(f_t/f)$ denote deviations of the respective variables from their steady state values s and f. The matrix Γ consists of zeros and ones, indicating whether a variable is directly hit by a shock or not. The $n_s \times n_s$ matrix $\tilde{\Psi}_1$ and the $n_f \times n_s$ matrix $\tilde{\Phi}_1$ capture the dynamics of the model: the elements of $\tilde{\Psi}_1$ approximate the elasticities of future states with respect to current states, whereby the entries in $\tilde{\Phi}_1$ approximate the elasticities of current decisions with respect to current states.
Both $\tilde{\Psi}_1$ and $\tilde{\Phi}_1$ are functions of the structural parameters of the model. Depending on the exact values of these parameters three cases are possible: there exists a unique solution (determinacy), there exist multiple solutions (indeterminacy), or there exists no solution of the linearized equilibrium system (nonexistence). As is common in the literature, we focus only on the first case and restrict the parameter space to rule out nonexistence or indeterminacy.
When computing the matrices $\tilde{\Psi}_1$ and $\tilde{\Phi}_1$ numerically, we follow the approach by Klein (2000), since it is particularly easy to implement.[12] Alternatively,

[11] This feature makes them inadequate for welfare analysis under uncertainty and for studying time-varying volatilities. See Kim and Kim (2003) and Fernandez-Villaverde and Rubio-Ramirez (forthcoming).

[12] Details can be found in our MATLAB codes which are available upon request.

we could use the approaches by Blanchard and Kahn (1980), Uhlig (1999), or Sims (2002), among others.

3.2 Second-order perturbation

Second-order perturbation extends linearization techniques by capturing second-order terms in the policy functions. The resulting law of motion for the jth state variable is given by

$$\hat{s}_{t+1}^{(j)} = \hat{\Psi}_0^{(j)}(\Theta) + \hat{\Psi}_1^{(j)}(\Theta)\hat{s}_t + \frac{1}{2}\hat{s}_t'\hat{\Psi}_2^{(j)}(\Theta)\hat{s}_t + \Gamma\varepsilon_t. \tag{27}$$

Similarly, the kth decision variable evolves according to

$$\hat{f}_{t+1}^{(k)} = \hat{\Phi}_0^{(k)}(\Theta) + \hat{\Phi}_1^{(k)}(\Theta)\hat{s}_t + \frac{1}{2}\hat{s}_t'\hat{\Phi}_2^{(k)}(\Theta)\hat{s}_t. \tag{28}$$

The vector $\hat{\Psi}_1^{(j)}$ corresponds exactly to the jth row in the matrix $\tilde{\Psi}_1$ obtained through linearization, and $\hat{\Phi}_1^{(k)}$ corresponds to the kth row in $\tilde{\Phi}_1$. The $n_s \times n_s$ matrices $\hat{\Psi}_2^{(j)}$ and $\hat{\Phi}_2^{(k)}$ capture the second-order terms of the policy functions. The constants $\hat{\Psi}_0^{(j)}(\Theta)$ and $\hat{\Phi}_0^{(k)}(\Theta)$ correct for precautionary behavior. Consequently, the unconditional means of the model's variables do no longer coincide with their steady state values in the case of the quadratic solution.

As is the case for linearization, several methods are available to numerically compute the matrices of the second-order approximation. Popular algorithms include Collard and Juillard (2001), Swanson, Anderson, and Levin (2003), and Schmitt-Grohe and Uribe (2004). We follow the latter approach.

4 The state-space representation

Assume that we observe time series of length T on N_Y macroeconomic variables, summarized by $\mathcal{Y}^T = \{\mathcal{Y}_t\}_{t=1}^T$.[13] Together with the data, our model forms a nonlinear state-space system:

$$X_{t+1} = \mathcal{H}(X_t, \varepsilon_t; \Theta) \qquad (transition\,equation) \tag{29}$$

$$\mathcal{Y}_t = \mathcal{G}(X_t, \nu_t) \qquad (measurement\,equation) \tag{30}$$

$X_t = (s_t',\, f_t')'$ collects the model's (state and decision) variables, whereas $\mathcal{H}$ is a nonlinear function that can be constructed from Ψ and Φ. $\mathcal{G}$ is a mapping which relates the model's variables to the observables. For the sake of

[13] Let us for the sake of notational simplicity introduce $\mathcal{Y}^0 = \emptyset$.

notational simplicity we make the common (and non-restrictive) assumption that the data are linear transformations of the model's variables, and hence $\mathcal{G}$ is a linear function. Finally, ν_t denotes a vector of measurement errors. For simplicity we will assume that $\nu_t \sim N(0, \Sigma_\nu)$, where Σ_ν is diagonal. Since the functions Ψ and Φ cannot be derived analytically, we have no closed-form solution for $\mathcal{H}$. However, using the approximate model solutions described in the previous section, we can again construct linear and quadratic approximations.

The linear state-space model
Recall that the first-order accurate solution of the model is given by the linear policy functions (25) and (26). Introducing $\hat{X}_t = (\hat{s}_t',\ \hat{f}_t')'$ and $\mathcal{G}(X_t, \nu_t) = \mathrm{G}X_t + \nu_t$ we can thus recast the linearized model in state-space form:

$$\begin{aligned} X_t &= \mathrm{H}(\Theta)X_{t-1} + \mathrm{J}(\Theta)\varepsilon_t & (31)\\ \mathcal{Y}_t &= \mathrm{G}X_t + \nu_t & (32) \end{aligned}$$

The matrices H and J can be constructed in a straightforward way from $\tilde{\Psi}_1$, $\tilde{\Phi}_1$ and Γ.

The quadratic state-space model
The model's second-order accurate policy functions (27) and (28) imply a quadratic transition equation, such that the model in state-space form is given by:

$$\begin{aligned} X_t &= \tilde{H}(X_{t-1}, \varepsilon_t; \Theta), & (33)\\ \mathcal{Y}_t &= \mathrm{G}X_t + \nu_t & (34) \end{aligned}$$

$\tilde{H}$ is a quadratic function that follows directly from (27) and (28).

5 Maximum Likelihood estimation

The state-space representations presented above can be used, together with filtering techniques, to conduct likelihood based inference. Being precise, we can estimate the model's structural parameters, Θ, as well as the measurement error variances, Σ_ν, by Maximum Likelihood or Bayesian methods. Our paper follows the classical approach.
Let us first describe how to construct the likelihood of a data sample $\mathcal{Y}^T$ for given Θ and Σ_ν. In the case of the linearized model, this will be achieved through the Kalman filter. In the quadratic case, we will use the particle filter as suggested by Fernandez-Villaverde and Rubio-Ramirez (forthcoming),

whereby we closely follow the implementation by An and Schorfheide (forthcoming). In particular, we heavily build on computer code developed by the latter authors. Finally, we briefly discuss the numerical maximization of the likelihood function with respect to Θ and Σ_ν.

5.1 The Kalman filter

Equations (31) and (32) constitute a linear state-space system with Gaussian errors. Hence, for given values of Θ and Σ_ν we can use the Kalman filter to construct the likelihood of the sample $\mathcal{Y}^T = \{\mathcal{Y}_t\}_{t=1}^T$,

$$\mathcal{L}(\mathcal{Y}^T|\Omega) = \prod_{t=1}^{T} p(\mathcal{Y}_t|\mathcal{Y}^{t-1};\Omega), \tag{35}$$

where $\Omega = \{\Theta, \Sigma_\nu\}$. An extensive description (including proofs) of the Kalman filter is provided in many econometrics textbooks, for example, in Chapter 13 of Hamilton (1994). We thus confine ourselves here to a brief illustration, which serves mainly to easy the comparison between the Kalman filter and the particle filter presented in the following section.
Our implementation of the Kalman filter consists of the following steps:

1. **Initialization.** We start by deriving a predictor of the first state, $X_{1|0}$[14], and an estimate of the corresponding prediction error covariance matrix, $\Sigma_{1|0}^X = E[(X_1 - X_{1|0})(X_1 - X_{1|0})']$. We follow the standard approach for stationary processes which builds on the steady state of the system. Being precise, we set $X_{1|0} = X^*$ and $\Sigma_{1|0}^X = \Sigma^*$ such that $X^* = \mathrm{H}X^*$ and

$$\Sigma^* = \mathrm{H}\Sigma^*\mathrm{H}' + \mathrm{J}\Sigma_\varepsilon\mathrm{J}' = [I - \mathrm{H}\otimes\mathrm{H}]^{-1}vec(\mathrm{J}\Sigma_\varepsilon\mathrm{J}').$$

   Then we set $t = 1$ and proceed.

2. **Forecasting.** In the beginning of period t, we use the state predictor $X_{t|t-1}$ together with the measurement equation (32) to compute the best linear predictor for $\mathcal{Y}_t$,

$$\mathcal{Y}_{t|t-1} = E(\mathcal{Y}_t|\mathcal{Y}^{t-1}) = \mathrm{G}X_{t|t-1}.$$

   After having observed $\mathcal{Y}_t$, we construct the forecast error

$$u_t = \mathcal{Y}_t - \mathcal{Y}_{t|t-1} = \mathcal{Y}_t - \mathrm{G}X_{t|t-1} = \nu_t + \mathrm{G}(X_t - X_{t|t-1}).$$

[14] In general, we use the notation $Z_{t|s} = E(Z_t|\mathcal{Y}^s)$ to denote the best predictor for a variable Z at time t, Z_t, conditional upon information available at time s.

Since the system (31),(32) is linear with Gaussian errors, u_t is normally distributed with mean zero and variance matrix $\Sigma_t^u = \Sigma_\nu + \mathrm{G}\Sigma_{t|t-1}^X \mathrm{G}'$. Furthermore, since $\mathcal{Y}_{t|t-1}$ is by construction the best linear predictor of $\mathcal{Y}_t$ given $\mathcal{Y}^{t-1}$, it follows that

$$p(\mathcal{Y}_t|\mathcal{Y}^{t-1};\Omega) = \frac{1}{\sqrt{(2\pi)^{N_Y}|\Sigma_t^u|}} \exp\left(-\frac{u_t'(\Sigma_t^u)^{-1}u_t}{2}\right). \tag{36}$$

Expression (36) points out that, in order to compute the likelihood (35), we need to compute the sequences $\{u_t\}_{t=1}^T$ and $\{\Sigma_t^u\}_{t=1}^T$. This requires to keep track of the state predictors and their variances, in particular, at every point in time t we need to compute $X_{t+1|t}$ and $\Sigma_{t+1|t}^X$ from $X_{t|t-1}$ and $\Sigma_{t|t-1}^X$.

3. **Updating.** The crucial step in deriving $X_{t+1|t}$ and $\Sigma_{t+1|t}^X$ is the updating of the state predictor, i.e., to compute the best predictors $X_{t|t}$ and $\Sigma_{t|t}^X$, from $X_{t|t-1}$, $\Sigma_{t|t-1}^X$, and the observed data $\mathcal{Y}_t$. As shown by Kalman (1960), this can be achieved according to

$$\begin{aligned} X_{t|t} &= X_{t|t-1} + K_t(\mathcal{Y}_t - \mathcal{Y}_{t|t-1}) = X_{t|t-1} + K_t(\mathcal{Y}_t - \mathrm{G}X_{t|t-1}), \\ \Sigma_{t|t}^X &= \Sigma_{t|t-1}^X - K_t\mathrm{G}\Sigma_{t|t-1}^X, \end{aligned}$$

where

$$K_t = \Sigma_{t|t-1}^X \mathrm{G}'(\mathrm{G}\Sigma_{t|t-1}^X\mathrm{G}' + \Sigma_\nu)^{-1}.$$

K_t is denoted the Kalman gain matrix. Now that $X_{t|t}$ and $\Sigma_{t|t}^X$ have been derived, we can obtain $X_{t+1|t} = \mathrm{H}X_{t|t}$ and $\Sigma_{t+1|t}^X = \mathrm{H}\Sigma_{t|t}^X\mathrm{H}' + \mathrm{J}\Sigma_\varepsilon\mathrm{J}'$ in a straightforward way.
Unless $t = T$ we set $t = t + 1$ and return to step 2.

4. **Likelihood construction.** We use the sequences $\{u_t\}_{t=1}^T$ and $\{\Sigma_t^u\}_{t=1}^T$ to construct the likelihood as

$$\mathcal{L}(\mathcal{Y}^T|\Theta,\Sigma_\nu) = \prod_{t=1}^T \frac{1}{\sqrt{(2\pi)^{N_Y}|\Sigma_t^u|}} \exp\left(-\frac{u_t'(\Sigma_t^u)^{-1}u_t}{2}\right). \tag{37}$$

5.2 The particle filter

The Kalman filter can no longer be applied to build the likelihood in a non-linear and/or non-Gaussian environment. Consequently, we need to resort to different methods to construct the likelihood implied by the quadratic model

(33) and (34). Fernandez-Villaverde and Rubio-Ramirez (forthcoming) suggest to use Sequential Monte Carlo methods, and introduce a particle filter that allows to obtain the likelihood function of a DSGE model solved with a non-linear method. We follow their approach and proceed as follows.
We start by rewriting the likelihood function (35) as

$$\mathcal{L}(\mathcal{Y}^T|\Omega) = \prod_{t=1}^{T} \int p(\mathcal{Y}_t|X_t, \mathcal{Y}^{t-1}; \Omega) p(X_t|\mathcal{Y}^{t-1}; \Omega) dX_t. \tag{38}$$

For the quadratic model we cannot compute this expression analytically, since a closed form solution for $p(X_t|\mathcal{Y}^{t-1};\Omega)$ is infeasible. However, we can use simulation methods to derive an approximation. More precisely, we can construct draws $\{\tilde{x}^i_{t|t-1}\}_{i=1}^N$ (a so called *swarm of particles*, hence the name *particle filter*) from each density in the sequence $\{p(X_t|\mathcal{Y}^{t-1};\Omega)\}_{t=1}^T$, which - by the law of large numbers - allows us to approximate the likelihood (38) with

$$\mathcal{L}(\mathcal{Y}^T|\Omega) \approx \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y}_t|\tilde{x}^i_{t|t-1}, \mathcal{Y}^{t-1}; \Omega). \tag{39}$$

Similar to the Kalman filter, this involves the following steps.

1. **Initialization.** We draw initial particles $\{\tilde{x}^i_{1|0}\}_{i=1}^N$ from the density $p(X_1;\Omega)$. We accomplish this in the following way, which is well documented in An (2005). First we generate $\{\tilde{x}^i_{-1}\}_{i=1}^N$, where $\tilde{x}^i_{-1}$ is set to the non-stochastic steady state for all $i = 1, ..., N$. Then we draw N realizations of the model's innovations, ε^i_0, $i = 1, ..., N$, and use the transition equation (33) to construct $\{\tilde{x}^i_{0|0}\}_{i=1}^N$. Since

$$p(X_{t+1}|\mathcal{Y}^t;\Omega) = \int p(X_{t+1}|X_t;\Omega) p(X_t|\mathcal{Y}^t;\Omega) dX_t \approx \frac{1}{N} \sum_{i=1}^{N} p(X_{t+1}|\tilde{x}^i_{t|t}) \tag{40}$$

   we can then use $\{\tilde{x}^i_{0|0}\}_{i=1}^N$ together with the transition equation (33) to sample N draws $\{\tilde{x}^i_{1|0}\}_{i=1}^N$ from the conditional density $p(X_1|\Omega)$. This amounts to generating one draw from $p(X_{t+1}|\tilde{x}^i_{t|t})$ for each i.
   We set $t = 1$ and proceed.

2. **Forecasting.** In the beginning of period t, we use the particles $\{\tilde{x}^i_{t|t-1}\}_{i=1}^N$ to construct forecasts $\{\mathcal{Y}^i_{t|t-1}\}_{i=1}^N$, where

$$\mathcal{Y}^i_{t|t-1} = \mathrm{G}\tilde{x}^i_{t|t-1}.$$

After having observed $\mathcal{Y}_t$, we construct the prediction errors $\{\xi_t^i\}_{i=1}^N$, where

$$\xi_t^i = \mathcal{Y}_t - \mathcal{Y}_{t|t-1}^i.$$

Note that, conditional on the state particle $\tilde{x}_{t|t-1}^i$, the uncertainty in ξ_t^i comes solely from measurement errors. Therefore, we can use $\{\xi_t^i\}_{i=1}^N$ to compute the densities $p(\mathcal{Y}_t|\tilde{x}_{t|t-1}^i, \mathcal{Y}^{t-1}; \Omega)$, $i = 1, ..., N$, according to

$$p(\mathcal{Y}_t|\tilde{x}_{t|t-1}^i, \mathcal{Y}^{t-1}; \Omega) = \frac{1}{\sqrt{(2\pi)^{N_Y}|\Sigma_\nu|}} \exp\left(-\frac{\xi_{t,i}'(\Sigma_\nu)^{-1}\xi_t}{2}\right). \quad (41)$$

This expression shows that, to construct the approximate likelihood function (42), we need particles $\{\tilde{x}_{t|t-1}^i\}_{i=1}^N$ from the densities $p(X_t|\mathcal{Y}^{t-1}; \Omega)$ for all $t = 1, ..., T$.
In the following step we illustrate how to use the particles $\{\tilde{x}_{t|t-1}^i\}_{i=1}^N$ together with the observations $\mathcal{Y}_t$ to draw particles $\{\tilde{x}_{t+1|t}^i\}_{i=1}^N$ from $p(X_{t+1}|\mathcal{Y}^t; \Omega)$.

3. **Updating.** We continue by computing the normalized weights

$$q_t^i = \frac{p(\mathcal{Y}_t|\tilde{x}_{t|t-1}^i, \mathcal{Y}^{t-1}; \Omega)}{\sum_{i=1}^N p(\mathcal{Y}_t|\tilde{x}_{t|t-1}^i, \mathcal{Y}^{t-1}; \Omega)}.$$

Note that for every particle $\tilde{x}_{t|t-1}^i$, q_t^i gives the relative likelihood of the observation $\mathcal{Y}_t$, conditional on the particle and on past observations. Then we generate particles $\{\tilde{x}_{t|t}^i\}_{i=1}^N$ by sampling N times with replacement from $\{\tilde{x}_{t|t-1}^i\}$ using the weights $\{q_t^i\}_{i=1}^N$. That is, we construct $\{\tilde{x}_{t|t}^i\}_{i=1}^N$ such that[15]

$$Prob\left(\tilde{x}_{t|t}^i = \tilde{x}_{t|t-1}^i\right) = q_t^i.$$

As emphasized in Corollary 5 of Fernandez-Villaverde and Rubio-Ramirez (forthcoming), the particles $\{\tilde{x}_{t|t}^i\}_{i=1}^N$ are draws from $p(X_t|\mathcal{Y}^t; \Omega)$. Due to (40), we can then use $\{\tilde{x}_{t|t}^i\}_{i=1}^N$ together with the transition equation (33) to generate particles $\{\tilde{x}_{t+1|t}^i\}_{i=1}^N$ from $p(X_{t+1}|\mathcal{Y}^t; \Omega)$.
Unless $t = T$ we set $t = t + 1$ and return to step 2.

4. **Likelihood construction.** We use the sequence $\{\{\xi_t^i\}_{i=1}^N\}_{t=1}^T$ to construct the likelihood as

$$\mathcal{L}(\mathcal{Y}^T|\Omega) \approx \prod_{t=1}^T \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{(2\pi)^{N_Y}|\Sigma_\nu|}} \exp\left(-\frac{\xi_t^{i\prime}(\Sigma_\nu)^{-1}\xi_t^i}{2}\right). \quad (42)$$

[15] We follow the suggestion by Fernandez-Villaverde and Rubio-Ramirez (forthcoming) and An (2005) and use the deterministic resampling approach by Kitagawa (1996).

5.3 Maximizing the likelihood function

Once we have constructed the likelihood function, either from the linear or the quadratic model, we need to numerically maximize it with respect to the parameters Ω. Often this turns out to be a daunting task, since the likelihood function of a DSGE model typically has many local maxima. Numerical maximization routines, in particular Gradient-based methods, are known to have serious difficulties not to get stuck in a local maximum. We therefore use a combination of the *simulated annealing* approach proposed by Goffe, Ferrier, and Rogers (1994) and the MATLAB built-in function *fminsearch* to maximize the likelihood function.[16]

6 The forecasting exercise

We start our forecasting exercise by generating artificial data from the non-linear New Keynesian model. To this end, we first calibrate the model. Our parameter choices are summarized in Table 1, and briefly discussed in Appendix C. Then we solve the model numerically using a third-order Galerkin projection method. With the solution at hand, we simulate the model and construct 100 data sets, each of which consists of 96 observations on five macroeconomic variables: output, real money holdings, inflation, nominal interest rates, and the labor tax rate.[17] Similarly, we construct 100 data sets from a VAR(2) model. Further details on the construction of our artificial data are provided in Appendix C.

The first 80 observations of each data set are used to estimate four competing models: the linear and quadratic versions of our New Keynesian DSGE model, and the following VAR(1) and VAR(2) models:

$$\hat{\mathcal{Y}}_t = A^1\hat{\mathcal{Y}}_{t-1} + v_t^1, \quad v_t^1 \sim N(0, \Sigma_1^V) \tag{43}$$

$$\hat{\mathcal{Y}}_t = A_1^2\hat{\mathcal{Y}}_{t-1} + A_2^2\hat{\mathcal{Y}}_{t-2} + v_t^2, \quad v_t^2 \sim N(0, \Sigma_2^V) \tag{44}$$

$\hat{\mathcal{Y}}_t$ denotes deviations of the data from their unconditional sample mean. Furthermore, Σ_1^V and Σ_2^V are assumed to be diagonal. We estimate the VAR

[16] Metaphorically speaking, we use simulated annealing to *find the highest mountain*, and use fminsearch to *climb up to its top*.

[17] Unfortunately, using 100 totally different data sets would go beyond our computing capacities, since the particle filter estimation is very time consuming and we cannot repeat it so many times. We thus use the following shortcut: we generate five realizations of the first 80 observations (which will be used for estimation), and for each of these, we construct 20 different realizations of the subsequent 16 periods. Obviously this is somewhat suboptimal, however, we do not find a better approach given our endowment with computing facilities.

parameters by Least Squares.
For the New Keynesian models, the parameters to be estimated are the structural parameters, Θ, and the measurement error variances, Σ_ν. As is common in the literature[18], we fix some parameters (which are typically weakly identified by the data) prior to estimation. These are the labor parameter η, the elasticity of substitution between intermediate goods, θ, and the parameters associated with capital, α, δ, and ϕ_k. The remaining 26 parameters are estimated by numerically maximizing the likelihood function implied by the Kalman and the particle filter, respectively. For the particle filter, we follow Fernandez-Villaverde and Rubio-Ramirez (forthcoming) and use $N = 80.000$ particles.
We then use each of the four models to compute 1 to 16 step ahead forecasts of the last 16 observations in our samples. The forecasts of the New Keynesian models are based on the final state predictor, $X_{T|T}$, implied by the Kalman and the particle filter, respectively. Obviously, the accuracy of the state prediction is crucial for the accuracy of the forecasts generated from the model. Therefore, once the parameters have been estimated using 80.000 particles, we increase the number of particles to 400.000 to compute $X_{T|T}$ from the quadratic model. This generates more precise estimates of the state, and since it has to be done only once after each estimation algorithm, it does not create a remarkable computational burden.
We evaluate the forecasts of all models using both univariate and multivariate measures. Our univariate measures of forecast accuracy are the mean forecast errors,

$$MFE(i,h) = \frac{1}{N^f}\sum_{s=1}^{N^f} e_s^i(h), \tag{45}$$

together with the root mean squared forecast errors,

$$RMSFE(i,h) = \sqrt{\frac{1}{N^f}\sum_{s=1}^{N^f} e_s^i(h)^2}. \tag{46}$$

In the above expressions, we use the following notation. N^f gives the overall number of available h step ahead forecasts ($N^f = 100$ in our application), $e_s(h)$ gives the vector of h-step ahead forecast errors from the data set indexed by s. The index $i = 1, ..., N_Y$ indicates the forecasted variable, and $e_s^i(h)$ denotes the ith element in the vector $e_s(h)$.
Our multivariate measures are the log-determinant statistics suggested by

[18] See, for example, Ireland (2004a).

Adolfson, Linde, and Villani (2005). The log-determinant statistic to evaluate the h step-ahead forecasting performance is given by $\log|\Omega(h)|$, where

$$\Omega(h) = \frac{1}{N^f}\sum_{s=1}^{N^f}\tilde{e}_s(h)\tilde{e}_s(h)', \tag{47}$$

$\tilde{e}_s(h) = M^{-1/2}e_s(h)$, and M denotes a positive definite scaling matrix. In our analysis, we use a diagonal scaling matrix with the average sample variances in our artificial data as diagonal elements.[19]

7 Results

We now present the results of our forecasting exercise. We first outline the case where the data are generated from the nonlinear New Keynesian DSGE model, and later present the case where the DGP is a vector autoregression.

7.1 The DGP is the nonlinear New Keynesian model

We first consider the scenario where the artificial data are generated from the nonlinear New Keynesian model, more precisely, the New Keynesian model solved by a third-order Galerkin projection method. Details on the construction of the data can be found in Appendix C.

We start the analysis by inspecting the parameter estimates associated with the New Keynesian models. Table 2 presents the minimum, maximum and average parameter estimates obtained from the linearized model, together with the true values which were used for simulation. The corresponding estimates obtained from the quadratic model are documented in Table 3. We observe that most of the parameters are reasonably well identified, and that the point estimates of the structural parameters, Θ, are very similar for the linearized and the quadratic model. The observation error variances, Σ_ν, however, are more precisely estimated by the particle filter used for the quadratic model.

The forecasting performances of our four competing models are illustrated in Figure 1 and Table 4. Figure 1 plots the absolute values of the mean forecast errors (column one), as well as root mean squared forecast errors (column two). Our first observation is that, compared to the linearized

[19] Alternatively, we could study the trace of $\Omega(h)$ instead of its determinant. However, since we use a diagonal scaling matrix, the trace statistic would reduce to a simple weighted average of the individual mean squared forecast errors, thus not incorporating information on correlation of forecast errors associated with different series. We therefore prefer the log-determinant statistic.

New Keynesian model, the quadratic model gives better predictions for the money series. This holds for both the average and the mean squared forecast errors, and for all forecast horizons. Furthermore, it yields better predictions for the labor tax rate, which is particularly interesting since the tax rate is a series with linear dynamics.[20] Finally, the quadratic model appears to give slightly better forecasts for output, whereas it does not provide better predictions for inflation and the interest rate, which are roughly equally well predicted by the linear model.
Comparing the New Keynesian model to the two VAR benchmarks by univariate measures illustrates the good predictive abilities of DSGE models. According to the RMSFE, the quadratic model beats both VAR models at short horizons, and does roughly equally well at longer horizons.

Let us now turn to the multivariate measures of forecast accuracy, summarized in Table 4. The log-determinant statistics display a clear winner among our four models: the quadratic New Keynesian model. It outperforms its three competitors at all forecast horizons. Interestingly, the relative advantage of the quadratic model over its linearized counterpart, as measured by the log-determinant statistics, arguably increases with the forecast horizon. This is particularly relevant for practitioners who use dynamic stochastic general equilibrium models for long-run forecasting.

Finally, let us close this section by briefly discussing the size of quality improvements that are achieved by moving to second-order approximations. At first sight, the differences in our accuracy measures seem not very big in magnitude. However, when comparing these numbers one has to keep in mind that our version of the New Keynesian model is very close to being linear. Our results show that even in this almost linear environment, improvements from second-order approximation are feasible. Most models used by practitioners are much *more nonlinear*, e.g., due to more complex utility functions or the presence of a broad variety of non-additive shocks. Within these environments, it is to be expected that accounting for nonlinearities will much more noticeably improve the predictive abilities of DSGE models.

7.2 The true DGP is a VAR(2) model

Let us now consider the case where the data generating process is given by a VAR(2) model. The exact description of the model is given in Appendix

[20]This result may be puzzling at first sight, however, the intuition behind it is straightforward. The quadratic model delivers on average more precise estimates of the parameters describing the evolution of tax rates, and delivers on average better predictions of the state of the system, as compared to the linearized model. Taken together, this results in better forecasting properties, even for the linear tax rate series.

C, equation (70). Again, we start by inspecting the parameter estimates of the linear and quadratic New Keynesian models, summarized in Tables 5 and 6. The misspecification of the New Keynesian model is obvious by mere inspection: the estimation algorithms of both the linear and quadratic model deliver unreasonable estimates for several parameters. The government spending parameter g, for example, is estimated at a unreasonably low value of almost zero. Similarly, the preference shock parameter a is in general estimated at an unreasonably high value.

Despite misspecification, the New Keynesian models turn out to have very good predictive abilities. This is pointed out by the MFE and RMSFE measures visualized in Figure 2, as well as the log-determinant measures summarized in Table 7. We observe that the vector autoregressive models now perform relatively well in forecasting output and real balances in the short-run, whereas they perform rather poor in the medium and the long run. Interestingly, the New Keynesian model seems to have some difficulties in predicting the money series in the short run, which holds particularly for the second-order approximated model. This suggests that the short run dynamics of real balances, generated by the VAR model, are at odds with the economic theory embedded in the New Keynesian model. Apart from that, however, the NK model seems flexible enough to fit the dynamics in the data, since it delivers forecasts that are often better than the corresponding forecasts from vector autoregressions.

Comparing the linear to the quadratic version of the New Keynesian model by univariate measures does not hint to whether nonlinearities play an important role in forecasting. Whereas the quadratic model is better in forecasting output, the linear model now provides better forecasts for the money series. The remaining series, i.e., inflation, the interest rate, and the tax rate, are almost equally well forecasted by both models. Studying the multivariate measures of forecast accuracy delivers sharper results. A pairwise comparison between our four models reveals that, overall, the quadratic New Keynesian model still features the best predictive abilities. According to the multivariate measures, it outperforms the VAR(2) model at 14 out of the 16 forecast horizons considered in our analysis, and dominates the VAR(1) model at all horizons. This result emphasizes that, even when the data are from a VAR model, DSGE models may generate better forecasts since they can often be more precisely estimated. Finally, the quadratic model also outperforms the linearized New Keynesian model. It provides better forecasts for 14 out of 16 horizons. In particular, it delivers again the best results for long forecast horizons.

7.3 Implications of our results

The previous subsections have outlined two scenarios where accounting for nonlinearities helps to improve the forecasting performance of a DSGE model. What are the implications of these results for practical applications such as Smets and Wouters (2004) and Adolfson, Linde, and Villani (2005)? Should these authors use quadratic instead of linear approximations to improve their models' forecasting properties?

Presuming that the models used by practitioners are very well specified, and that most models used are much more nonlinear than the model analyzed in this paper, our answer is yes. In order to fully exploit the model's predictive abilities, researchers should not restrict attention to linearized economies but use quadratic models.[21] It is important, however, not to believe that quadratic models always deliver better results than linearized models. When the model is at odds with the dynamics of the data, using quadratic approximations may deteriorate the fit and forecasting performance of the model. This is due to the low numerical efficiency of the particle filter under misspecification. To illustrate the problem, Table 8 summarizes the *effective sample size* measures, proposed by Fernandez-Villaverde and Rubio-Ramirez (forthcoming) to check for a depletion of the sample problem in the particle filter. We observe that, when our data are taken from the VAR(2) model, the *effective sample size* is substantially lower as compared to when the data are generated from the New Keynesian model. In our application, this has reduced the relative advantage of the quadratic over the linearized model. In other applications, it could easily reverse the ranking of linear and quadratic models. Either way, in practical applications one should (if necessary) address this problem by using substantially more particles for estimating the quadratic model, or by resorting to numerically more efficient methods than the basic particle filter applied in this paper.

8 Summary and conclusions

This paper has analyzed the predictive abilities of a New Keynesian DSGE model. Using artificial time series generated from the New Keynesian model and a VAR(2) model, we have demonstrated that DSGE models exhibit very good predictive abilities, even when they are (obviously) misspecified. Furthermore, and more importantly, we have shown that capturing second-order

[21] The additional computational burden associated with this approach should not deter practitioners, since particle filtering is fast enough to be implemented on good desktop PCs and to be applied to the class of models needed for serious policy analysis. See Fernandez-Villaverde and Rubio-Ramirez (forthcoming).

terms in the policy functions improves the New Keynesian model's forecasting performance. Our results were derived in an almost linear environment, such that we are confident that they carry over to a broad variety of dynamic stochastic general equilibrium models. Our findings thus suggest that, in order fully exploit the predictive abilities of DSGE models, practitioners should not restrict attention to linearized economies.

References

Adolfson, M., J. Linde, and M. Villani (2005): "Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model," Working Paper Series 190, Sveriges Riksbank (Central Bank of Sweden).

Amisano, G., and O. Tristani (2006): "Euro area inflation persistence in an estimated nonlinear," Computing in Economics and Finance 2006 347, Society for Computational Economics.

An, S. (2005): "Bayesian Estimation of DSGE Models: Lessons from Second-Order Approximations," Discussion paper, University of Pennsylvania.

An, S., and F. Schorfheide (forthcoming): "Bayesian Analysis of DSGE Models," *Econometric Reviews*.

Blanchard, O. J., and C. M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–11.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Clements, M. P., and D. F. Hendry (1998): *Forecasting Economic Time Series*. Cambridge University Press.

——— (1999): *Forecasting Non-Stationary Economic Time Series*. The MIT press.

Collard, F., and M. Juillard (2001): "A Higher-Order Taylor Expansion Approach to Simulation of Stochastic Forward-Looking Models with an Application to a Nonlinear Phillips Curve Model," *Computational Economics*, 17(2-3), 125–39.

Del Negro, M., and F. Schorfheide (2003): "Take your model bowling: forecasting with general equilibrium models," *Economic Review*, (Q4), 35–50.

——— (2004): "Priors from General Equilibrium Models for VARS," *International Economic Review*, 45(2), 643–673.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2004): "On the fit and forecasting performance of new Keynesian models," Working Paper 2004-37, Federal Reserve Bank of Atlanta.

Dib, A., M. Gammoudi, and K. Moran (2006): "Forecasting Canadian Time Series with the New Keynesian Model," Working Papers 06-4, Bank of Canada.

Diebold, F. X. (1998): "The Past, Present, and Future of Macroeconomic Forecasting," *Journal of Economic Perspectives*, 12(2), 175–92.

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez (2005): "Estimating dynamic equilibrium economies: linear versus nonlinear likelihood," *Journal of Applied Econometrics*, 20(7), 891–910.

——— (forthcoming): "Estimating Macroeconomic Models: A Likelihood Approach," *The Review of Economic Studies*.

Fernandez-Villaverde, J., J. F. Rubio-Ramirez, and M. S. Santos (2006): "Convergence Properties of the Likelihood of Computed Dynamic Models," *Econometrica*, 74(1), 93–119.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994): "Global optimization of statistical functions with simulated annealing," *Journal of Econometrics*, 60(1-2), 65–99.

Hamilton, J. D. (1994): *Time series analysis*. Princeton University Press, Princeton, NJ.

Hendry, D. F. (1995): "Econometrics and Business Cycle Empirics," *Economic Journal*, 105(433), 1622–36.

Ireland, P. N. (1997): "A small, structural, quarterly model for monetary policy evaluation," *Carnegie-Rochester Conference Series on Public Policy*, 47, 83–108.

——— (2004a): "A method for taking models to the data," *Journal of Economic Dynamics and Control*, 28(6), 1205–1226.

——— (2004b): "Technology Shocks in the New Keynesian Model," *The Review of Economics and Statistics*, 86(4), 923–936.

Judd, K. L. (1992): "Projection methods for solving aggregate growth models," *Journal of Economic Theory*, 58(2), 410–452.

Kalman, R. E. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D), 35–45.

Kim, J. (2000): "Constructing and estimating a realistic optimizing model of monetary policy," *Journal of Monetary Economics*, 45(2), 329–359.

Kim, J., and S. H. Kim (2003): "Spurious welfare reversals in international business cycle models," *Journal of International Economics*, 60(2), 471–500.

Kitagawa, G. (1996): "Monte Carlo Filter and Smoother for Non-Gaussian Non-linear State Space Models," *Journal of Computational and Graphical Statistics*, 5(1), 1–25.

Klein, P. (2000): "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.

Schmitt-Grohe, S., and M. Uribe (2004): "Solving dynamic general equilibrium models using a second-order approximation to the policy function," *Journal of Economic Dynamics and Control*, 28(4), 755–775.

Sims, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2), 1–20.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

——— (2004): "Forecasting with a Bayesian DSGE Model: An Application to the Euro Area," *Journal of Common Market Studies*, 42(4), 841–867.

Stokey, N. L., and R. E. Lucas (1989): *Recursive Methods in Economic Dynamics* . Harvard University Press.

Swanson, E., G. Anderson, and A. Levin (2003): "Higher-Order Solutions to Dynamic, Discrete-Time Rational Expectations Models: Methods and an Application to Optimal Monetary Policy," Discussion Paper 64, Society for Computational Economics.

Uhlig, H. (1999): "A toolkit for analysing nonlinear dynamic stochastic models easily," in *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimon, and A. Scott, pp. 30–61. Oxford University Press.

WOODFORD, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

A Derivation of equilibrium conditions

This appendix derives the model's equilibrium conditions from first-order conditions associated with the agents' optimization problems, as well as a monetary policy rule and the stochastic laws of motion for the exogeneous variables.

Let us first consider equations (7)-(11), and (18)-(19). Using the method of Lagrange, we can derive them from the household's problem. The Lagrangian is given

$$\begin{aligned}\mathcal{L}_h \;= E_0 \;& \sum_{t=0}^{\infty}\beta^t\Bigg[a_t\frac{\gamma}{\gamma-1}\log\left(c_t^{\frac{\gamma-1}{\gamma}}+b_t^{\frac{1}{\gamma}}\left(M_t/P_t\right)^{\frac{\gamma-1}{\gamma}}\right)+\eta\log(1-h_t)\Bigg] \\ & +\lambda_t\Bigg[\frac{M_{t-1}+B_{t-1}+(1-\tau_t)W_th_t+Q_tk_t+D_t+T_t}{P_t}- \\ & \left(c_t+k_{t+1}-(1-\delta)k_t+\frac{\phi_k}{2}\Big(\frac{k_{t+1}}{k_t}-1\Big)^2k_t+\frac{B_t/i_t+M_t}{P_t}\right)\Bigg]\end{aligned} \tag{48}$$

where λ_t denotes the Lagrange multiplier. Differentiating the Lagrangian with respect to the household's decision variables, i.e. c_t, h_t, M_t/P_t, B_t/P_t, k_{t+1}, and with respect to the Lagrange multiplier, λ_t, yields the following system of first order conditions, which must hold for all time periods $t \in \{0,1,2,...\}$:

$$0 = a_t c_t^{-\frac{1}{\gamma}} - \lambda_t\left[c_t^{\frac{\gamma-1}{\gamma}}+b_t^{\frac{1}{\gamma}}\left(\frac{M_t}{P_t}\right)^{\frac{\gamma-1}{\gamma}}\right] \tag{49}$$

$$0 = \eta - \lambda_t\left(\frac{W_t}{P_t}\right)(1-h_t)(1-\tau_t) \tag{50}$$

$$0 = a_t b_t^{\frac{1}{\gamma}}\left(\frac{M_t}{P_t}\right)^{-\frac{1}{\gamma}} - \left[\lambda_t-\beta E_t\left(\frac{\lambda_{t+1}P_t}{P_{t+1}}\right)\right]\left[c_t^{\frac{\gamma-1}{\gamma}}+b_t^{\frac{1}{\gamma}}\left(\frac{M_t}{P_t}\right)^{\frac{\gamma-1}{\gamma}}\right] \tag{51}$$

$$0 = \lambda_t - \beta i_t E_t\left(\frac{\lambda_{t+1}P_t}{P_{t+1}}\right) \tag{52}$$

$$\begin{aligned}0 = {} & \lambda_t\left[1+\phi_k\left(\frac{k_{t+1}}{k_t}-1\right)\right]-\beta E_t\left[\lambda_{t+1}\left(\frac{Q_{t+1}}{P_{t+1}}+1-\delta\right)\right]+ \\ & \frac{\beta\phi_k}{2}E_t\left[\lambda_{t+1}\left(\frac{k_{t+2}}{k_{t+1}}-1\right)^2\right]-\beta\phi_k E_t\left[\lambda_{t+1}\left(\frac{k_{t+2}}{k_{t+1}}-1\right)\left(\frac{k_{t+2}}{k_{t+1}}\right)\right]\end{aligned} \tag{53}$$

$$0 = \frac{M_{t-1}+B_{t-1}+(1-\tau_t)W_th_t+Q_tk_t+D_t+T_t}{P_t}- \tag{54}$$

$$\left[c_t + k_{t+1} - (1-\delta)k_t + \frac{\phi_k}{2}\left(\frac{k_{t+1}}{k_t} - 1\right)^2 k_t + \frac{B_t/i_t + M_t}{P_t}\right]$$

Introducing $m_t = M_t/P_t$, $w_t = W_t/P_t$, $q_t = Q_t/P_t$, and $\pi_t = P_t/P_{t-1}$, equations (49)-(53) can be rewritten as

$$\begin{aligned}
0 &= c_t^{\frac{1}{\gamma}} \lambda_t [c_t^{\frac{\gamma-1}{\gamma}} + b_t^{\frac{1}{\gamma}} m_t^{\frac{\gamma-1}{\gamma}}] - a_t && (55)\\
0 &= \lambda_t w_t (1-h_t)(1-\tau_t) - \eta && (56)\\
0 &= m_t \left(1 - \frac{1}{i_t}\right)^{\gamma} - b_t c_t && (57)\\
0 &= \beta i_t E_t(\lambda_{t+1}/\pi_{t+1}) - \lambda_t && (58)\\
0 &= \beta E_t\left[\lambda_{t+1}\left(q_{t+1} + 1 - \delta\right)\right] - \frac{\beta\phi_k}{2} E_t\left[\lambda_{t+1}\left(\frac{k_{t+2}}{k_{t+1}} - 1\right)^2\right]\\
&+ \beta\phi_k E_t\left[\lambda_{t+1}\left(\frac{k_{t+2}}{k_{t+1}} - 1\right)\left(\frac{k_{t+2}}{k_{t+1}}\right)\right] - \phi_k\lambda_t\left(\frac{k_{t+1}}{k_t} - 1\right) - \lambda_t && (59)
\end{aligned}$$

It is easy to see that equations (55)-(59) correspond exactly to (7)-(11). To derive (18), we first introduce investment $x_t = k_{t+1} - (1-\delta)k_t$, and we set B_t equal to zero for all t (without loss of generality due to the Ricardian equivalence in our model). Then we use the (symmetric) intermediate-goods-producing firm's dividend equation,

$$D_t = P_t y_t - W_t h_t - Q_t k_t - \frac{\phi_p}{2}\left[\frac{P_t/P_{t-1}}{\pi} - 1\right]^2 y_t P_t, \qquad (60)$$

as well as the government's budget constraint (6), to rewrite (54) as

$$0 = c_t + x_t + g + \frac{\phi_k}{2}\left(\frac{x_t}{k_t} - \delta\right)^2 k_t + \frac{\phi_p}{2}\left(\frac{\pi_t}{\pi} - 1\right)^2 y_t - y_t. \qquad (61)$$

This expression corresponds to equation (18). Along the way, we have derived equation (19) from the definition of investment.

The optimality conditions associated with the intermediate firms can be used to derive equations (12)-(16). First, we observe that the production technology together with symmetry imply (12). Similarly, the definition of dividend payments together with symmetry imply (13). To derive equations (14)-(16), we make again use of the Lagrangian method. To this end, let the Lagrangian associated with intermediate firm j be given by

$$\mathcal{L}^j_{int} = E_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{P_t}\Big[P_t(j) y_t(j) - W_t h_t(j) - Q_t k_t(j) -$$

$$\frac{\phi_p}{2}\left(\frac{P_t(j)/P_{t-1}(j)}{\pi}-1\right)^2 y_t P_t\Bigg] - \omega_t(j)\left[k_t(j)^\alpha [z_t h_t(j)]^{1-\alpha} - \left(\frac{P_t(j)}{P_t}\right)^{-\theta} y_t\right]$$

The first derivatives with respect to the firm's choice variables $h_t(j)$, $k_t(j)$ and $P_t(j)$, yield a set of equilibrium conditions

$$\begin{aligned}
0 &= \lambda_t \frac{W_t}{P_t} - \omega_t(j)(1-\alpha)k_t(j)^\alpha (z_t)^{1-\alpha} h_t(j)^{-\alpha} && (62)\\
0 &= \lambda_t \frac{Q_t}{P_t} - \omega_t(j)\alpha k_t(j)^{\alpha-1}(z_t h_t(j))^{1-\alpha} && (63)\\
0 &= (1-\theta)\lambda_t\Big[\frac{P_t(j)}{P_t}\Big]^{-\theta}\frac{y_t}{P_t} + \theta\omega_t(j)\Big[\frac{P_t(j)}{P_t}\Big]^{-\theta-1}\frac{y_t}{P_t} && (64)\\
&\quad -\lambda_t\phi_p\Big[\frac{P_t(j)}{\pi P_{t-1}(j)}-1\Big]\frac{y_t}{\pi P_{t-1}(j)} + \beta\phi_p E_t\lambda_{t+1}\Big[\frac{P_{t+1}(j)}{\pi P_t(j)}-1\Big]\Big[\frac{P_{t+1}(j)y_{t+1}}{\pi P_t(j)^2}\Big]
\end{aligned}$$

Multiplying the last equation by P_t/y_t yields

$$\begin{aligned}
0 &= (1-\theta)\lambda_t\Big[\frac{P_t(j)}{P_t}\Big]^{-\theta} + \theta\omega_t(j)\Big[\frac{P_t(j)}{P_t}\Big]^{-\theta-1} && (65)\\
&\quad -\lambda_t\phi_p\Big[\frac{P_t(j)}{\pi P_{t-1}(j)}-1\Big]\frac{P_t}{\pi P_{t-1}(j)} + \beta\phi_p E_t\lambda_{t+1}\Big[\frac{P_{t+1}(j)}{\pi P_t(j)}-1\Big]\Big[\frac{P_{t+1}(j)P_t y_{t+1}}{\pi P_t(j)^2 y_t}\Big]
\end{aligned}$$

Using $m_t = M_t/P_t$, $w_t = W_t/P_t$, $q_t = Q_t/P_t$, and $\pi_t = P_t/P_{t-1}$, and assuming symmetry, we can rewrite the above expressions as

$$\begin{aligned}
0 &= \omega_t(1-\alpha)y_t - \lambda_t w_t h_t && (66)\\
0 &= \omega_t\alpha y_t - \lambda_t q_t k_t && (67)\\
0 &= (1-\theta)\lambda_t + \theta\omega_t - \lambda_t\phi_p\Big(\frac{\pi_t}{\pi}-1\Big)\frac{\pi_t}{\pi}\\
&\qquad +\beta\phi_p E_t\left\{\lambda_{t+1}\Big(\frac{\pi_{t+1}}{\pi}-1\Big)\frac{\pi_{t+1}}{\pi}\frac{y_{t+1}}{y_t}\right\} && (68)
\end{aligned}$$

which correspond to (14)-(16).
We complete the derivation of equilibrium conditions by stating that, in equilibrium, the monetary policy rule (17) must be satisfied, and the exogenous variables must follow their respective laws of motion, given by (20)-(23).

B The non-stochastic steady state

In the absence of disturbances, our model economy converges to a non-stochastic steady state in which all of the variables remain constant over time. Dropping time indices and expectations, and setting all innovations to zero, the steady state of our model is characterized by the system of equations:

$$\begin{aligned}
y &= c + \delta k + g \\
a &= c^{\frac{1}{\gamma}} \lambda [c^{\frac{\gamma-1}{\gamma}} + b^{\frac{1}{\gamma}} m^{\frac{\gamma-1}{\gamma}}] \\
\eta &= \lambda w (1-h)(1-\tau) \\
\psi\lambda &= \beta r (\lambda/\pi) \\
bc &= m \left(1 - \frac{1}{r}\right)^{\gamma} \\
\psi\lambda &= \beta \left[\lambda \left(q + 1 - \delta\right)\right] \\
y &= k^{\alpha} [zh]^{1-\alpha} \\
d &= y - wh - qk \\
\lambda w h &= \omega (1-\alpha) y \\
\lambda q k &= \omega \alpha y \\
0 &= (1-\theta)\lambda + \theta\omega
\end{aligned}$$

By construction of the stochastic processes, the steady state values of a_t, b_t, z_t and τ_t correspond to the parameters a, b, z and τ, respectively. The system of equations above uniquely determines the remaining 12 steady state values y, c, m, h, w, q, k, d, π, i, λ and ω. These are given by:

$$\begin{aligned}
i &= \pi/\beta \\
q &= 1/\beta - 1 - \delta \\
\omega &= [(\theta-1)/\theta]\lambda \\
m &= b \left[\frac{i}{i-1}\right]^{\gamma} c \\
c &= \frac{a}{\lambda} \left[1 + b \left(\frac{i}{i-1}\right)^{\gamma-1}\right]^{-1} \\
k &= \frac{\omega\alpha y}{\lambda q} = \frac{\alpha y}{q} \frac{\theta - 1}{\theta} \\
y &= \left[1 - \delta \frac{\alpha}{q} \frac{\theta - 1}{\theta}\right]^{-1} (c + g)
\end{aligned}$$

$$
\begin{aligned}
h &= \frac{1}{z}\left(\frac{y}{k^{\alpha}}\right)^{\frac{1}{1-\alpha}} \\
w &= \frac{(1-\alpha)y}{h}\frac{\theta-1}{\theta} \\
d &= y - wh - qk
\end{aligned}
$$

and

$$
\lambda = \frac{\eta + (1-\alpha)(1-\tau)\left[\frac{\theta}{\theta-1} - \delta\frac{\alpha}{q}\right]^{-1} a\left[1 + b\left(\frac{r}{r-1}\right)^{\gamma-1}\right]^{-1}}{(1-\alpha)(1-\tau)\left(\frac{\theta-1}{\theta}\right)^{\frac{1}{1-\alpha}} z\left(\frac{\alpha}{q}\right)^{\frac{\alpha}{1-\alpha}} - (1-\alpha)(1-\tau)\left[\frac{\theta}{\theta-1} - \delta\frac{\alpha}{q}\right]^{-1} g}
$$

C Artificial data

This appendix provides details on the construction of our artificial data. We first describe how to obtain data from the nonlinear New Keynesian model, more precisely, from the model solved by a third-order Galerkin projection method. Then we illustrate how we construct data from the VAR(2) model.

The first step in generating artificial data from the nonlinear NK model is to assert values to its structural parameters, Θ, and to solve the model using a nonlinear method. The parameter values we select are described in Table 1. Most of these parameters are taken (or very similar to parameters) from related studies, in particular, Ireland (2004b) and Dib, Gammoudi, and Moran (2006). Whenever parameter values are not available in the related literature, we select values that generate plausible implications for the model's variables. This *casual* approach is unproblematic in our application, we believe, since we use these parameters only to simulate artificial data and not to judge the model against the real world.

Having "calibrated" the model's parameters, we solve the model numerically. Since high-order perturbation methods are difficult to implement, as they require the manipulation of huge matrices and there is hardly any computer code available to build on, we choose to solve the model with a third-order Galerkin projection method. We use a complete basis of Chebyshev polynomials up to an order of three to approximate the policy functions for the future capital stock, $k_{t+1} = \tilde{\Psi}_k(s_t;\kappa)$, current output, $y_t = \tilde{\Phi}_y(s_t;\kappa)$, and current inflation, $\pi_t = \tilde{\Phi}_\pi(s_t;\kappa)$. We use the second-order perturbation solution to compute an initial estimate of κ. Given k_{t+1}, y_t, and π_t we derive the remaining decision variables by solving a linear system of equations. We summarize all policy functions as $\tilde{\Psi}(s_t;\kappa)$ and $\tilde{\Phi}(s_t;\kappa)$. The parameter

vector κ of the policy functions is finally determined such that

$$\int_{\Omega} ER(\tilde{\Phi}(\tilde{\Psi}(s_t, \varepsilon_{t+1}; \kappa); \kappa), \tilde{\Phi}(s_t; \kappa), \tilde{\Psi}(s_t, \varepsilon_{t+1}; \kappa), s_t, \varepsilon_{t+1}) \omega_i(s_t) ds_t = 0, \quad (69)$$

where $\omega_i(s_t)$ denote weighting functions, which, since we apply a Galerkin method, coincide with the regressors used in the approximate policy functions. We evaluate the expectation in (69) by Monte Carlo methods instead of the commonly used quadrature methods, since our model features many exogenous state variables such that quadrature methods would be computationally too demanding. A detailed outline of projection methods is provided, among others, by Judd (1992). Further details on our model's solution can be found in the MATLAB codes which are available upon request.

With the model solution at hand, we simulate time series of the five variables to be included in the data set, $\{y_t, m_t, \pi_t, i_t, \tau_t\}_{t=-9}^{80}$, starting at the steady state. We then drop the first 10 observations, such that our sample is given by $\{y_t, m_t, \pi_t, i_t, \tau_t\}_{t=1}^{80}$. Then we simulate 20 different realizations for the subsequent 16 periods, $\{y_t^j, m_t^j, \pi_t^j, i_t^j, \tau_t^j\}_{t=81}^{96}$, where $j = 1, ..., 20$. We use the first 80 observations together with the last 20×16 observations to construct 20 data sets of length 96. Overall, we conduct this exercise five times, such that we finally have a total of 100 data sets.
The construction of data from the VAR(2) model is very similar. The only difference is that, instead of using the New Keynesian model solution we simulate data from a vector autoregressive model given by

$$\hat{\mathcal{Y}}_t = A_1 \hat{\mathcal{Y}}_{t-1} + A_2 \hat{\mathcal{Y}}_{t-2} + v_t, \quad v_t^2 \sim N(0, \Sigma_2^V) \quad (70)$$

where

$$A_1 = \begin{pmatrix} 0.88 & 0.15 & 0.01 & -0.02 & 0.140 \\ -0.01 & 0.89 & 0 & 0 & -0.040 \\ -1.76 & -4.41 & 0.78 & 0.24 & 0.260 \\ -0.33 & 1.28 & 0.38 & 0.69 & 0.140 \\ 0.07 & 0.08 & -0.01 & 0 & 0.970 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} -0.04 & -0.38 & -0.04 & 0.04 & -0.160 \\ 0.01 & 0.07 & 0 & 0 & 0.040 \\ 0.41 & 0.2 & -0.17 & 0.11 & -0.310 \\ 0.58 & 0.72 & -0.28 & 0.13 & -0.230 \\ -0.09 & -0.17 & 0.01 & 0 & -0.020 \end{pmatrix},$$

and

$$\Sigma_2^V = \begin{pmatrix} 0.0165^2 & 0 & 0 & 0 & 0 \\ 0 & 0.0493^2 & 0 & 0 & 0 \\ 0 & 0 & 0.0030^2 & 0 & 0 \\ 0 & 0 & 0 & 0.0028^2 & 0 \\ 0 & 0 & 0 & 0 & 0.0101^2 \end{pmatrix}.$$

Again, further details are provided in our MATLAB codes.

D Tables and Figures

Table 1: Parameter values used for simulation

Parameter	Value
β	0.99
γ	0.2
ϕ_p	70
π	1.01
a	1
b	0.5
τ	0.3
ρ_a	0.94
ρ_b	0.95
ρ_z	0.92
ρ_τ	0.95
σ_a	0.03
σ_b	0.02
σ_z	0.02
σ_τ	0.01
σ_i	0.005
ρ_r	0.8
ρ_y	0.05
ρ_π	0.8
σ_y^ν	0.7
σ_m^ν	2
σ_π^ν	0.08
σ_i^ν	0.08
σ_τ^ν	0.5
g	1000
z	4000

Table 2: Parameter estimates: linearized model, Kalman filter; DGP=NK

	Min	Max	Mean	True
β	0.986	0.992	0.989	0.990
γ	0.172	0.308	0.238	0.200
ϕ_p	52.747	83.301	70.456	70.000
π	1.008	1.010	1.009	1.010
a	0.673	0.952	0.801	1.000
b	0.417	0.821	0.523	0.500
τ	0.291	0.312	0.302	0.300
ρ_a	0.938	0.981	0.953	0.940
ρ_b	0.605	0.989	0.886	0.950
ρ_z	0.814	0.940	0.880	0.920
ρ_τ	0.845	0.973	0.930	0.950
σ_a	0.007	0.036	0.026	0.030
σ_b	0.010	0.046	0.022	0.020
σ_z	0.022	0.028	0.025	0.020
σ_τ	0.010	0.013	0.011	0.010
σ_i	0.004	0.006	0.005	0.005
ρ_r	0.698	1.000	0.910	0.800
ρ_y	0.000	0.093	0.053	0.050
ρ_π	0.552	1.183	0.916	0.800
σ_y^ν	0.621	0.789	0.687	0.700
σ_m^ν	1.018	2.067	1.660	2.000
σ_π^ν	0.002	0.116	0.077	0.080
σ_i^ν	0.047	0.101	0.078	0.080
σ_τ^ν	0.000	0.581	0.330	0.500
g	1039.130	1698.833	1354.574	1000.000
z	3294.753	5683.011	4365.607	4000.000

Table 3: Parameter estimates: quadratic model, particle filter; DGP=NK

	Min	Max	Mean	True
β	0.986	0.992	0.989	0.990
γ	0.174	0.311	0.241	0.200
ϕ_p	53.397	83.406	70.875	70.000
π	1.008	1.010	1.009	1.010
a	0.672	0.958	0.806	1.000
b	0.425	0.820	0.530	0.500
τ	0.291	0.307	0.301	0.300
ρ_a	0.938	0.981	0.954	0.940
ρ_b	0.606	0.989	0.887	0.950
ρ_z	0.818	0.940	0.882	0.920
ρ_τ	0.859	0.973	0.932	0.950
σ_a	0.007	0.037	0.026	0.030
σ_b	0.010	0.046	0.023	0.020
σ_z	0.022	0.029	0.025	0.020
σ_τ	0.010	0.013	0.012	0.010
σ_i	0.004	0.006	0.005	0.005
ρ_r	0.698	0.994	0.904	0.800
ρ_y	0.000	0.094	0.053	0.050
ρ_π	0.556	1.151	0.911	0.800
σ_y^ν	0.697	0.711	0.702	0.700
σ_m^ν	1.956	2.127	2.030	2.000
σ_π^ν	0.080	0.081	0.080	0.080
σ_i^ν	0.080	0.082	0.081	0.080
σ_τ^ν	0.501	0.528	0.508	0.500
g	1017.437	1696.412	1347.875	1000.000
z	3289.574	5665.490	4385.257	4000.000

Table 4: Log-determinant statistics; DGP=NK

h	Linear	Quadratic	VAR(1)	VAR(2)	RW
1	-7.1028	-7.1677	-6.9330	-7.0535	-6.6249
2	-5.3223	-5.4210	-4.8704	-4.9693	-4.4398
3	-4.7827	-4.8765	-4.3407	-4.4417	-3.7203
4	-4.2073	-4.3121	-3.5416	-3.7734	-3.0608
5	-3.4717	-3.5676	-2.7656	-3.1688	-2.3236
6	-3.0246	-3.1387	-2.5237	-2.8991	-1.5790
7	-2.3946	-2.4983	-1.9100	-2.2515	-0.9690
8	-2.2438	-2.3564	-1.7311	-2.0300	-0.5623
9	-2.0850	-2.2035	-1.6432	-1.8783	-0.2710
10	-2.1521	-2.2937	-1.8407	-2.0596	-0.2208
11	-1.7178	-1.8298	-1.3457	-1.6437	0.1585
12	-1.6663	-1.8023	-1.4624	-1.6335	0.2952
13	-1.6914	-1.8160	-1.4565	-1.6994	0.2497
14	-1.0568	-1.2348	-0.9041	-1.0996	0.7856
15	-1.1374	-1.2671	-1.0369	-1.0806	1.0153
16	-1.2082	-1.3563	-1.0468	-1.1555	1.0576

Table 5: Parameter estimates: linearized model, Kalman filter; DGP=VAR

	Min	Max	Mean
β	0.988	0.993	0.990
γ	0.050	0.189	0.114
ϕ_p	61.086	263.288	131.580
π	1.009	1.012	1.011
a	1.268	9.147	3.146
b	0.356	0.622	0.513
τ	0.298	0.307	0.302
ρ_a	0.830	0.909	0.876
ρ_b	0.810	0.970	0.913
ρ_z	0.658	0.924	0.838
ρ_τ	0.802	0.957	0.906
σ_a	0.031	0.059	0.043
σ_b	0.054	0.064	0.060
σ_z	0.016	0.046	0.026
σ_τ	0.009	0.016	0.012
σ_i	0.002	0.003	0.003
ρ_r	0.679	0.782	0.737
ρ_y	0.011	0.055	0.039
ρ_π	0.290	0.559	0.479
σ_y^ν	0.003	1.029	0.611
σ_m^ν	0.252	2.215	1.457
σ_π^ν	0.002	0.148	0.037
σ_i^ν	0.104	0.182	0.152
σ_τ^ν	0.000	0.657	0.293
g	0.000	3.210	0.734
z	1662.787	4656.575	3458.745

Table 6: Parameter estimates: quadratic model, particle filter; DGP=VAR

	Min	Max	Mean
β	0.988	0.993	0.990
γ	0.050	0.190	0.114
ϕ_p	61.400	265.624	132.347
π	1.009	1.013	1.011
a	1.265	9.251	3.163
b	0.358	0.624	0.516
τ	0.298	0.304	0.301
ρ_a	0.829	0.910	0.876
ρ_b	0.811	0.970	0.914
ρ_z	0.657	0.925	0.839
ρ_τ	0.793	0.958	0.904
σ_a	0.031	0.059	0.044
σ_b	0.054	0.064	0.060
σ_z	0.016	0.046	0.026
σ_τ	0.009	0.017	0.012
σ_i	0.002	0.004	0.003
ρ_r	0.672	0.780	0.736
ρ_y	0.011	0.056	0.039
ρ_π	0.297	0.561	0.483
σ_y^ν	0.696	0.710	0.702
σ_m^ν	1.989	2.024	2.010
σ_π^ν	0.078	0.081	0.080
σ_i^ν	0.079	0.083	0.081
σ_τ^ν	0.486	0.528	0.504
g	0.000	3.189	0.730
z	1667.425	4654.331	3456.876

Table 7: Log-determinant statistics; DGP=VAR

h	Linear NK	Quadratic NK	VAR(1)	VAR(2)	RW
1	-6.6419	-6.7573	-6.4561	-7.1032	-6.2303
2	-4.7736	-4.8183	-4.4462	-4.8278	-3.9804
3	-3.5939	-3.6088	-3.142	-3.0704	-2.6318
4	-2.787	-2.7288	-2.2119	-1.9978	-1.9142
5	-1.8602	-1.8413	-1.2824	-1.0819	-0.9683
6	-1.3541	-1.4106	-0.9033	-0.6155	-0.1875
7	-0.7005	-0.817	-0.3624	-0.1463	-0.5159
8	-0.4572	-0.5657	0.0241	0.3445	0.7066
9	-0.2525	-0.3481	0.211	0.5398	0.9326
10	0.0198	-0.067	0.5615	0.8869	1.3344
11	0.3083	0.1816	0.6734	1.0464	1.6231
12	0.6111	0.4525	0.857	1.3146	1.9338
13	0.9056	0.6975	0.9865	1.4821	2.3876
14	0.9133	0.6641	0.9505	1.4031	2.4581
15	0.9265	0.7097	0.9765	1.4085	2.4630
16	1.0157	0.7597	0.9663	1.5044	2.8083

Table 8: Effective sample size

	Min	Max	Mean
DGP=NK	2441	9850	4842
DGP=VAR	38	2020	880

Figure 1: *Univariate forecast accuracy measures; DGP=NK*

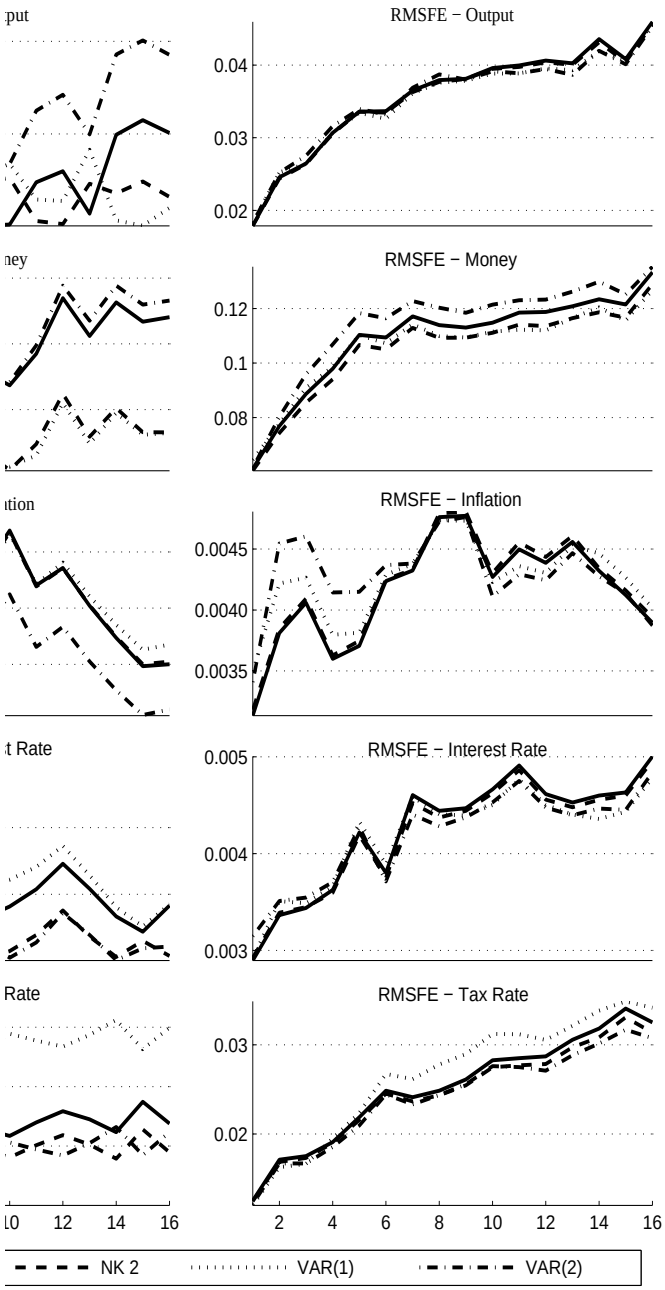

Figure 2: *Univariate forecast accuracy measures, DGP=VAR*