A New Class of Nonlinear Times Series Models for the Evaluation of DSGE Models

S. Borağan Aruoba* — *University of Maryland*
Luigi Bocola — *UPenn*
Frank Schorfheide — *UPenn, CEPR, NBER*

December 9, 2011

Preliminary Version

Abstract

A new class of nonlinear time series models that can be used to evaluate the fit of DSGE models that have been solved with second-order perturbation methods is developed. We use such a nonlinear model to construct a predictive check for a simple New Keynesian DSGE model. We find that the DSGE model is not able to reproduce some of the apparent nonlinearities in U.S. data. (JEL C11, C32, C52, E32)

*Correspondence: B. Aruoba: University of Maryland, Department of Economics, University of Maryland, College Park, MD 20742. Email: aruoba@econ.umd.edu. L. Bocola, F. Schorfheide: University of Pennsylvania, Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104. Email: lbocola@sas.upenn.edu (Bocola), schorf@ssc.upenn.edu (Schorfheide). We thank Frank Diebold, Karel Mertens, as well as seminar participants at the 2011 FRB-NBER DSGE Conference, the University of Pennsylvania, Washington University of St. Louis, University of Rochester for helpful comments. Aruoba and Schorfheide gratefully acknowledge financial support from the National Science Foundation under Grant SES 1061725.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are now widely used for empirical research in macroeconomics as well as for forecasting and quantitative policy analysis in central banks. In these models, decision rules of economic agents are derived from assumptions about agents' preferences and production technologies and some fundamental principles such as intertemporal optimization, rational expectations, and competitive equilibrium. In practice, this means that the functional forms and parameters of equations that describe the behavior of economic agents are tightly restricted by equilibrium conditions. Consequently, a careful evaluation of the DSGE model-implied restrictions is an important aspect of empirical research.

In the past, much of the empirical work was based on linearized DSGE models which, from an econometric perspective, take the form of restricted linear state space models. A natural benchmark for the evaluation of such models is provided by vector autoregressions (VARs) that relax the cross-coefficient restrictions. In fact, there exists an extensive literature that develops and applies methods to evaluate DSGE models based on comparisons with VARs, e.g., Cogley and Nason (1994), Christiano, Eichenbaum, and Evans (2005), Del Negro, Schorfheide, Smets, and Wouters (2007), and Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007).

Starting with the work of Fernández-Villaverde and Rubio-Ramírez (2007), an increasing number of papers that estimate nonlinear, as opposed to linearized DSGE models have been written. Linear approximations tend to be inadequate in settings where risk matters, e.g. asset pricing, welfare comparisons, exposure to potentially large shocks, and in environments with non-convexities, such as occasionally binding constraints. While nonlinear DSGE models can, in principle, be compared to linear vector autoregressions, such comparisons only reveal whether the nonlinear model captures means and autocovariance patterns of observed macroeconomic data, but not whether it captures dynamics in higher-order moments. Thus, the comparison is unable to shed light on the question whether the nonlinearities generated by the structural model coincide with the nonlinearities apparent in the data.

The objective of this paper is to develop a class of time series models that mimic non-linearities of DSGE models and can serve as a benchmark for model evaluations. More specifically, motivated by the popular second-order perturbation approximations of DSGE model dynamics, we consider autoregressive models that involve quadratic terms of lagged endogenous variables as well as interactions between current period innovations and lagged endogenous variables. While our ultimate goal is to develop structural multivariate nonlinear autoregressive models, we focus in this paper on a class of univariate models, that we refer to as QAR(p,q) models, where "Q" stands for quadratic.[1]

After documenting some of the theoretical properties of the QAR models, the first step of the empirical analysis is to fit QAR models to U.S. growth of real Gross Domestic Product (GDP), inflation, and interest rate data. According to our estimates post-1983 GDP growth exhibits nonlinear conditional mean dynamics that imply that large deviations from the mean lower the conditional expectation of output growth. Moreover, whenever output growth is low, i.e. in recessions, its volatility tends to be high. With respect to the Federal Funds rate, the estimated QAR model implies that interest rate volatility is large in periods in with the level of the interest rate is high. Post-1983 inflation did not exhibit significant nonlinearities.

The second step of the empirical analysis consists of the estimation of a prototypical small-scale New Keynesian DSGE model that has been solved by second-order perturbation methods. To do so, we use the same data set as in the estimation of the univariate QAR models. The parameter estimates for the DSGE models are consistent with estimates that have been reported elsewhere in the literature.

The final, and most important step of the analysis is to conduct a posterior predictive check of the DSGE model. To do so, we simulate multiple trajectories from the fitted DSGE model and estimate the QAR model on each of the trajectories. The predictive check amounts to assessing how far the QAR estimates obtained from the actual data lie in the tails of the predictive distribution. The main empirical findings based on post-1983 U.S. data are the following. Our predictive checks indicate that output growth exhibits conditional

[1]The abbreviation QAR has previously been used for Quantile Autoregressions, see Koenker and Xiao (2006)

mean dynamics that our estimated DSGE model is unable to reproduce. Moreover, we find strong evidence for conditional variance dynamics in the interest rate series that the DSGE model is unable to generate. These deficiencies cannot be detected by comparing the nonlinear DSGE model to linear autoregressive models, highlighting the limitations of existing methods of evaluating DSGE models.

Our work is related to several branches of the literature. There exists a large body of work on nonlinear time series models, including regime switching models, e.g. Hamilton (1989) and Sims and Zha (2006), time-varying coefficient models, e.g. Cogley and Sargent (2002) and Primiceri (2005), threshold and smooth-transition autoregressive models, e.g. Tong and Lim (1980) and Teräsvirta (1994), bilinear models, e.g. Granger and Andersen (1978) and Rao (1981), and generalized autoregressive models, e.g. Mittnik (1990). However, none of these model classes seem to be directly useable for our purposes since either the nonlinearities do not match the nonlinearities of DSGE models solved with higher-order perturbation methods or the models have undesirable instability properties.

The proposed QAR family is most closely related to generalized autoregressions (GAR) in the sense that the conditional mean of the dependent variable y_t is a polynomial function of its lags. Our QAR models also involve interactions between lagged dependent variables y_{t-j} and innovations u_t, which is a defining property of bilinear models. In order to ensure that the QAR model is stationary in a region of the parameter space that can be characterized in a straightforward manner, we adapt the idea of pruning, e.g. Kim, Kim, Schaumburg, and Sims (2008), that has been used in the literature on the simulation of DSGE models solved with higher-order perturbation methods. In a nutshell, pruning involves replacing nonlinear functions of y_{t-j} by nonlinear functions of a hidden state variable s_t. This hidden state s_t follows a law of motion that can be viewed as a linear approximation of the law of motion of y_t. As a consequence, the QAR model has a state-space representation in which the innovations of the measurement equation and the state-transition equation are perfectly correlated. To the extent that both conditional mean and variance depend on quadratic functions of the innovations u_t our model is also related to the class of (G)ARCH-M models, e.g. Engle, Lilien, and Robins (1987) and Grier and Perry (1996). Finally, the QAR model can be viewed as a set of tight restrictions on the coefficients of a Volterra (1930) representation of a nonlinear time series.

As mentioned above, there exists an abundant literature that develops methods to evaluate DSGE models based on comparisons with more flexible and densely parameterized time series models. In this paper we use so-called posterior predictive checks to evaluate a prototypical DSGE model. A general discussion of the role of predictive checks in Bayesian analysis can be found in Lancaster (2004) and Geweke (2005). Canova (1994) is the first paper that uses predictive checks to assess implications of a DSGE model. While Canova (1994)'s checks were based on the prior predictive distribution, we use posterior predictive checks in this paper as, for instance, in Chang, Doh, and Schorfheide (2007).

The remainder of the paper is organized as follows. The QAR model is introduced in Section 2. We discuss some of its theoretical properties as well as a Metropolis-within-Gibbs sampler that can be used to implement posterior inference for the parameters of the QAR model. Section 3 reviews the simple New Keynesian model evaluated in this paper. The empirical analysis is presented in Section 4. First we estimate the QAR model on output growth, inflation, and interest rate data for the U.S. and discuss evidence for nonlinearities. Second, we implement the predictive checks and examine the extent to which the nonlinearities generated by the DSGE model mimic the nonlinearities in the U.S. data. Finally, Section 5 concludes. An online Appendix contains detailed derivations of the properties of the QAR model, as well as details of the Markov-Chain-Monte-Carlo (MCMC) methods employed in this paper.

Throughout this paper we use $Y_{t_0:t_1}$ to denote the sequence of observations or random variables $\{y_{t_0}, \dots, y_{t_1}\}$. If no ambiguity arises, we sometimes drop the time subscripts and abbreviate $Y_{1:T}$ by Y. If θ is the parameter vector, then we use $p(\theta)$ to denote the prior density, $p(Y|\theta)$ is the likelihood function, and $p(\theta|Y)$ the posterior density. We use *iid* to abbreviate independently and identically distributed, $N(\mu, \Sigma)$ denotes a multivariate normal distribution with mean μ and covariance matrix Σ, and $IG(s, \nu)$ refers to an Inverse Gamma with location parameter s and degrees of freedom ν. $\mathcal{I}\{a \geq b\}$ denotes the indicator function that is equal to one if $a \geq b$ and equal to zero otherwise.

2 Quadratic Autoregressive Models

While the goal of this research is to develop a nonlinear generalization of the widely used structural VAR models, the analysis in this paper is restricted to the univariate case. Quadratic autoregressive models can be motivated in several different ways. Our motivation is based on techniques that are used to approximate DSGE model dynamics by second-order perturbation methods. Suppose that y_t follows a nonlinear process that can be described by

$$y_t = f(y_{t-1}, \sigma u_t), \quad u_t \overset{iid}{\sim} N(0,1). \tag{1}$$

Moreover, assume that the process has a unique deterministic steady state that solves the equation $y_* = f(y_*, 0)$. Imagine that (1) describes a DSGE-model implied equilibrium law of motion of y_t. A second-order perturbation approximation of the equilibrium dynamics replaces the nonlinear law of of motion (1) by a Taylor series approximation around y_*

$$\begin{aligned} y_t - y_* &= f^{(1,0)}(y_*, 0)(y_{t-1} - y_*) + f^{(0,1)}(y_*, 0)\sigma u_t \\ &+\frac{1}{2} f^{(2,0)}(y_*, 0)(y_{t-1} - y_*)^2 + f^{(1,1)}(y_*, 0)(y_{t-1} - y_*)\sigma u_t \\ &+\frac{1}{2} f^{(0,2)}(y_*, 0)(\sigma u_t)^2 + \text{higher order terms}, \end{aligned} \tag{2}$$

where $f^{(i,j)}(y,u)$ denotes the (i,j)'th derivative of $f(y,u)$. In this approximation the scale of the fluctuations of $y_t - y_*$ is proportional to σ, which means that the higher-order-terms are of order σ^3 and dominated by the first- and second-order terms as $\sigma \longrightarrow 0$.

From a time series analysis perspective, the approximation (2) suggests to generalize the standard AR(1) model by including powers of the lagged dependent variable and interactions between the lagged dependent variable and the innovation:

$$y_t - \phi_0 = \phi_1(y_{t-1} - \phi_0) + \phi_2(y_{t-1} - \phi_0)^2 + (1 + \gamma(y_{t-1} - \phi_0))\sigma u_t, \quad u_t \sim N(0,1). \tag{3}$$

Here we omitted the term $(\sigma u_t)^2$ to preserve the conditional Normal distribution of y_t. For $\gamma = 0$ the model in (3) is a special case of a generalized autoregression (GAR) as discussed in Mittnik (1990). Unfortunately, (3) has two undesirable features. First, the quadratic model has two deterministic steady states:

$$y_* = \phi_0^{(1)} \quad \text{and} \quad y_*^{(2)} = \phi_0 + \frac{1 - \phi_1}{\phi_2}.$$

Second, the model has explosive dynamics even for $|\phi_1| < 1$. Consider the special case of $\phi_0 = \gamma = 0$ and let $\phi_2 > 0$. Suppose that $y_{t-1} = y_*^{(2)} + \tilde{y}$, where $\tilde{y} > 0$. Then

$$\Delta y_t = (\phi_1 - 1)y_{t-1} + \phi_2 y_{t-1}^2 = (1 - \phi_1)\tilde{y} + \phi_2 \tilde{y}^2 > 0 \; \forall \tilde{y} > 0.$$

Depending on the parametrization, even moderate shocks can push y_t above the second steady state and trigger the explosive dynamics. Notice that the multiplicity of steady states and the explosiveness can arise even if the underlying nonlinear model (1) has a unique steady state and exhibits strictly stationary dynamics. In this case, the undesirable properties would simply be an artifact of the Taylor series approximation (2). In the remainder of this section we will modify (3) to eliminate its undesirable properties, which leads us to a new class of time series models that we label QAR models (Section 2.1). We subsequently characterize some of its properties (Section 2.2), describe an MCMC algorithm to implement Bayesian inference (Section 2.3), and discuss generalizations of the basic specification (Section 2.4) and potential alternatives to the approach taken in this paper (Section 2.5).

2.1 Specification of the QAR(1,1) Model

The undesirable features of (3) had been recognized in the literature that discusses the simulation of DSGE models that have been solved with second-order perturbation methods. Suppose one recursively computes y_t based on (3) under the assumption $y_* = 0$ and $\sigma = 1$, starting from $y_0 = y_{-1} = 0$. Then,

$$\begin{aligned}
y_1 &= u_1 \\
y_2 &= \big(u_2 + \phi_1 u_1\big) + \phi_2 u_1^2 \\
y_3 &= \big(u_3 + \phi_1 u_2 + \phi_1^2 u_1\big) + \big(\phi_2 u_2^2 + \phi_1\phi_2(1+\phi_1)u_1^2 + 2\phi_1\phi_2 u_1 u_2\big) \\
& \phi_1\phi_2^2 u_1^3 + \phi_2^2 u_2 u_1^2 + \phi_2^2 u_1^4 \\
&\vdots
\end{aligned}$$

It is easy to see that if $u_1 > 1$ and u_1^{2j} diverges faster than ϕ_2^j converges to zero, the path of y_t is explosive even if the AR(1) coefficient ϕ_1 is less than one in absolute value. Recognizing this problem, Kim, Kim, Schaumburg, and Sims (2008) proposed to eliminate higher-than-second-order terms that arise in the forward simulation of (3). Their so-called pruning amounts to replacing (3) by the following state-space model

$$\begin{aligned} y_t &= \phi_0 + \phi_1(y_{t-1} - \phi_0) + \phi_2 s_{t-1}^2 + (1 + \gamma s_{t-1})\sigma u_t \\ s_t &= \phi_1 s_{t-1} + \sigma u_t. \end{aligned} \tag{4}$$

The latent state s_t is supposed to capture the dynamics associated with the first-order approximation of the nonlinear model (1) and replaces the second-order terms in (3). Simulating (4) forward using the same parameterization and initialization as above yields

$$\begin{aligned} \tilde{y}_1 &= u_1 \\ \tilde{y}_2 &= \big(u_2 + \phi_1 u_1\big) + \phi_2 u_1^2 \\ \tilde{y}_3 &= \big(u_3 + \phi_1 u_2 + \phi_1^2 u_1\big) + \big(\phi_2 u_2^2 + \phi_1 \phi_2 (1 + \phi_1) u_1^2 + 2\phi_1 \phi_2 u_1 u_2\big) \\ &\vdots \end{aligned}$$

Since powers of u_{t-j} of order greater than two are eliminated, the system is guaranteed to be non-explosive if $|\phi_1| < 1$.

While Kim, Kim, Schaumburg, and Sims (2008) used (4) as a device to simulate second-order approximated DSGE models, we treat it as a nonlinear time series model that generalizes the linear dynamics of an AR(1) model. We refer to this model as $QAR(1,1)$, where the first number indicates the number of lags in the conditional mean function and the second number indicates the number of lags that interact with the innovation u_t. Notice that the AR(1) model can be obtained from the QAR(1,1) model by setting $\phi_2 = \gamma = 0$.

2.2 Some Properties of the QAR(1,1) Model

To complete the specification of the QAR(1,1) model in (4) we assume that the distribution of the initial values in period $t = \tau$ have distribution F_τ, that the innovations u_t are normally distributed, and that $|\phi_1|$ is less than one:

$$(y_\tau, s_\tau) \sim F_\tau, \quad u_t \overset{iid}{\sim} N(0,1), \quad |\phi_1| < 1. \tag{5}$$

If the marginal distribution of s_τ is $N\big(0, \sigma^2/(1-\phi_1)^2\big)$, then the process s_τ, $t \geq \tau$, is strictly stationary under the restriction $|\phi_1| < 1$. In turn, the vector process $z_t = [s_{t-1}, s_{t-1}^2, u_t]'$ is strictly stationary and we can rewrite the law of motion of y_t in (4) as

$$y_t = \phi_0 + \phi_1(y_{t-1} - \phi_0) + g(z_t) = \phi_0 + \sum_{j=0}^{\infty} \phi_1^j g(z_{t-j}). \tag{6}$$

This representation highlights that y_t is a stationary process. Since $g(z_t)$ is a nonlinear function of u_t and its history, the process is, however, not linear in u_t anymore. In fact, under the assumption that y_t was initialized in the infinite past ($\tau \longrightarrow -\infty$), we obtain the following representation:

$$y_t = \phi_0 + \sigma \sum_{j=0}^{\infty} \phi_1^j u_{t-j} + \sigma \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \left(\gamma \mathcal{I}\{l > j\} \phi_1^{l-j} + \phi_2 \sum_{k=0}^{\min\{j,l\}} \phi_1^{j+l-k} \right) u_{t-j} u_{t-l}. \tag{7}$$

(7) can be interpreted as a discrete-time Volterra series expansion, in which we impose tight parametric restrictions on the Volterra kernels of order one and two and set the kernels of order larger than two to zero.[2]

By taking expectations of the left-hand-side and right-hand-side of (4) one can deduce that the time-invariant mean of y_t has to satisfy

$$\mathbb{E}[y_t - \phi_0] = \phi_1 \mathbb{E}[y_{t-1} - \phi_0] + \phi_2 \mathbb{E}[s_{t-1}^2] + (1 + \gamma \mathbb{E}[s_{t-1}]) \sigma \mathbb{E}[u_t].$$

Using the formulas for the unconditional mean and variance of s_t and recognizing that u_t and s_{t-1} are independent, we obtain

$$\mathbb{E}[y_t] = \phi_0 + \frac{\phi_2 \sigma^2}{(1-\phi_1)(1-\phi_1^2)}. \tag{8}$$

Thus, the nonlinearity creates a wedge between the mean of the process and its deterministic steady steady state $\lim_{\sigma \longrightarrow 0} \mathbb{E}[y_t] = \phi_0$. If $\phi_2 > 0$ and $0 \leq \phi_1 < 1$ then $\mathbb{E}[y_t] > \phi_0$. Explicit formulas for the variances and covariances of the y_t process are derived in the online Appendix.

To summarize, replacing the generalized autoregression (2) by our so-called QAR(1,1) model defined in (4) has eliminated two undesirable properties of the GAR specification.

[2] The infinite sequences of coefficients on terms $\{u_{t-j}\}_{j \geq 0}$, $\{u_{t-j}u_{t-l}\}_{j\geq 0, l \geq 0}$, $\{u_{t-j}u_{t-l}u_{t-k}\}_{j \geq 0, l \geq 0, k \geq 0}$, etc. are called Volterra kernels.

Our QAR(1,1) model has a unique time-invariant mean and it exhibits stationary dynamics provided the parameters that govern the first-order dynamics satisfy the typical stationarity conditions for linear autoregressive models. Moreover, (7) implies that the QAR(1,1) model can be viewed as a restricted second-order Volterra series approximation of a nonlinear process y_t with a convenient and parsimonious state-space representation given by (4).

2.3 Posterior Inference for the QAR(1,1) Model

We will estimate the QAR(1,1) model using Bayesian methods, which requires us to specify a joint distribution for parameters and observations. The QAR(1,1) model can be viewed as a nonlinear state-space model, where the first equation in (4) is the measurement equation and the second equation is the state-transition equation. However, this state-space model has a special structure that simplifies the analysis: the innovations in the measurement and state-transition equation are perfectly correlated.

Let $\theta = [\phi_0, \phi_1, \phi_2, \gamma, \sigma^2]'$ and recall the notation $Y_{t_1:t_2} = \{y_{t_1}, \dots, y_{t_2}\}$ and $S_{t_1:t_2} = \{s_{t_1}, \dots, s_{t_2}\}$. The joint distribution of observation and state in period t conditional on time $t-1$ information is given by

$$p(y_t, s_t | Y_{0:t-1}, S_{0:t-1}, \theta) = p(y_t | Y_{0:t-1}, S_{0:t-1}, \theta) p(s_t | y_t, Y_{0:t-1}, S_{0:t-1}, \theta). \tag{9}$$

Since we assumed $u_t \overset{iid}{\sim} N(0,1)$, the first density on the right-hand-side is a Normal density, whereas the second density is simply a pointmass at

$$s_t = \phi_1 s_{t-1} + \frac{(y_t - \phi_0) - \phi_1(y_{t-1} - \phi_0) - \phi_2 s_{t-1}^2}{1 + \gamma s_{t-1}}, \tag{10}$$

because conditional on $(y_t, Y_{0:t-1}, S_{0:t-1})$ the latent state s_t is known. By induction, if s_1 is known conditional on (y_1, y_0, s_0), it follows that $S_{1:t}$ is known conditional on $(Y_{0:t}, s_0)$. Thus, the likelihood function can be computed recursively given (y_0, s_0):

$$\begin{aligned} p(Y_{1:T} | y_0, s_0, \theta) &= (2\pi)^{-T/2} |\sigma^2|^{-T/2} \left(\prod_{t=1}^{T} |1 + \gamma s_{t-1}|^{-1} \right) \qquad (11) \\ &\quad \times \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \frac{(y_t - \phi_0 - \phi_1(y_{t-1} - \phi_0) - \phi_2 s_{t-1}^2)^2}{(1 + \gamma s_{t-1})^2} \right\}, \\ s_t &= \phi_1 s_{t-1} + \frac{y_t - \phi_0 - \phi_1(y_{t-1} - \phi_0) - \phi_2 s_{t-1}^2}{1 + \gamma s_{t-1}}, \quad t = 1, \dots, T \end{aligned}$$

In principle, one can integrate out (y_0, s_0) numerically using the model implied conditional distribution $p(s_0|y_0,\theta)$. In our empirical analysis we decided to take the short cut of conditioning on the observation y_0 and setting $s_0 = y_0 - \phi_0$.

While the dimension of θ is small enough such that one could use a single-block random-walk Metropolis (RWM) algorithm to generate draws from the posterior of θ, we decided to consider a Metropolis-within-Gibbs algorithm that groups the parameters into three blocks and is likely to perform better if further lags are added to the model specification. Let $\theta_1 = [\gamma, \sigma]'$, $\theta_2 = [\phi_1, \phi_2]$, and $\theta_3 = \phi_0$. Thus, the first subvector collects the conditional variance parameters, the second subvector the autoregressive parameters, and the last subvector describes the deterministic steady state. Our prior density can be factorized as follows

$$p(\theta) = p(\theta_1)p(\theta_2)p(\theta_3|\theta_1,\theta_2) = p(\gamma)p(\sigma)p(\phi_1,\phi_2)p(\phi_0|\sigma,\phi_1,\phi_2),$$

where

$$\begin{aligned} \gamma &\sim N(\underline{\gamma}, \underline{V}_\gamma), \quad \sigma \sim IG(\underline{s}, \underline{\nu}), \quad \phi_1 \sim \mathcal{I}\{|\phi_1| < 1\} N(\underline{\phi}_1, \underline{V}_{\phi_1}), \\ \phi_2 &\sim N(\underline{\phi}_2, \underline{V}_{\phi_2}), \quad \phi_0|(\sigma,\phi_1,\phi_2) \sim N\big(\underline{\phi}_0(\sigma,\phi_1,\phi_2), \underline{V}_{\phi_0}\big). \end{aligned}$$

The prior distribution for ϕ_1 is truncated to ensure stationarity of the QAR(1,1) model. Further details on the parametrization of the prior will be provided in Section 4.

We now describe the Metropolis-within-Gibbs sampler that is used to generate draws from the posterior distribution of θ, iterating over the conditional distributions of the subvectors θ_1, θ_2, and θ_3.

MCMC Algorithm for QAR(1,1): For $i = 1$ to n, generate draws $\theta_1^{(i)}$, $\theta_2^{(i)}$, and $\theta_3^{(i)}$ as follows.

(i) Draw $\theta_1^{(i)}|(\theta_2^{(i-1)}, \theta_3^{(i-1)})$. We use a random-walk Metropolis step based on a candidate draw ϑ_1 from the proposal distribution $N(\theta_1^{(i-1)}, \kappa_1\Omega_{11})$.

(ii) Draw $\theta_2^{(i)}|(\theta_1^{(i)}, \theta_3^{(i-1)})$. Based on $(\theta_1^{(i)}, \theta_2^{(i-1)}, \theta_3^{(i-1)})$, compute $S_{0,T}$ according to (10). Omitting the i and $i-1$ superscripts, we specify the auxiliary regression

$$\tilde{y}_t = \frac{y_t - \phi_0}{1+\gamma s_{t-1}} = [\phi_1,\ \phi_2]' \begin{bmatrix} \frac{y_{t-1}-\phi_0}{1+\gamma s_{t-1}} \\ \frac{s_{t-1}^2}{1+\gamma s_{t-1}} \end{bmatrix} + \sigma v_t = x_t'\vartheta_2 + \sigma v_t. \tag{12}$$

Using the obvious matrix notation, let $\bar{\vartheta}_2 = (X'X)^{-1}X'\tilde{Y}$ be the OLS estimator of ϑ_2 in (12) and $\bar{V}_{\vartheta_2} = (\sigma^{(i)})^2(X'X)^{-1}$. We use a Metropolis-Hastings step based on a candidate draw ϑ_2 from the proposal distribution $N(\bar{\vartheta}_2, \kappa_2\bar{V}_{\vartheta_2})$.

(iii) Draw $\theta_3^{(i)}|(\theta_1^{(i)}, \theta_2^{(i)})$. Based on $(\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i-1)})$, compute $S_{0,T}$ according to (10). Omitting the i and $i-1$ superscripts, we specify the auxiliary regression

$$\tilde{y}_t = \frac{y_t - \phi_1 y_{t-1} - \phi_2 s_{t-1}^2}{1+\gamma s_{t-1}} = \phi_0 \frac{1-\phi_1}{1+\gamma s_{t-1}} + \sigma v_t = x_t'\vartheta_3 + \sigma v_t. \tag{13}$$

Using the obvious matrix notation, let $\bar{\vartheta}_3 = (X'X)^{-1}X'\tilde{Y}$ be the OLS estimator of ϑ_3 in (13) and $\bar{V}_{\vartheta_2} = (\sigma^{(i)})^2(X'X)^{-1}$. We use a Metropolis-Hastings step based on a candidate draw ϑ_3 from the proposal distribution $N(\bar{\vartheta}_3, \kappa_3\bar{V}_{\vartheta_3})$. $\square$

The constants κ_1, κ_2, and κ_3 scale the variance of the proposal distributions in the three steps of the algorithm and can be used to fine-tune the acceptance rates. The covariance matrix Ω_{11} is obtained as follows. Prior to running the algorithm we use a maximization routine to find the mode $\tilde{\theta}$ of the posterior distribution. We then use numerical second derivatives to obtain the Hessian at the posterior mode. Let Ω be the negative of the inverse Hessian and Ω_{11} the submatrix corresponding to the subvector θ_1 of θ. We choose the initial values $\theta_2^{(0)}$ and $\theta_3^{(0)}$ in the neighborhood of $\tilde{\theta}$.

2.4 Generalizations

The QAR(1,1) model has a straightforward generalization in which we include additional lag terms:

$$\begin{aligned} y_t &= \phi_0 + \sum_{l=1}^{p} \phi_{1.l}(y_{t-l} - \phi_0) + \sum_{l=1}^{p}\sum_{m=l}^{p} \phi_{2.lm} s_{t-l}s_{t-m} \qquad (14) \\ &\quad + \left(1 + \sum_{l=1}^{q} \gamma_l s_{t-l}\right)\sigma u_t \\ s_t &= \sum_{l=1}^{p} \phi_{1.l} s_{t-l} + \sigma u_t. \end{aligned}$$

We refer to (14) as QAR(p,q) model. As in the standard AR(p) model, the stationarity of y_t is governed by the roots of the lag polynomial $1 - \sum_{l=1}^{p} \phi_{1.l} z^l$. The quadratic terms generate an additional $p(p+1)/2$ coefficients in the conditional mean equation for y_t. The MCMC Algorithm described in Section 2.3 can be extended to handle posterior inference for the QAR(p,q) model in a straightforward manner. However, since the number of coefficients grows at rate p^2, a careful choice of prior distributions is required even for moderate values of p to cope with the dimensionality problem. The QAR model can also be extended to the vector case, which is an extension that we are pursuing in ongoing research. The empirical analysis presented in Section 4 is based on the QAR(1,1) specification.

2.5 Alternative Modeling Approaches

In closing we briefly mention two alternatives to the proposed QAR models. Grier and Perry (1996, 2000) have estimated GARCH-M models on macroeconomic time series. GARCH-M models provide a generalization of the ARCH-M models proposed by Engle, Lilien, and Robins (1987) and can be written as

$$\begin{aligned} y_t &= \phi_0 + \phi_1(y_{t-1} - \phi_0) + \phi_2(\sigma_t^2 - \sigma^2) + \sigma_t u_t \\ \sigma_t^2 - \sigma^2 &= \gamma_1(u_{t-1}^2 - \sigma^2) + \gamma_2(\sigma_{t-1}^2 - \sigma^2). \end{aligned}$$

Under suitable parameter restrictions y_t can be expressed as a nonlinear function of u_t and its lags. As in the case of the QAR model, y_t depends on the sequence $\{u_{t-j}^2\}$. In addition, the term $\sigma_t u_t$ introduces interactions between u_t and u_{t-j}^2, $j > 1$. However, coefficients on terms of the form $u_{t-j}u_{t-l}$, $j \neq l$ are restricted to be zero. From our perspective, the biggest drawback of the GARCH-M model is that nonlinear conditional mean dynamics are tied to the volatility dynamics: in the absence of conditional heteroskedasticity the dynamics of y_t are linear. The QAR model is much less restrictive in this regard: y_t can be conditionally homoskedastic ($\gamma = 0$) but at the same time have nonlinear conditional mean dynamics, that is, $\phi_2 \neq 0$.

In recent years, many authors have modeled macroeconomic time series by autoregressive models with time-varying parameters.[3] In the context of an AR(1) model, one would assume that the AR(1) coefficient ϕ_t would follow an exogenous process, e.g. $\phi_t = \phi_{t-1} + \eta_t$, where

[3]Del Negro and Schorfheide (2011) provide a brief survey of this literature.

$\eta_t \stackrel{iid}{\sim} N(0, \omega^2)$. To understand the relationship between time-varying parameter models and our approach suppose the nonlinear dynamics of y_t are described by a separable model of the form

$$y_t = f(y_{t-1}) + \sigma u_t = \phi(y_{t-1})y_{t-1} + \sigma u_t.$$

The function $\phi(y_{t-1})$ could be interpreted as a time-varying parameter of an AR(1) model. In a Bayesian framework, one could use a Gaussian process prior for the function $\phi(y)$ with the property that $\mathbb{E}[\phi(y)] = \bar{\phi}$ and a covariance kernel with the property $cov(\phi(y), \phi(\tilde{y})|\lambda) = \lambda|y - \tilde{y}|$. Here λ is a hyperparameter that controls the *a priori* variation in the function $\phi(y)$. For $\lambda = 0$ the model reduces to a constant coefficient model $\phi(y) = \bar{\phi}$. In this setup the function $\phi(y)$, and thus the time-varying coefficient $\phi(y_{t-1})$, is estimated non-parametrically. Defining $\phi_t = \phi(y_{t-1})$ it is easy to see that the conditional distribution $\phi_t|(\phi_{1:t-1}, Y_{1:t-1})$ is quite different from a time-varying parameter model that assumes that ϕ_t evolves according to a random walk, independent of the past history $Y_{1:t-1}$. While the non-parametric estimation of the function $\phi(y_{t-1})$ is intriguing, the curse of dimensionality makes it impractical for applications in which y_t is a vector process that depends on multiple lags. Thus, we decided to work with a polynomial approximation of the function $f(y_{t-1})$.

3 A Prototypical New Keynesian DSGE Model

The DSGE model we consider is the small-scale New Keynesian model studied in An and Schorfheide (2007). The model economy consists of a final good producing firm, a continuum of intermediate goods producing firms, a representative household, and a monetary as well as a fiscal authority. This model has become a benchmark specification for the analysis of monetary policy and is analyzed in detail, for instance, in Woodford (2003). To keep the model specification simple, we abstract from wage rigidities and capital accumulation.

3.1 The Agents and Their Decision Problems

Final Good Producers. The perfectly competitive, representative, final good producing firm combines a continuum of intermediate goods indexed by $j \in [0,1]$ using the technology

$$\mathcal{Y}_t = \left(\int_0^1 \mathcal{Y}_t(j)^{1-\nu} dj \right)^{\frac{1}{1-\nu}}. \tag{15}$$

The firm takes input prices $P_t(j)$ and output prices P_t as given. Profit maximization implies that the demand for intermediate goods is

$$\mathcal{Y}_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-1/\nu} \mathcal{Y}_t. \tag{16}$$

The relationship between intermediate goods prices and the price of the final good is

$$P_t = \left(\int_0^1 P_t(j)^{\frac{\nu-1}{\nu}} dj \right)^{\frac{\nu}{\nu-1}}. \tag{17}$$

Intermediate Goods Producers. Intermediate good j is produced by a monopolist who has access to the following linear production technology:

$$\mathcal{Y}_t(j) = A_t N_t(j), \tag{18}$$

where A_t is an exogenous productivity process that is common to all firms and $N_t(j)$ is the labor input of firm j. Labor is hired in a perfectly competitive factor market at the real wage W_t. Firms face nominal rigidities in terms of quadratic price adjustment costs

$$AC_t(j) = \frac{\varphi}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - \pi \right)^2 \mathcal{Y}_t(j), \tag{19}$$

where φ governs the price stickiness in the economy and π is the steady state inflation rate associated with the final good. Firm j chooses its labor input $N_t(j)$ and the price $P_t(j)$ to maximize the present value of future profits

$$\mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^s Q_{t+s|t} \left(\frac{P_{t+s}(j)}{P_{t+s}} \mathcal{Y}_{t+s}(j) - W_{t+s} N_{t+s}(j) - AC_{t+s}(j) \right) \right]. \tag{20}$$

Here, $Q_{t+s|t}$ is the time t value of a unit of the consumption good in period $t+s$ to the household, which is treated as exogenous by the firm.

Representative Household. The representative household derives utility from consumption C_t relative to a habit stock and disutility from hours worked H_t. We assume that the habit stock is given by the level of technology A_t. This assumption ensures that the economy evolves along a balanced growth path even if the utility function is additively separable in consumption and hours. The household maximizes

$$\mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^s \left(\frac{(C_{t+s}/A_{t+s})^{1-\tau} - 1}{1-\tau} - \chi_H H_{t+s} \right) \right], \tag{21}$$

where β is the discount factor, $1/\tau$ is the intertemporal elasticity of substitution, and χ_H is a scale factor that determines the steady state level of hours worked. Since we do not use data on hours we set $\chi_H = 1$. The household supplies perfectly elastic labor services to the firms taking the real wage W_t as given. The household has access to a domestic bond market where nominal government bonds B_t are traded that pay (gross) interest R_t. Furthermore, it receives aggregate residual real profits D_t from the firms and has to pay lump-sum taxes T_t. Thus, the household's budget constraint is of the form

$$P_t C_t + B_t + T_t = P_t W_t H_t + R_{t-1} B_{t-1} + P_t D_t + P_t SC_t, \tag{22}$$

where SC_t is the net cash inflow from trading a full set of state-contingent securities. The usual transversality condition on asset accumulation applies, which rules out Ponzi schemes.

Monetary and Fiscal Policy. Monetary policy is described by an interest rate feedback rule of the form

$$R_t = R_t^{*\,1-\rho_R} R_{t-1}^{\rho_R} e^{\epsilon_{R,t}}, \tag{23}$$

where $\epsilon_{R,t}$ is a monetary policy shock and R_t^* is the (nominal) target rate. Our specification of R_t^* implies that the central bank reacts to inflation and deviations of output growth from its equilibrium steady state γ:

$$R_t^* = r\pi^* \left(\frac{\pi_t}{\pi^*} \right)^{\psi_1} \left(\frac{\mathcal{Y}_t}{\gamma \mathcal{Y}_{t-1}} \right)^{\psi_2}. \tag{24}$$

Here r is the steady state real interest rate, π_t is the gross inflation rate defined as $\pi_t = P_t/P_{t-1}$, and π^* is the target inflation rate, which in equilibrium coincides with the steady state inflation rate. The fiscal authority consumes a fraction ζ_t of aggregate output $\mathcal{Y}_t$, where

$\zeta_t \in [0, 1]$ follows an exogenous process. The government levies a lump-sum tax (subsidy) to finance any shortfalls in government revenues (or to rebate any surplus).

Exogenous Processes. The model economy is perturbed by three exogenous processes. Aggregate productivity evolves according to

$$\ln A_t = \ln \gamma + \ln A_{t-1} + \ln z_t, \quad \text{where} \quad \ln z_t = \rho_z \ln z_{t-1} + \epsilon_{z,t}. \tag{25}$$

Thus, on average technology grows at the rate γ and z_t captures exogenous fluctuations of the technology growth rate. Define $g_t = 1/(1 - \zeta_t)$. We assume that

$$\ln g_t = (1 - \rho_g) \ln g + \rho_g \ln g_{t-1} + \epsilon_{g,t}. \tag{26}$$

Finally, the monetary policy shock $\epsilon_{R,t}$ is assumed to be serially uncorrelated. The three innovations are independent of each other at all leads and lags and are normally distributed with means zero and standard deviations σ_z, σ_g, and σ_R, respectively.

3.2 Equilibrium Relationships and Model Solution

We consider the symmetric equilibrium in which all intermediate goods producing firms make identical choices so that the j subscript can be omitted. Since the non-stationary technology process A_t induces a stochastic trend in output and consumption, it is convenient to express the model in terms of detrended variables $c_t = C_t/A_t$ and $y_t = \mathcal{Y}_t/A_t$. The model economy has a unique steady state in terms of the detrended variables that is attained if the innovations $\epsilon_{R,t}$, $\epsilon_{g,t}$, and $\epsilon_{z,t}$ are zero at all times. The steady state inflation π equals the target rate π^* and

$$r = \frac{\gamma}{\beta}, \quad R = r\pi^*, \quad c = (1 - \nu)^{1/\tau}, \quad \text{and} \quad y = g(1 - \nu)^{1/\tau}.$$

Let $\hat{x}_t = \ln(x_t/x)$ denote the percentage deviation of a variable x_t from its steady state x. Then the model can be expressed as

$$\begin{aligned}
1 &= \beta \mathbb{E}_t \left[e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{R}_t - \hat{z}_{t+1} - \hat{\pi}_{t+1}} \right] \qquad (27)\\
\frac{1-\nu}{\nu \varphi \pi^2} \left(e^{\tau \hat{c}_t} - 1 \right) &= \left(e^{\hat{\pi}_t} - 1 \right) \left[\left(1 - \frac{1}{2\nu} \right) e^{\hat{\pi}_t} + \frac{1}{2\nu} \right] \\
&\quad - \beta \, \mathbb{E}_t \left[\left(e^{\hat{\pi}_{t+1}} - 1 \right) e^{-\tau \hat{c}_{t+1} + \tau \hat{c}_t + \hat{y}_{t+1} - \hat{y}_t + \hat{\pi}_{t+1}} \right] \\
e^{\hat{c}_t - \hat{y}_t} &= e^{-\hat{g}_t} - \frac{\varphi \pi^2 g}{2} \left(e^{\hat{\pi}_t} - 1 \right)^2 \\
\hat{R}_t &= \rho_R \hat{R}_{t-1} + (1-\rho_R)\psi_1 \hat{\pi}_t + (1-\rho_R)\psi_2 \left(\Delta \hat{y}_t + \hat{z}_t \right) + \epsilon_{R,t} \\
\hat{g}_t &= \rho_g \hat{g}_{t-1} + \epsilon_{g,t} \\
\hat{z}_t &= \rho_z \hat{z}_{t-1} + \epsilon_{z,t} \\
\Delta \hat{y}_t &= y_t - y_{t-1}.
\end{aligned}$$

The equilibrium conditions can be re-parameterized in terms of $\kappa = \tau(1-\nu)/(\nu\pi^2\varphi)$, which corresponds to the slope of the New Keynesian Phillips curve in the context of a first-order approximation of the equilibrium conditions.

We use a second-order perturbation to solve for the policy functions of the model characterized by the equilibrium conditions (27). The state variables of the model are $s_t = [s_t^{\text{end}}, s_t^{\text{exo}}] = [\hat{y}_{t-1}, \hat{R}_{t-1}, \epsilon_{R,t}, \hat{g}_t, \hat{z}_t]'$, while the control variables are $c_t = [\hat{c}_t, \Delta\hat{y}_t, \hat{\pi}_t]'$.[4] The approximate solution of the model has the form:

$$\begin{aligned}
c_{i,t} &= C_{1i}(\theta) + C_{2ij}(\theta) s_{j,t} + \frac{1}{2} C_{3ijk}(\theta) s_{j,t} s_{k,t} \qquad (28)\\
s_{i,t+1}^{\text{end}} &= S_{1i}^{\text{end}}(\theta) + S_{2ij}^{\text{end}}(\theta) s_{j,t} + \frac{1}{2} S_{3ijk}^{\text{end}}(\theta) s_{j,t} s_{k,t} \\
s_{i,t+1}^{\text{exo}} &= S_{2i}^{\text{exo}}(\theta) s_{i,t}^{\text{exo}} + S_{3i}^{\text{exo}}(\theta) \epsilon_{i,t+1}.
\end{aligned}$$

Here we are using tensor notation which expresses the product of $n \times n$ matrices $A = BC$ as $A_{ij} = B_{ik}C_{kj} = \sum_{k=1}^{n} b_{ik} c_{kj}$, where b_{ik} and c_{kj} denote individual elements of the matrices B and C. The system matrices are intended to be functions of the vector of structural

[4]This is a slight abuse of notation. Note that the s_t here is different from the s_t associated with QAR model in Section 2. Moreover, we are now using c_t to denote the vector of control variables instead of detrended consumption.

parameters θ, while $\epsilon_t = [\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t}]'$ is the vector of structural shocks. The last equation in 28 reflects the restriction that the exogenous disturbances follow independent univariate processes.

Real GDP growth, the Inflation rate and the Interest rate can be expressed as linear functions of state and control variables of this model. For example, real GDP growth can be written in terms of model variables as:

$$\Delta \ln \mathcal{Y}_t = \gamma + \Delta \hat{y}_t + \hat{z}_t.$$

Define $x_t = [c_t', s_t']'$. and let $INFL_t$ and FFR_t denote net inflation and interest rates. The measurement equation for the vector of observations $y_t = [\Delta \ln \mathcal{Y}_t, INFL_t, FFR_t]'$ takes the form:[5]

$$y_{i,t} = A_{1i}(\theta) + A_{2ij}(\theta) x_{j,t} + e_{i,t}, \tag{29}$$

where e_t is a vector of iid Gaussian measurement errors. Thus, (28) and (29) form a nonlinear state-space model.

4 Empirical Analysis

The empirical analysis proceeds in three steps. First, we present estimates of the QAR(1,1) model for U.S. output growth, CPI inflation, and the Federal Funds Rate (Section 4.1). Second, the small-scale New Keynesian DSGE model is estimated based on the same output growth, inflation, and interest rate data, that were used in the first step (Section 4.2). Finally, posterior predictive checks are implemented to assess whether the nonlinearities captured in the second-order-approximated DSGE model are commensurate with the nonlinearities in U.S. data captured by the QAR(1,1) model (Section 4.3). The data are quarterly and collected from a variety of sources. The data for inflation come from the Bureau of Labor Statistics (CPI-U: All Items, seasonally adjusted, 1984=100). The data on the Federal Funds Rate come from the Board of Governors of the Federal Reserve System. We average reported

[5] In slight abuse of notation we now use y_t to denote the vector of observables instead of detrended output in the DSGE model.

monthly figures in order to obtain a quarterly counterpart. Finally, data on real GDP per capita (chained 2005 dollars) come from the Bureau of Economic Analysis. Unless indicated otherwise, the estimation sample ranges from 1984:Q1 to 2010:Q4 and our priors are based on the inspection of a pre-sample from 1974:Q1 to 1983:Q4.

4.1 Estimation of QAR(1,1) Model on U.S. Data

We begin by fitting the QAR(1,1) model to U.S. time series of real GDP growth (GDP), inflation rates (INFL), and the Federal Funds rate (FFR). The prior distribution is summarized in Table 1. We use standard normal distributions for the parameters that govern the nonlinearities, ϕ_2 and γ. The prior distributions for ϕ_1, the first-order autoregressive coefficient, are centered at the pre-sample first-order autocorrelations of the three time series. The location parameter $\underline{s}$ for the inverse Gamma distribution of σ corresponds to the residual standard deviation associated with the pre-sample estimation of an AR(1) model. Finally, the prior mean of ϕ_0 is specified such that the implied $\mathbb{E}[y_t]$ of the QAR(1,1) model, see (8), corresponds to the pre-sample mean of the respective time series.

Insert Table 1 Here

The MCMC algorithm described in Section 2.3 is used to implement the posterior inference. Table 2 summarizes the posterior distributions for the QAR(1,1) parameters. Recall that ϕ_1 captures the first-order autocorrelation of the observed time series. The posterior mean estimates are 0.34 for GDP Growth, 0.82 for Inflation, and 0.98 for the Federal Funds Rate. Most important for our analysis are the parameters ϕ_2 and γ, which cover nonlinearities in the conditional mean as well as conditional heteroskedasticity. Only GDP growth seems to exhibit nonlinear conditional mean dynamics with a 90% posterior credible set for ϕ_2 that ranges from -0.39 to -0.12. A strong conditional heteroskedasticity is captured by the estimate of γ for the Federal Funds rate. The posterior credible set ranges from 0.10 to 0.17. According to this estimate interest rate volatility is larger in episodes of high interest rates. GDP growth also exhibits conditional heteroskedasticity. The posterior mean of γ for output growth is -0.11, meaning that output growth volatility increases in recessions, when the hidden state s_t of the QAR(1,1) model is less than zero and the growth is below its long-run average level.

Insert Table 2 Here

We are reporting log marginal data densities for QAR(1,1) and AR(1) models of GDP growth, inflation, and interest rates. The priors for the AR(1) are identical to those used for the QAR(1,1) model, with the exception that ϕ_2 and γ are set to zero. Under equal prior probabilities, the difference in log marginal data density between two models has the interpretation of log posterior odds. In line with the posterior estimates discussed above, we find that the QAR(1,1) specification is preferred to the linear AR(1) model for GDP growth and the Federal Funds rate, whereas the AR(1) model slightly dominates the QAR(1,1) model for the inflation series. To shed more light on this finding we provide scatter plots of the data as well as density plots for the ergodic distributions associated with the estimated AR(1) and QAR(1,1) models in Figures 1 to 3.[6]

Insert Figures 1, 2, and 3 Here

The top panels of Figure 1 depicts scatter plots of lagged versus current GDP growth together with a linear and a quadratic sample regression function. For the post 1984 sample (top right panel) the quadratic regression function essentially picks up the low GDP growth observations of the most recent recession. Compared to the GDP growth rates observed during the post-Great-Moderation period from 1984 to 2006, these observations are fairly extreme and contribute to the significance of the quadratic term in the regression function. If the sample is extended to the pre-Great-Moderation period from 1966 to 1984 the nonlinearity in the regression function is dampened.

The bottom panels of Figure 1 show non-parametric kernel density plots computed from the observed U.S. data as well as data simulated from the estimated QAR(1,1) and AR(1) models. For each of the two time series models we simulate a very long trajectory of observations conditional on the posterior mean parameter estimates to approximate the ergodic

[6] The graphical depiction is based on figures prepared by Karel Mertens for a discussion of this paper.

distribution. For the long sample the AR(1) and QAR(1,1) densities look essentially identical. For the short sample the QAR(1,1) model generates a slightly leptokurtic and skewed distribution that more closely resembles the density of the actual U.S. growth rates.

Results for inflation are plotted in Figure 2. As discussed above, based on the short sample the estimates of ϕ_2 and γ for inflation are essentially zero. Correspondingly, the linear and quadratic regression lines in the top-right panel and the ergodic distributions of the estimated AR(1) and QAR(1,1) models in the bottom-right panel are non-distinguishable. Interestingly, if the QAR(1,1) fitted to the long inflation sample from 1966 to 2010, then estimate of the quadratic term in the conditional mean equation is non-zero. In turn the ergodic distribution associated with the estimated model is skewed, with a long right tail that captures the high inflation rates in the late 1970s. The QAR(1,1) is a lot more successful in mimicking the shape of the density estimated from U.S. inflation data. The most striking feature of the nominal interest rate data in Figure 3 is the skewness generated by the zero lower bound. The QAR(1,1) model captures this skewness through a positive coefficient γ, which implies that the interest rate volatility is low whenever the level of the interest rate is small.

Insert Figure 4 Here

To document the role of the nonlinear terms for the dynamics of GDP growth and interest rates, we construct posterior statistics for two functionals of the QAR(1,1) parameters. The top panel of Figure 4 depicts the time path of the conditional heteroskedasticity of GDP growth and the Federal Funds rate, given by $(1 + \gamma s_{t-1})\sigma$. While the variation in the volatility of the GDP innovations is modest, it is in general larger during recessions than during expansions. The top panel of Figure 4 displays the elasticity

$$\mathbb{E}_t\left[\frac{\partial y_{t+1}}{\partial u_t}\right] = \phi_1(1 + \gamma s_{t-1})\sigma + 2\phi_2 s_t \sigma.$$

This elasticity measures the impact of a one-standard deviation shock u_t on y_t in the following quarter. For a simple AR(1) this elasticity is constant. In our QAR(1,1) the elasticity is time-varying for two reasons: the conditional heteroskedasticity generated by $\gamma \neq 0$ (shown in the top panel) as well as the nonlinear conditional mean dynamics due to $\phi_2 \neq 0$. The elasticity peaks in the early periods of recessions. Recessions are periods in which GDP growth is below average and the hidden state s_t is potentially large in absolute value. The negative coefficient ϕ_2 on s_{t-1}^2 implies that output growth is lower than predicted by the linear dynamics of an AR(1) model. Moreover, the negative value of γ combined with a negative s_{t-1} amplifies the impact of the shock. Thus, the parameter estimates can generate a deep and fast recession.

Qualitatively, our results for GDP growth are in line with findings by Brunner (1997), who estimated three nonlinear models for real Gross National Product (GNP): a self-exciting threshold autoregressive (SETAR) model, a state-dependent model of conditional variance, and a semi-nonparametric model with time-varying volatility. Based on a sample from 1947 to 1990 the author obtained strong evidence of countercyclical volatility, that is recessions are periods of high volatility. Moreover, Brunner (1997) detects nonlinear conditional mean dynamics: according to the impulse responses of the SETAR model, the effects of a negative shock accumulate faster than those of a positive shock.

4.2 DSGE Model Estimation on U.S. Data

The second step in the empirical analysis consists of estimating the DSGE model based on the same data that was used to estimate the QAR(1,1) data. The prior distribution for the DSGE model parameters is summarized in Table 4. The parameters ν and $1/g$ are fixed at 0.10 and 0.85, respectively. Using a first-order approximation of the DSGE model, neither parameter is identifiable from the estimation sample. It turned out that even under a second-order approximation our sample remains uninformative. We use pre-sample evidence to quantify *a priori* beliefs about the average growth rate of the economy as well as average inflation and real interest rates. Our prior for κ encompasses values that imply an essentially flat as well as a fairly steep New Keynesian Phillips curve. One notable feature of the prior distribution is the fairly small prior standard deviation for ψ, the inflation coefficient in the interest rate equation. It has been documented in the literature that this coefficient is not very well identified. A tight prior tends to stabilize the posterior simulator.

Insert Table 4 Here

As discussed in Section 3, a second-order approximation is used to solve the DSGE model, which leads to the nonlinear state-space representation given by (28) and (29). We use the particle filter developed in Fernández-Villaverde and Rubio-Ramírez (2007) to evaluate the likelihood function of the DSGE model. To facilitate the likelihood evaluation with the particle filter the measurement equation contains mean-zero *iid* Gaussian measurement errors. The measurement error variances are set equal to 10% of the sample variances of GDP growth, inflation, and interest rates. Posterior inference is implemented with a single-block Random-Walk Metropolis algorithm, described in detail in An and Schorfheide (2007).

Insert Table 5 Here

Posterior summary statistics for the DSGE model parameters are reported in 5. In the second and third column we report posterior means and credible intervals based on the estimation of the nonlinear version of the DSGE model, whereas the last column contains posterior means based on a linearized version of the DSGE model. By and large the parameter estimates are similar to estimates that have been reported elsewhere in the literature. The estimated slope of the Phillips curve is fairly small, implying that monetary policy shocks tend to have a large effect on output. The estimate of τ implies that households have approximately logarithmic preferences. The estimated serial correlation of technology shocks is small, $\hat{\rho}_z = 0.26$, which is consistent with the small autocorrelation of output growth. The government spending shock is more persistent, but its autocorrelation is not as close to unity as in many estimated DSGE models. Due to the tight prior distribution, the posterior mean of the inflation coefficient in the monetary policy rule essentially equals its prior mean of 1.5. According to our estimates, the central bank reacts quite forcefully to output growth movements, $\hat{\psi}_2 = 1.51$, and engages in moderate interest rate smoothing, $\hat{\rho}_r = 0.54$.

4.3 Posterior Predictive Checks

The key part of the empirical analysis is a set of posterior predictive checks. The posterior predictive checks is implemented with the following algorithm.

Posterior Predictive Checks. Let $\theta^{(i)}$ denote the i'th draw from the posterior distribution of the DSGE model parameter θ. For $i = 1$ to n

(i) Conditional on $\theta^{(i)}$ simulate a pre-sample of length T_0 and an estimation sample of size T from the DSGE model. The second-order approximated is simulated using the pruning algorithm described in Kim, Kim, Schaumburg, and Sims (2008). A Gaussian *iid* measurement error is added to the simulated data. The measurement error variance is identical to the one imposed during the estimation of the DSGE model. Denote the simulated data by $Y^{(i)}_{-T_0+1:T}$.

(ii) Based on the simulated trajectory $Y^{(i)}_{-T_0+1:T}$ compute sample statistics $\mathcal{S}\big(Y^{(i)}_{-T_0+1:T}\big)$. These statistics include sample mean, sample variance, sample autocorrelation, as well as posterior means for the coefficients of the QAR(1,1) model. The latter are obtained by eliciting a prior for the QAR(1,1) model based on the presample $Y^{(i)}_{-T_0+1:0}$ and computing the posterior based on $Y^{(i)}_{1:T}$.

Use a nonparametric Kernel density estimator to approximate the predictive density $p(\mathcal{S}(\cdot)|Y_{1:T})$ and examine how far the actual value $\mathcal{S}(Y_{1:T})$, computed from U.S. data, lies in the tail of its predictive distribution. □

Insert Figure 5 Here

The results from the predictive checks based on the three sample moments are summarized in Figure 5. The solid blue lines indicate the predictive densities and the dashed red vertical lines signify the sample moments computed from the actual data. The estimated nonlinear DSGE model is able to reproduce the observed sample means, standard deviations, and autocorrelations of GDP growth and inflation. The sample moments computed from the actual data lie in the center of the respective posterior predictive distributions. With respect to the interest rate series the fit is not quite as good: the actual sample standard deviation and autocorrelation of the Federal Funds rate lie in the far right tail of their predictive distributions.

Insert Figures 6, 7, and 8 Here

We now turn our attention to the predictive checks based on the posterior means of the QAR(1,1) coefficients, which are presented in Figures 6 to 8. In particular, we will focus on the parameters ϕ_2 and γ, which generate the nonlinearities in the QAR(1,1) model. Consider the results for the GDP growth series, depicted in Figure 6. For U.S. data the posterior mean of ϕ_2 and γ are, respectively, -0.25 and -0.11. While the estimated New Keynesian model assigns significant probability to values of $\hat{\gamma}$ between 0 and -0.2, the value of $\hat{\phi}_2$ obtained from actual data lies far in the left tail of the predictive distribution generated by the estimated DSGE model. Thus, there seem to be conditional mean dynamics present in the data that the DSGE model is unable to capture.

The opposite result is obtained for the Federal Funds rate in Figure 8. The ϕ_2 estimates obtained from both actual as well as simulated data are essentially zero and the DSGE model passes the $\hat{\phi}_2$-based predictive check. With respect to γ, however, model-simulated and actual data look quite different. There is strong evidence in the actual data for a positive $\hat{\gamma}$, but none in the simulated data. More precisely, $\hat{\gamma}$ based on U.S. data lies far in the right tail of its predictive distribution. The DSGE model is unable to generate the conditional variance dynamics that appear to be present in the data. Finally, for inflation the estimates of the QAR(1,1) based on actual data seem to line up fairly well with the estimates based on the model generated data.

Our results highlight the limitations of *linear* time series model as tools to assess the fit of DSGE models solved with higher-order perturbation methods. Indeed, we have seen that the estimated model solved with second-order perturbation generates trajectories for the variables of interests that respect the first and second moments of the data. But we have also seen that the same trajectories do not always generate empirically plausible values for the QAR(1,1) model's parameters. Indeed, the simple small-scale New Keynesian model considered in this paper does not have any channel to generate secular and cyclical movements in the conditional volatility of the Federal Funds and GDP growth series, and the GDP growth series, nor can it generate asymmetric movements across the business cycle in GDP growth. We were able to detect these problems after inserting quadratic terms into a standard AR(1) model, and we believe that this result offers a strong rationale for our exercise.

5 Conclusion

Building on the ideas of generalized autoregressive models, bilinear models, and second-order Volterra expansions this paper proposes a new class of nonlinear time series models that can be used to assess nonlinear DSGE models. So far, we have explored a fairly restricted specification from this class, the QAR(1,1) model. In future research we are planning to consider more complex specifications that can replace multivariate vector autoregressive models. Using the QAR(1,1) model to construct predictive checks, we found that GDP growth and the Federal Funds rate exhibit nonlinear dynamics that cannot be reproduced with a small-scale nonlinear New Keynesian DSGE model. To some extent this result is not surprising as it is well known in the literature that the dynamics of the canonical New Keynesian model are essentially linear, provided the shocks are not too large, and that one would need to introduce stochastic volatility processes for the structural shocks to generate conditional heteroskedasticity. However, we think that the tools developed in this paper will be very useful to evaluate other nonlinear DSGE models.

References

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26(2-4), 113–172.

Brunner, A. D. (1997): "On the Dynamic Properties of Asymmetric Models of Real GNP," *Review of Economics and Statistics*, 79(2), 321–326.

Canova, F. (1994): "Statistical Inference in Calibrated Models," *Journal of Applied Econometrics*, 9, S123–144.

Chang, Y., T. Doh, and F. Schorfheide (2007): "Non-stationary Hours in a DSGE Model," *Journal of Money, Credit, and Banking*, 39(6), 1357–1373.

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.

Cogley, T., and J. M. Nason (1994): "Testing the Implications of Long-Run Neutrality for Monetary Business Cycle Models," *Journal of Applied Econometrics*, 9, S37–70.

Cogley, T., and T. J. Sargent (2002): "Evolving Post-World War II U.S. Inflation Dynamics," in *NBER Macroeconomics Annual 2001*, ed. by B. S. Bernanke, and K. Rogoff, vol. 16, pp. 331–88. MIT Press, Cambridge.

Del Negro, M., and F. Schorfheide (2011): "Bayesian Macroeconometrics," in *The Oxford Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. van Dijk, pp. 293–389. Oxford University Press.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit of New Keynesian Models," *Journal of Business and Economic Statistics*, 25(2), 123–162.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987): "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55(2), 391–407.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74(4), 1059–1087.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97(3), 1021–1026.

Geweke, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Hoboken, New Jersey.

Granger, C. W., and A. P. Andersen (1978): *Introduction to Bilinear Time Series Models*. Vandenhoeck & Ruprecht, Göttingen.

Grier, K. B., and M. J. Perry (1996): "Inflation, Inflation Uncertainty, and Relative Price Dispersion: Evidence from Bivariate GARCH-M Models," *Journal of Monetary Economics*, 38, 391–405.

——— (2000): "The Effects of Real and Nominal Uncertainty on Inflation and Output Growth: Some GARCH-M Evidence," *Journal of Applied Econometrics*, 15, 45–58.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econemetrica*, 57(2), 357–384.

Kim, J., S. Kim, E. Schaumburg, and C. A. Sims (2008): "Calculating and Using Second-Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models," *Journal of Economic Dynamics and Control*, 32, 3397–3414.

Koenker, R., and Z. Xiao (2006): "Quantile Autoregression," *Journal of the American Statistical Association*, 101(475), 980–990.

Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*. Blackwell Publishing.

Mittnik, S. (1990): "Modeling Nonlinear Processes with Generalized Autoregressions," *Applied Mathematical Letters*, 3(4), 71–74.

Primiceri, G. E. (2005): "Time Varying VARs and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

Rao, S. T. (1981): "On the Theory of Bilinear Time Series Models," *Journal of the Royal Statistical Society Series B*, 43(2), 244–255.

Sims, C. A., and T. Zha (2006): "Were There Regime Switches in U.S. Monetary Policy?," *American Economic Review*, 96(1), 54–81.

Teräsvirta, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89(425), 208–218.

Tong, H., and K. Lim (1980): "Threshold Autoregression, Limit Cycles and Cyclical Data," *Journal of the Royal Statistical Society Series B*, 42(3), 245–292.

Volterra, V. (1930): *Theory of Functionals and of Integral and Integro-Differential Equations*. Dover Publications.

Woodford, M. (2003): *Interest and Prices*. Princeton University Press.

Table 1: Prior Distribution for QAR(1,1) Model

	GDP Growth	Inflation	Federal Funds Rate
ϕ_1	$N^{\dagger}(0.25, 0.5)$	$N^{\dagger}(0.96, 0.5)$	$N^{\dagger}(0.93, 0.5)$
ϕ_2	$N(0, 0.1)$	$N(0, 0.1)$	$N(0, 0.1)$
γ	$N(0, 0.1)$	$N(0, 0.1)$	$N(0, 0.1)$
σ	$IG(1.03, 4)$	$IG(0.80, 4)$	$IG(1.22, 4)$
ϕ_0	$N\big(0.57, 2\big)$	$N\big(5.14, 2\big)$	$N\big(6.69, 2\big)$

Notes: The location parameters for the priors of ϕ_0, ϕ_1, and σ correspond to pre-sample mean, autocorrelation, and AR(1) innovation standard deviation. The pre-sample ranges from 1974:Q1 to 1983:Q4. For the estimation of the QAR(1,1) model based on simulated data we parameterize the prior based on simulated pre-samples of 40 observations. (†) The prior for ϕ_1 is truncated to ensure stationarity. The IG distribution is parameterized such that $p_{IG}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$.

Table 2: Posterior Estimates for QAR(1,1) Model

Data	ϕ_0	ϕ_1	ϕ_2	γ	σ
GDP Growth	0.48	0.34	-0.25	-0.11	0.55
	[0.35,0.61]	[0.20,0.51]	[-0.39,-0.12]	[-0.23,-0.01]	[0.49,0.62]
Inflation	2.80	0.82	0.00	-0.01	0.84
	[1.28,4.05]	[0.67,0.99]	[-0.06,0.03]	[-0.06,0.09]	[0.67,0.98]
Federal Funds Rate	6.17	0.98	-.003	0.12	0.78
	[4.46,7.33]	[0.97,0.99]	[-.006,-.001]	[0.10,0.17]	[0.64,0.91]

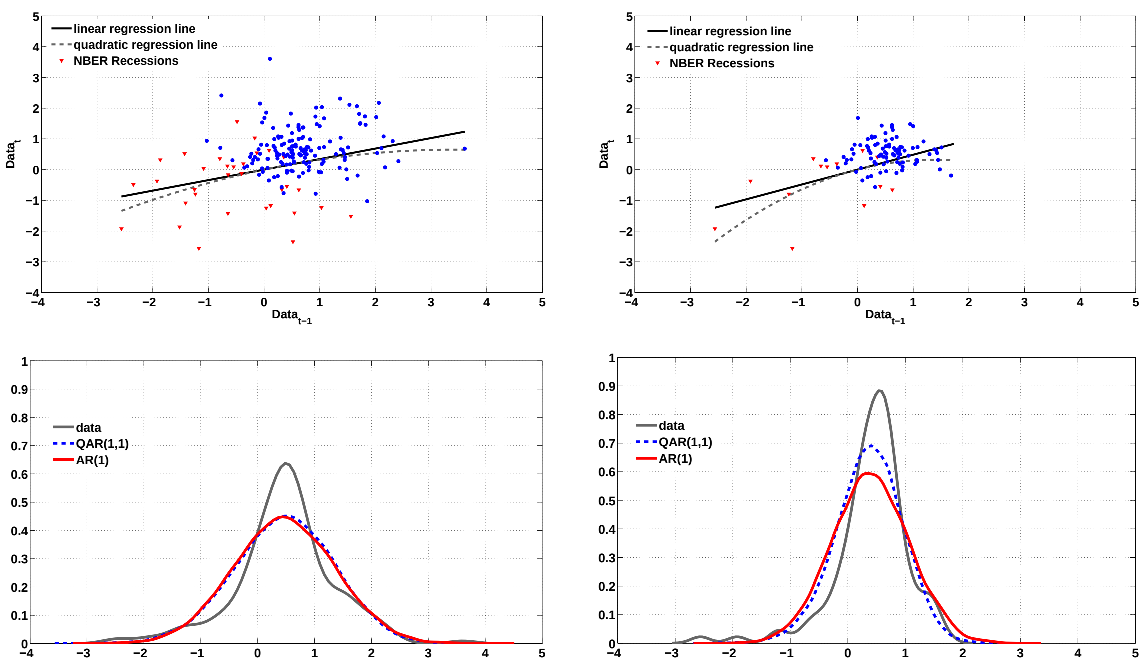
Notes: Estimation sample is 1984:Q1 to 2010:Q4. We report posterior means as well as 90% equal-tail-probability credible sets.

Table 3: Log Marginal Data Density: QAR(1,1) and AR(1)

	GDP	Inflation	FFR
QAR(1,1)	-91.63	-141.66	-82.78
AR(1)	-98.15	-140.99	-96.70

Figure 1: A Look at the Data: GDP Growth

1966:Q1 - 2010:Q4 | 1984:Q1 - 2010:Q4

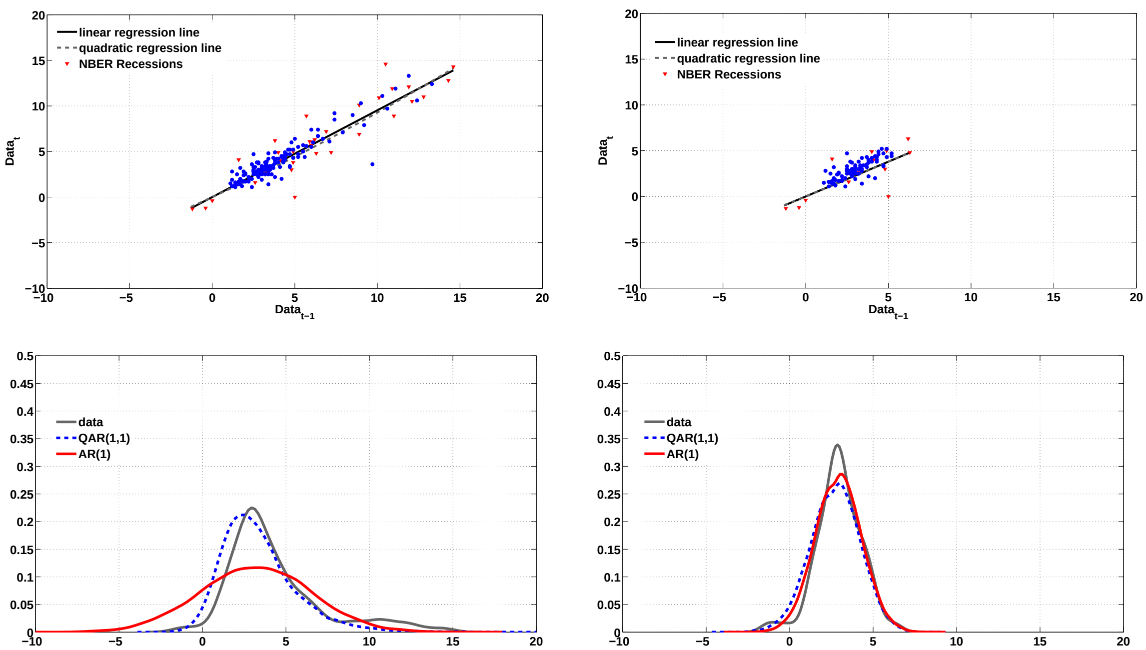


$\hat{\phi}_2 = -0.04$, $\hat{\gamma} = -0.06$ | $\hat{\phi}_2 = -0.25$, $\hat{\gamma} = -0.11$

Notes: The top panels depict scatter plots of time $t-1$ versus time t observations. If $t-1$ was designated as recession period by the NBER the point in the scatter plot appears in red. We overlay linear and quadratic sample regression lines. The bottom panels depict kernel density plots for (i) the observed data; (ii) the ergodic distribution associated with the posterior mean estimates of the QAR(1,1) model; (iii) the ergodic distribution associated with the posterior mean estimates of the AR(1,1) model.

Figure 2: A Look at the Data: Inflation

1966:Q1 - 2010:Q4 1984:Q1 - 2010:Q4

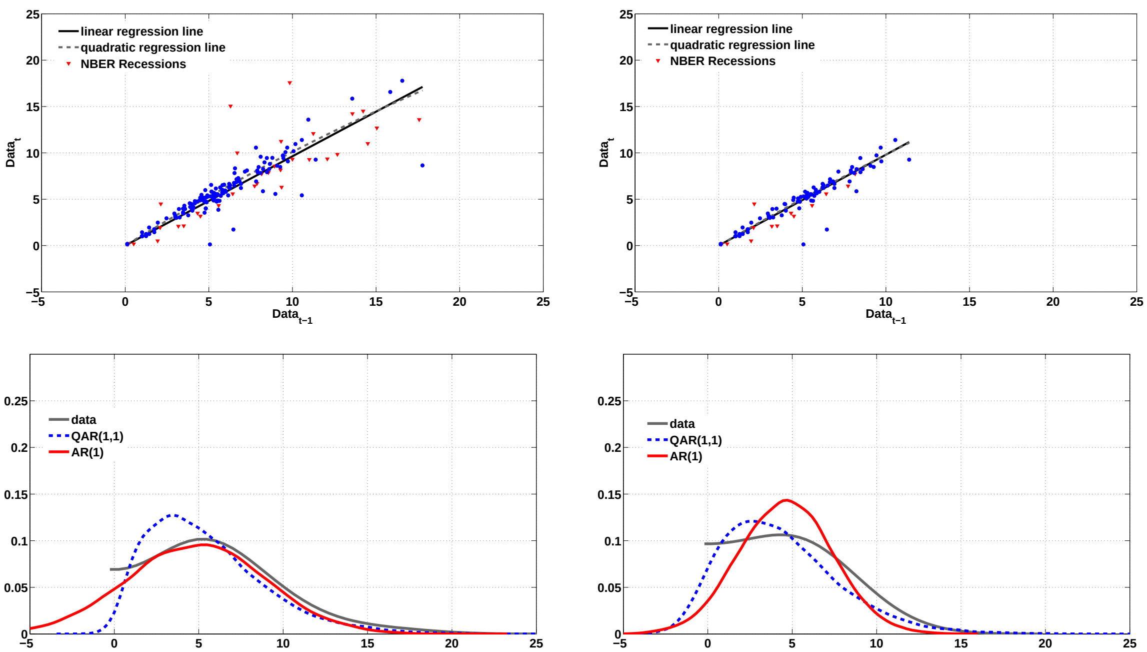


$\hat{\phi}_2 = 0.04,\ \hat{\gamma} = 0.01$ $\hat{\phi}_2 = 0.00,\ \hat{\gamma} = 0.01$

Notes: See Figure 1

Figure 3: A Look at the Data: Federal Funds Rate

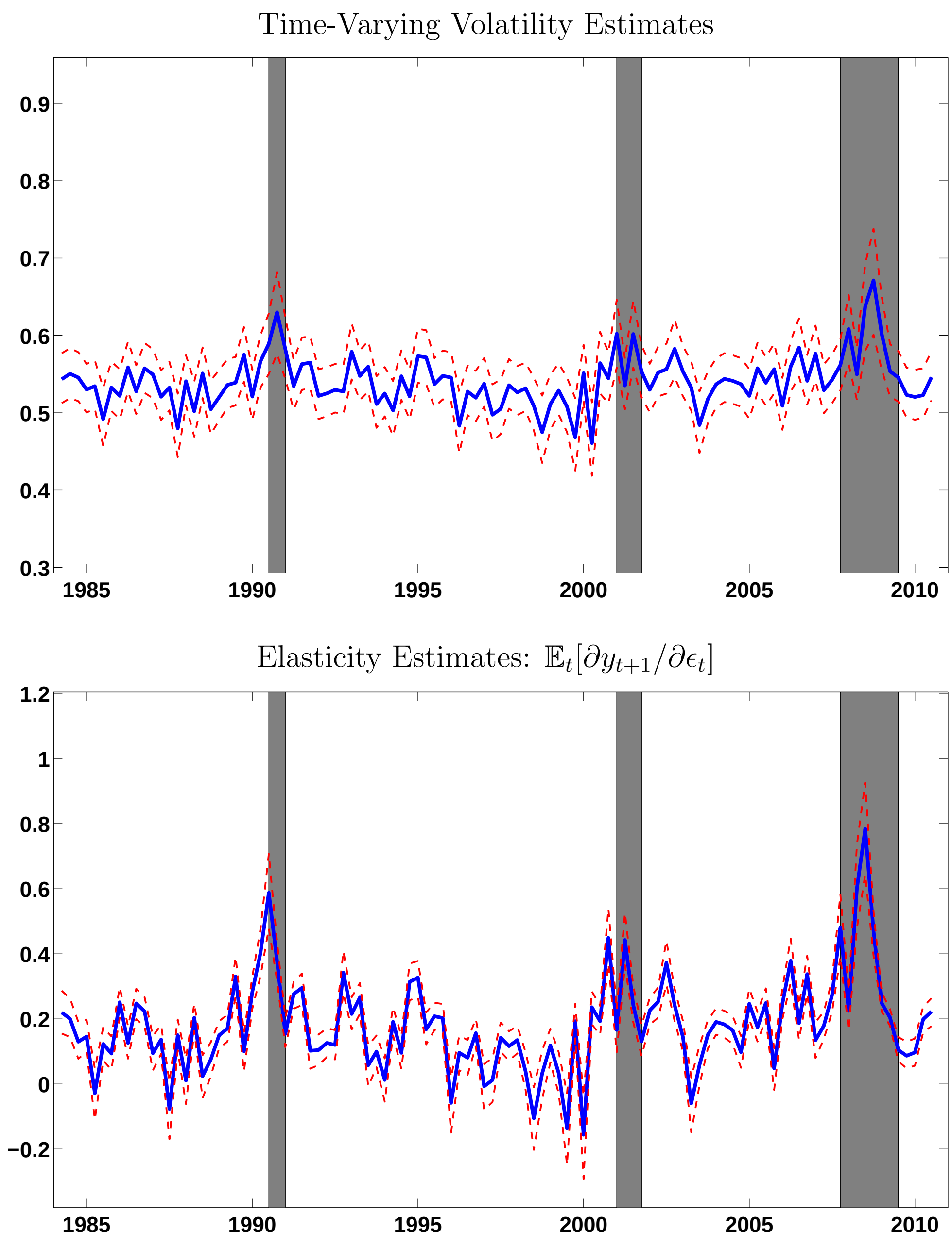
1966:Q1 - 2010:Q4 | 1984:Q1 - 2010:Q4



$\hat{\phi}_2 = 0.00$, $\hat{\gamma} = 0.15$ | $\hat{\phi}_2 = 0.00$, $\hat{\gamma} = 0.12$

Notes: See Figure 1

Figure 4: Nonlinearities of GDP Growth



Notes: Top panel: posterior median and 60% credible intervals (equal tail probability) for $(1+\gamma s_{t-1})\sigma$. Bottom panel: posterior median and 60% credible intervals (equal tail probability) for $\mathbb{E}_t[\partial y_{t+1}/\partial u_t] = \phi_1(1+\gamma s_{t-1})\sigma + 2\phi_2 s_t \sigma$. Estimation sample is 1984:Q1 to 2010:Q4. Shaded areas indicate NBER recessions.

Table 4: Prior for Structural Parameters of DSGE Model

Parameter	Distribution	Para (1)	Para (2)
τ	Gamma	2.00	0.50
κ	Gamma	0.30	0.20
ψ_1	Gamma	1.50	0.05
ψ_2	Gamma	0.50	0.25
ρ_r	Beta	0.50	0.20
ρ_g	Beta	0.80	0.10
ρ_z	Beta	0.20	0.15
r^A	Gamma	0.80	0.50
π^A	Gamma	4.00	2.00
γ^Q	Normal	0.40	0.20
$100\sigma_r$	InvGamma	0.30	4.00
$100\sigma_g$	InvGamma	0.40	4.00
$100\sigma_z$	InvGamma	0.40	4.00
ν	fixed	0.10	
$1/g$	fixed	0.85	

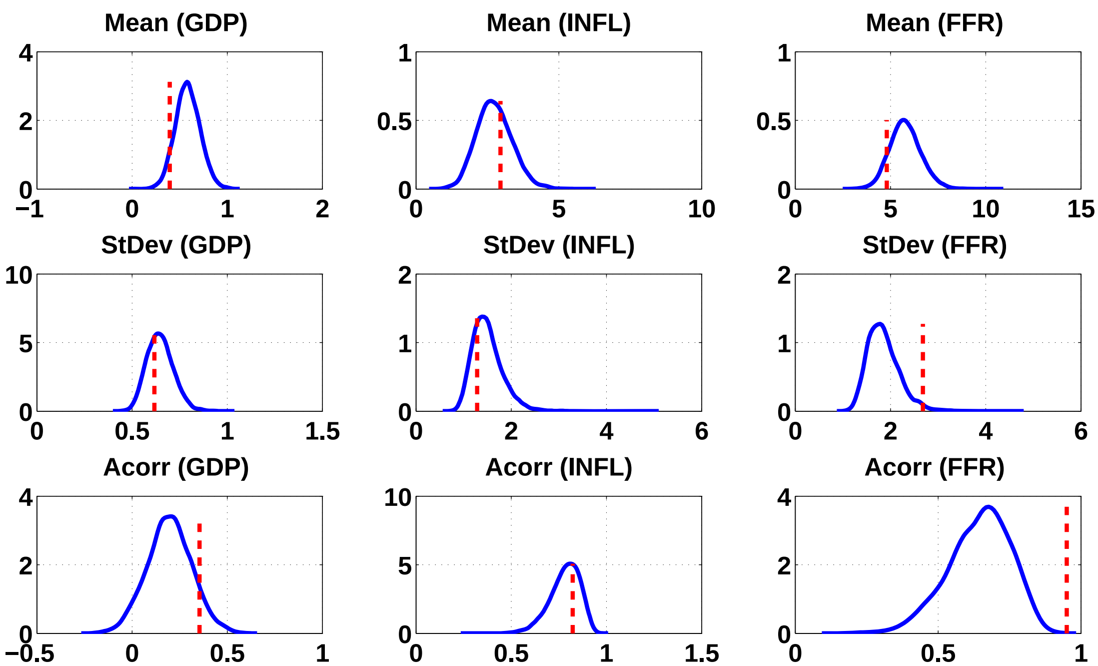
Notes: Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; and s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region.

Table 5: Posterior Estimates for DSGE Model Parameters

	Nonlinear		Linearized
Parameter	Mean	90% Interval	Mean
τ	1.05	[0.80, 1.35]	1.11
κ	0.03	[0.02, 0.05]	0.02
ψ_1	1.50	[1.43, 1.58]	1.49
ψ_2	1.51	[1.29, 1,78]	1.65
ρ_r	0.54	[0.38, 0.68]	0.56
ρ_g	0.89	[0.85, 0.92]	0.93
ρ_z	0.26	[0.10, 0.47]	0.26
r^A	0.70	[0.30, 1.08]	0.45
π^A	2.76	[2.74, 2.79]	2.78
γ^Q	0.57	[0.46, 0.71]	0.48
$100\sigma_r$	0.33	[0.24, 0.45]	0.36
$100\sigma_g$	0.88	[0.67, 1.16]	0.98
$100\sigma_z$	0.75	[0.51, 1.04]	0.85

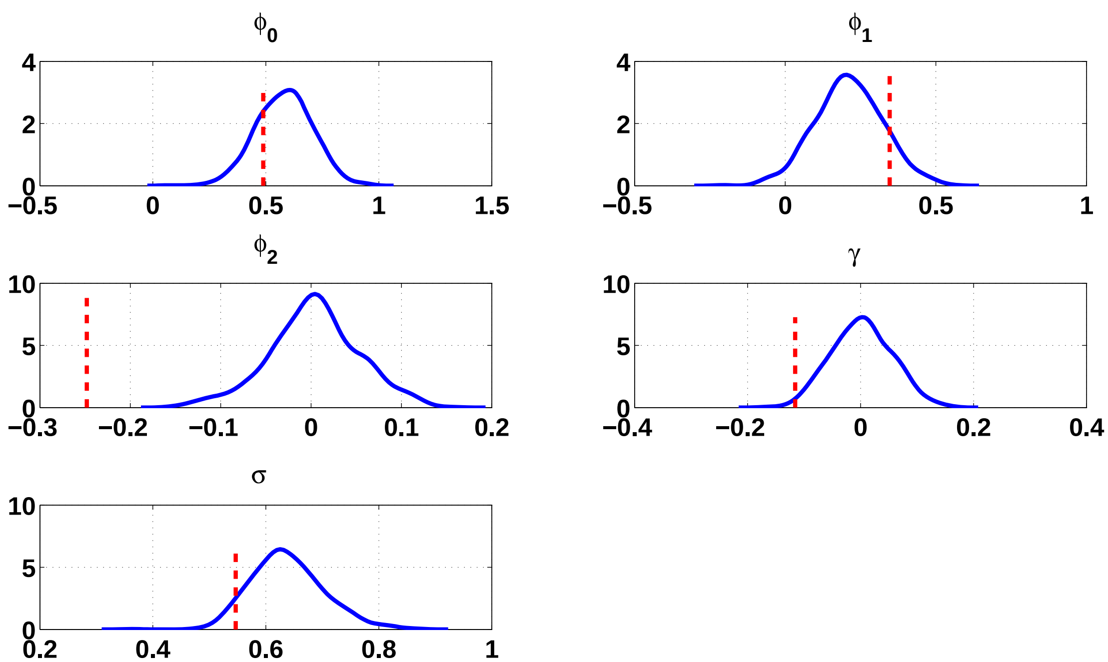
Notes: Estimation sample is 1984:Q1 to 2010:Q4. As 90% credible interval we are reporting the 5th and 95th percentile of the posterior distribution.

Figure 5: Predictive Check Based on Sample Moments



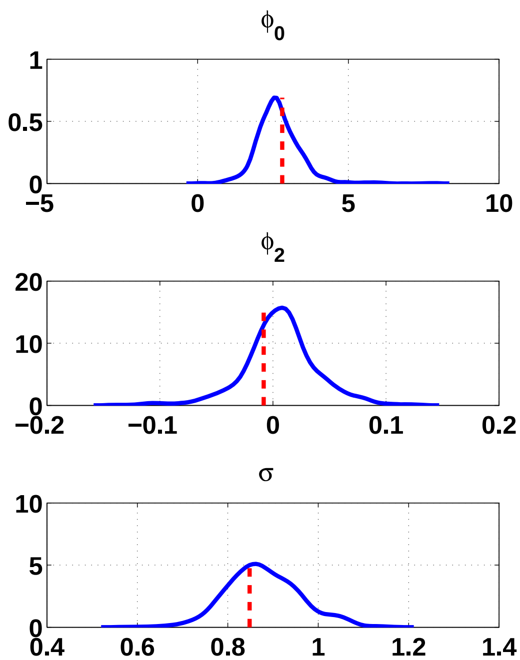
Notes: Vertical dashed lines indicate the sample means, standard deviations, and autocorrelations of U.S. GDP growth (GDP), inflation (INFL), and the Federal Funds Rate (INT) for the period 1984:Q1 to 2010:Q4. Solid lines are kernel density estimates of the DSGE model-implied posterior predictive densities for these sample moments. StDev = standard deviation, ACorr = autocorrelation.

Figure 6: Predictive Checks: QAR(1,1) for GDP Growth



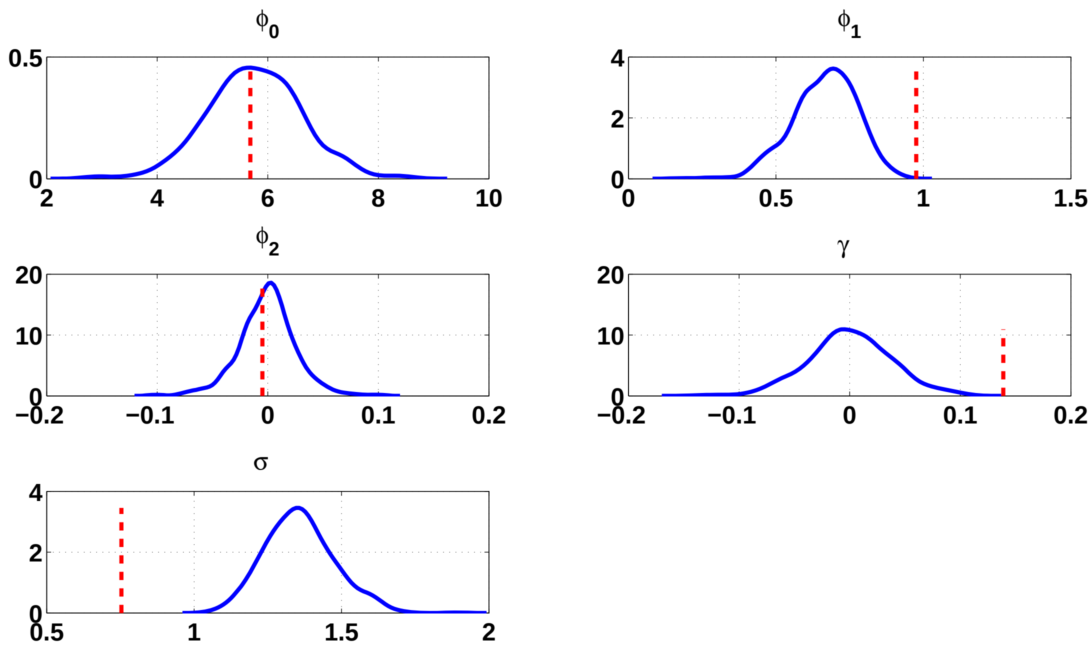
Notes: Vertical dashed lines indicate posterior mean estimates of the QAR(1,1) parameters based on U.S. data from 1984:Q1 to 2010:Q4. Solid lines are kernel density estimates of the DSGE model-implied posterior predictive densities for these estimates.

Figure 7: Predictive Checks: QAR(1,1) for Inflation



Notes: See Figure 6.

Figure 8: Predictive Checks: QAR(1,1) for Federal Funds Rate



Notes: See Figure 6.

Online Appendix: A New Class of Nonlinear Times Series Models for the Evaluation of DSGE Models

S. Borağan Aruoba, Luigi Bocola, and Frank Schorfheide

A Moments of the QAR(1,1) Model

This section derives important moments for the QAR(1,1) model given by

$$y_t = \phi_1 y_{t-1} + \phi_2 s_{t-1}^2 + (1 + \gamma s_{t-1})\sigma\epsilon_t, \quad \epsilon_t \sim iidN(0,1) \tag{A.1}$$

$$s_t = \phi_1 s_{t-1} + \sigma\epsilon_t, \quad |\phi_1| < 1. \tag{A.2}$$

The process s_t in (A.2) is linear and has a moving average representation of the from

$$s_t = \sigma \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}.$$

The first two moments of s_t are given by

$$\mathbb{E}[s_t] = 0 \quad \text{and} \quad \mathbb{E}[s_t^2] = \frac{\sigma^2}{1-\phi_1^2}.$$

Since the innovations ϵ_t are *iid* standard normal variates, we obtain that

$$\mathbb{E}[s_t^3] = \sum_{j=0}^{\infty} \phi_1^{3j} \mathbb{E}[\epsilon_{t-j}^3] = 0, \quad \mathbb{E}[s_t^4] = \sum_{j=0}^{\infty} \phi_1^{4j} \mathbb{E}[\epsilon_{t-j}^4] = \frac{3\sigma^4}{1-\phi_1^4}.$$

We proceed by calculating the time-invariant mean of the process y_t. Taking expectations on both sides of (A.1) we obtain

$$\begin{aligned} \mathbb{E}[y_t] &= \phi_1 \mathbb{E}[y_{t-1}] + \phi_2 \mathbb{E}[s_{t-1}^2] + (1 + \gamma \mathbb{E}[s_{t-1}])\sigma \mathbb{E}[\epsilon_t] \\ &= \phi_1 \mathbb{E}[y_{t-1}] + \frac{\phi_2 \sigma^2}{1-\phi_1^2}. \end{aligned}$$

Here we used the expression for $\mathbb{E}[s_{t-1}^2]$ obtained previously as well as the fact that ϵ_t and s_{t-1} are independent. In turn,

$$\mathbb{E}[y_t] = \frac{\phi_2 \sigma^2}{(1-\phi_1)(1-\phi_1^2)}. \tag{A.3}$$

Before we can calculate the uncentered second moment of y_t we need to derive $\mathbb{E}[y_t s_t^2]$:

$$\begin{aligned}
\mathbb{E}[y_t s_t^2] &= \mathbb{E}\big[(\phi_1 y_{t-1} + \phi_2 s_{t-1}^2 + (1+\gamma s_{t-1})\sigma\epsilon_t)(\phi_1^2 s_{t-1}^2 + 2\phi_1\sigma s_{t-1}\epsilon_t + \sigma^2\epsilon_t^2)\big] \\
&= \mathbb{E}\big[\phi_1^3 y_{t-1} s_{t-1}^2 + \phi_1^2\phi_2 s_{t-1}^4 + \sigma\phi_1^2(1+\gamma s_{t-1})s_{t-1}^2\epsilon_t \\
&\quad +2\phi_1^2\sigma y_{t-1}s_{t-1}\epsilon_t + 2\phi_1\phi_2\sigma s_{t-1}^3\epsilon_t + 2\phi_1\sigma^2(1+\gamma s_{t-1})s_{t-1}\epsilon_t \\
&\quad +\phi_1\sigma^2\epsilon_t^2 + \phi_2\sigma^2 s_{t-1}^2\epsilon_t^2 + \sigma^3(1+\gamma s_{t-1})\epsilon_t^3\big] \\
&= \phi_1^3\mathbb{E}[y_{t-1}s_{t-1}^2] + \phi_1^2\phi_2\mathbb{E}[s_{t-1}^4] + 2\phi_1\gamma\sigma^2\mathbb{E}[s_{t-1}^2] \\
&\quad +\phi_1\sigma^2\mathbb{E}[y_{t-1}] + \phi_2\sigma^2\mathbb{E}[s_{t-1}^2].
\end{aligned}$$

Plugging in the moments of s_t and solving for $\mathbb{E}[y_t s_t^2]$ under the assumption that the moment is time invariant, we obtain

$$\mathbb{E}[y_t s_t^2] = \frac{1}{1-\phi_1^3}\left[\frac{3\phi_1^2\phi_2\sigma^4}{1-\phi_1^4} + \frac{\sigma^2(\phi_2\sigma^2 + 2\phi_1\gamma\sigma)}{1-\phi_1^2} + \frac{\phi_1\phi_2\sigma^4}{(1-\phi_1)(1-\phi_2^2)}\right]. \tag{A.4}$$

We are now in a position to obtain the uncentered second moment of y_t:

$$\begin{aligned}
\mathbb{E}[y_t^2] &= \mathbb{E}\left[(\phi_1 y_{t-1} + \phi_2 s_{t-1}^2 + \sigma(1+\gamma s_{t-1})\epsilon_t)^2\right] \\
&= \mathbb{E}\big[\phi_1^2 y_{t-1}^2 + \phi_2^2 s_{t-1}^4 + \sigma^2(1+\gamma s_{t-1})^2\epsilon_t^2 \\
&\quad 2\phi_1\phi_2 y_{t-1}s_{t-1}^2 + 2\phi_2\sigma s_{t-1}^2(1+\gamma s_{t-1})\epsilon_t + 2\phi_1\sigma(1+\gamma s_{t-1})y_{t-1}\epsilon_t\big] \\
&= \phi_1^2\mathbb{E}[y_{t-1}^2] + \phi_2^2\mathbb{E}[s_{t-1}^4] + \sigma^2(1+\gamma^2\mathbb{E}[s_{t-1}^2]) + 2\phi_1\phi_2\mathbb{E}[y_{t-1}s_{t-1}^2].
\end{aligned}$$

Thus,

$$\mathbb{E}[y_t^2] = \frac{1}{1-\phi_1^2}\Big[\phi_2^2\mathbb{E}[s_{t-1}^4] + \sigma^2(1+\gamma^2\mathbb{E}[s_{t-1}^2]) + 2\phi_1\phi_2\mathbb{E}[y_{t-1}s_{t-1}^2]\Big]. \tag{A.5}$$

Notice that no further restriction on ϕ_1 is necessary to guarantee that the second moment is time invariant. Interestingly,

$$\begin{aligned}
\mathbb{E}[y_t s_t] &= \mathbb{E}\big[(\phi_1 y_{t-1} + \phi_2 s_{t-1}^2 + (1+\gamma s_{t-1})\sigma\epsilon_t)(\phi_1 s_{t-1} + \sigma\epsilon_t)\big] \\
&= \mathbb{E}\big[\phi_1^2 y_{t-1}s_{t-1} + \phi_1\phi_2 s_{t-1}^3 + \phi_1\sigma(1+\gamma s_{t-1})s_{t-1}\epsilon_t \\
&\quad \phi_1\sigma y_{t-1}\epsilon_t + \phi_2\sigma s_{t-1}^2\epsilon_t + \sigma^2(1+\gamma s_{t-1})\epsilon_t^2\big] \\
&= \phi_1^2\mathbb{E}[y_{t-1}s_{t-1}] + \sigma^2
\end{aligned}$$

All other terms drop out because $\mathbb{E}[\epsilon_t] = \mathbb{E}[s_t] = \mathbb{E}[s_t^3] = 0$. Thus, solving for $\mathbb{E}[y_t s_t]$ leads to the "first-order" variance expression

$$\mathbb{E}[y_t s_t] = \mathbb{E}[s_t^2] = \frac{\sigma^2}{1-\phi_1^2}.$$

The next step is to compute the (uncentered) autocovariances of y_t. Consider $\mathbb{E}[y_t y_{t-1}]$:

$$\begin{aligned}\mathbb{E}[y_t y_{t-1}] &= \mathbb{E}\big[(\phi_1 y_{t-1} + \phi_2 s_{t-1}^2 + (1+\gamma s_{t-1})\sigma\epsilon_t) y_{t-1}\big] \\ &= \phi_1 \mathbb{E}[y_{t-1}^2] + \phi_2 \mathbb{E}[s_{t-1}^2 y_{t-1}].\end{aligned}$$

In general, higher-order autocovariances can be computed recursively:

$$\begin{aligned}\mathbb{E}[y_t y_{t-h}] &= \mathbb{E}\big[(\phi_1 y_{t-1} + \phi_2 s_{t-1}^2 + (1+\gamma s_{t-1})\sigma\epsilon_t) y_{t-h}\big] \\ &= \phi_1 \mathbb{E}[y_{t-1} y_{t-h}] + \phi_2 \mathbb{E}[s_{t-1}^2 y_{t-h}].\end{aligned}$$

The expectation $\mathbb{E}[s_{t-1}^2 y_{t-h}]$ can also be calculated recursively:

$$\begin{aligned}\mathbb{E}[s_t^2 y_{t-h}] &= \mathbb{E}\big[(\phi_1 s_{t-1} + \sigma\epsilon_t)^2 y_{t-h}\big] \\ &= \phi_1^2 \mathbb{E}[s_{t-1}^2 y_{t-h}] + \sigma^2 \mathbb{E}[y_{t-h}].\end{aligned}$$

(*** Try to simplify some of the formulas by considering centered moments. ***)

In order to obtain a proposal distribution for γ and σ^2 in the MCMC algorithm the following calculation is useful. Let

$$u_t = (1 = \gamma s_{t-1})\epsilon_t.$$

Then,

$$\begin{aligned}\mathbb{E}[u_t^2] &= \sigma^2(1 + \gamma^2 \mathbb{E}[s_{t-1}^2]) \\ \mathbb{E}[u_t^2 s_{t-1}^2] &= \sigma^2(\mathbb{E}[s_{t-1}^2] + \gamma^2 \mathbb{E}[s_{t-1}^4])\end{aligned}$$

This system of moment conditions can be solved for σ^2 and $\gamma^2\sigma^2$:

$$\begin{bmatrix} \sigma^2 \\ \gamma^2\sigma^2 \end{bmatrix} = \begin{bmatrix} 1 & \mathbb{E}[s_{t-1}^2] \\ \mathbb{E}[s_{t-1}^2] & \mathbb{E}[s_{t-1}^4] \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}[u_t^2] \\ \mathbb{E}[u_t^2 s_{t-1}^2] \end{bmatrix}.$$

Thus, preliminary estimates of σ^2 and $\gamma^2\sigma^2$ can be obtained by estimating the auxiliary regression

$$u_t^2 = (\sigma^2) + (\sigma^2\gamma^2)s_{t-1}^2 + \text{residuals}. \tag{A.6}$$

The auxiliary regression does not determine the sign of γ. In order to choose the sign of γ the following additional moment condition is helpful:

$$\mathbb{E}[u_t^2 s_{t-1}] = \sigma^2 \mathbb{E}[(1 + 2\gamma s_{t-1} + \gamma^2 s_{t-1}^2)s_{t-1}] = 2\sigma^2\gamma \mathbb{E}[s_{t-1}^2].$$

Thus, a negative correlation between u_t^2 and s_{t-1} indicates that γ should be negative.