

Discussion
Papers

Deutsches Institut für Wirtschaftsforschung

2013

A Noncausal Autoregressive Model
with Time-Varying Parameters

An Application to U.S. Inflation

Markku Lanne and Jani Luoto

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2013

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<http://www.diw.de>

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<http://ideas.repec.org/s/diw/diwwpp.html>
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

A Noncausal Autoregressive Model with Time-Varying Parameters: An Application to U.S. Inflation*

by

Markku Lanne

Jani Luoto

University of Helsinki

University of Helsinki

Abstract

We propose a noncausal autoregressive model with time-varying parameters, and apply it to U.S. postwar inflation. The model fits the data well, and the results suggest that inflation persistence follows from future expectations. Persistence has declined in the early 1980's and slightly increased again in the late 1990's. Estimates of the new Keynesian Phillips curve indicate that current inflation also depends on past inflation although future expectations dominate. The implied trend inflation estimate evolves smoothly and is well aligned with survey expectations. There is evidence in favor of the variation of trend inflation following from the underlying marginal cost that drives inflation.

JEL Classification: C22, C51, C53, E31

*We thank the Academy of Finland and the OP-Pohjola Group Research Foundation for financial support. This research was completed while the first author was visiting the German Institute for Economic Research in Berlin, whose hospitality is gratefully acknowledged.

1 Introduction

Different univariate time series models have been successfully applied to describe inflation dynamics and to forecast inflation. Recently the family of models considered in this literature has been extended to noncausal autoregressive (AR) models that, in contrast to conventional causal AR models, allow for explicit dependence on future values. In particular, Lanne and Saikkonen (2011) introduced a new noncausal AR model and applied it to U.S. consumer price inflation, while Lanne, Luoma, and Luoto (2012) considered the Bayesian analysis of this model with an application to the same inflation series. The general conclusion of these papers is that despite being highly autocorrelated, inflation is dominated by future expectations, with past values having a lesser effect. In addition to accurately describing inflation dynamics, noncausal autoregressive models have also proved superior in forecasting. Lanne, Luoma, and Luoto (2012), and Lanne, Luoto, and Saikkonen (2012) demonstrate significant improvement over conventional causal AR models in the accuracy of forecasts of U.S. consumer price and GDP inflation, respectively, especially at longer forecast horizons.

In this paper, we extend the noncausal autoregressive model of Lanne and Saikkonen (2011) by allowing for time-varying parameters. While our extension may turn out to be useful in modeling many economic time series, we primarily focus on inflation for two reasons. First, in view of the good performance of the noncausal autoregressive models for inflation in the previous literature, it is interesting to see whether anything can be gained by allowing for time-varying parameters. Second, models with time-varying parameters are increasingly being employed in the empirical literature on inflation. This is especially the case when the primary concern is modeling trend inflation, usually interpreted as public's perception of the inflation goal of the central bank (see, e.g., Cogley and Sbordone (2008), and Cogley et al. (2010)).

Because of its complexity, we estimate the new model by Bayesian methods, as has typically been done also in the previous empirical literature on models with time-varying parameters. The estimation method is built upon our previous work on Bayesian analysis of noncausal AR models (see Lanne, Luoma, and Luoto (2012)). In

particular, we propose a non-standard simulation algorithm for the joint estimation of the time-varying parameters and future innovations. The non-standard features of the algorithm are related to the future innovations required for the estimation, but given the future innovations, the standard methods for state space models, including Kalman filtering and simulation smoothing, can be used for all the parameters (see, e.g., Cogley and Sargent (2005)).

In addition to estimation and forecasting in the unrestricted noncausal AR model with time-varying parameters, we also consider the estimation of the new Keynesian Phillips curve (NKPC) based on the new model, by placing additional restrictions along the lines of Lanne and Luoto (2013), who used a noncausal AR model with constant parameters to this end. A central problem in the estimation of the NKPC is that the model depends on an unobserved marginal cost variable that is difficult to measure, but estimation based on the noncausal AR model has the advantage that no marginal cost proxy is needed, but the variable is latent. However, assuming constancy of the parameters of the AR model, as in Lanne and Luoto (2013), leads to ignoring the effect of structural breaks due to technological changes, among other things, that may have taken place over time. Therefore, it is interesting to see, to what extent having time-varying parameters in the AR model affects the general conclusions.

Our empirical results concern postwar quarterly U.S. consumer price inflation. A purely noncausal AR model is selected for this series, as was also the case with the corresponding constant-parameter models in Lanne and Saikkonen (2011), and Lanne, Luoma and Luoto (2012). In accordance with previous empirical results, inflation persistence seems to have decreased considerably since the early 1980's, with a slight increase in persistence again since the late 1990's. In forecasting, the varying-parameter noncausal model turns out to be superior to both the corresponding constant-parameter noncausal AR model and the varying-parameter causal AR model by a relatively wide margin.

Even though the purely noncausal AR model is selected, in estimating the (hybrid) NKPC, we allow for dependence on past inflation to facilitate gauging its importance in the new Keynesian model. When more structure is imposed in the form of the

NKPC restrictions, we find that both past inflation and the (implicit) marginal cost also have a role in the determination of U.S. inflation, in accordance with the new Keynesian theory. The NKPC estimates suggest that changes in persistence have been inherited from the process driving inflation instead of following from changes in the parameters of the NKPC, in contrast to some recent literature, including Cogley and Sbordone (2008). Moreover, not allowing for time variation seems to exaggerate the importance of inflation expectations in determining current inflation.

Our model also lends itself to estimating the evolution of trend inflation. As recently pointed out by Chan et al. (2012), the common practice in the literature of using a driftless random walk specification has the counter-intuitive implication that trend inflation becomes unbounded. As a solution to this problem, Chan et al. (2012) impose explicit bounds on the trend inflation; in the framework of the noncausal AR model, the NKPC restrictions seem to work in a way similar to such bounds. In particular, while the trend inflation implied by the unrestricted noncausal AR model turns out to follow observed inflation quite closely in the same way as many previous estimates (including those based on the UC-SV model of Stock and Watson (2007) and the Trend-SV model of Chan et al. (2012)), the evolution of trend inflation implied by our model incorporating the NKPC restrictions is smooth and follows survey expectations quite accurately.

The rest of the paper is structured as follows. In Section 2, we describe the noncausal AR model with time-varying parameters and specify the priors of its parameters. In Section 3, the details of estimation and inference are presented, while in Section 4, the estimation of the NKPC based on a noncausal AR model with time-varying parameters is discussed. Section 5 contains the empirical results. Finally, Section 6 concludes. The details on the estimation algorithm as well as discussion on its efficiency are deferred to the appendix.

2 Model

In this section, we first describe the noncausal AR model with time-varying parameters, which is a generalization of the constant-parameter noncausal AR model of

Lanne and Saikkonen (2011). As the model will be estimated by Bayesian methods, after its description, we discuss the hierarchical prior distributions for the latent time-varying parameters to be used in the subsequent empirical application to U.S. inflation in Section 5. Hierarchical priors have typically been employed in the estimation of latent-variable models, and the specific priors we use are common in the macroeconomic literature on time-varying-parameter (TVP) models (see, e.g., Primiceri (2005), Koop and Potter (2011), and Chan et al. (2012)). The popularity of the Bayesian approach in the econometric analysis of such models follows from the fact that it facilitates reducing the problem to a set of tractable sub problems.

2.1 Noncausal AR Model with Time-Varying Parameters

Let us assume that inflation y_t (or some other stochastic process) is generated by

$$\phi_t(B)\varphi_t(B^{-1})y_t = c_t + \epsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where $\phi_t(B) = 1 - \phi_{t1}B - \dots - \phi_{tr}B^r$, $\varphi_t(B^{-1}) = 1 - \varphi_{t1}B^{-1} - \dots - \varphi_{tr}B^{-s}$, B is the usual backward shift operator, c_t denotes the time varying intercept, and ϵ_t is a white noise process with mean zero and time-varying scale parameter σ_t^2 .¹ We use the abbreviation TVP-AR(r, s) for the varying-parameter model defined by (1). If $\varphi_{t1} = \dots = \varphi_{ts} = 0$ for all $t = 1, \dots, T$, model (1) reduces to the causal TPV-AR($r, 0$) model with y_t depending on its past but not future values. The more interesting cases arise, when this restriction does not hold. If $\phi_{t1} = \dots = \phi_{tr} = 0$ for all $t = 1, \dots, T$, we have the purely noncausal TVP-AR($0, s$) model with dependence on future values only. In the mixed TVP-AR(r, s) case where neither restriction holds, y_t depends on its past as well as future values.

¹In the noncausal AR model with constant parameters, a non-Gaussian error distribution is required for identification (see Lanne and Saikkonen (2011)). While no proof is given, our experimentation with Gaussian errors indicate that this is not the case when (part of) the parameters are time varying. Nevertheless, in Section 5, following Lanne and Saikkonen (2011), and Lanne, Luoma, and Luoto (2012), a Student's t distribution (instead of a Gaussian distribution typically entertained in the previous literature) will be assumed for improved fit. This family of distributions has also recently been introduced into macroeconomic models by Chib and Ramamurthy (2011) and Cúrdia et al. (2012).

We assume that the polynomials $\phi_t(z)$ and $\varphi_t(z)$ have their zeros outside the unit circle for $t = 1, \dots, T$, so that

$$\phi_t(z) \neq 0 \text{ for } |z| \leq 1 \text{ and } \varphi_t(z) \neq 0 \text{ for } |z| \leq 1. \quad (2)$$

In what follows, it will sometimes be convenient to express the model in terms of the processes $u_t = \varphi_t(B^{-1})y_t$ and $v_t = \phi_t(B)y_t$. The former condition in (2) guarantees that the process u_t has the following backward moving average presentation,

$$u_t = \sum_{j=0}^{\infty} \alpha_{tj} w_{t-j}, \quad (3)$$

where $w_t = c_t + \epsilon_t$ and α_{tj} is the coefficient of z^j in the power series expansion of $\phi_t(B)^{-1}$. Similarly, by the condition that $\varphi_t(z)$ has its zeros outside the unit circle, the process v_t has the forward moving average representation

$$v_t = \sum_{j=0}^{\infty} \beta_{tj} w_{t+j}, \quad (4)$$

with β_{tj} the coefficient of z^j in the power series expansion of $\varphi_t(B^{-1})^{-1}$.

For further interpretation, we may express y_t in terms of past values and future errors. By writing (1) as

$$\phi_t(B) \varphi_t(B^{-1}) \left(y_t - [\phi_t(B)]^{-1} [\varphi_t(B^{-1})]^{-1} c_t \right) = \epsilon_t,$$

denoting the time-varying mean of y_t as $\mu_t \equiv [\phi_t(B)]^{-1} [\varphi_t(B^{-1})]^{-1} c_t$ and multiplying the resulting equation by $\varphi_t(B^{-1})^{-1}$, we obtain

$$\begin{aligned} y_t - \mu_t &= \phi_{t,1} (y_{t-1} - \mu_{t-1}) + \dots + \phi_{t,r} (y_{t-r} - \mu_{t-r}) \\ &\quad + \sum_{j=0}^{\infty} \beta_{tj} \epsilon_{t+j}, \end{aligned} \quad (5)$$

with β_{tj} as in (4). This shows that, in the general case, (demeaned) y_t depends on its own lags and an infinite number of future errors. Because the errors are dependent on past values of y_t , the conditional expectation of the last term is, in general, different from zero, and, hence, also the conditional expectation of y_t depends on future errors. Unfortunately, this also complicates the computation of forecasts, but to that end, we suggest the following straightforward extension of the procedures that Lanne, Luoma,

Luoto (2012), and Lanne, Luoto, and Saikkonen (2012) proposed for the constant-parameter noncausal AR model. Multiplying (1) by $\varphi_t (B^{-1})^{-1}$, using (4), and leading the resulting equation by $h > 0$ periods, we obtain

$$y_{T+h} = \phi_{T+h,1}y_{T+h-1} + \cdots + \phi_{T+h,r}y_{T+h-r} + \sum_{j=0}^{\infty} \beta_{T+h,j}w_{T+h+j}, \quad (6)$$

which can be used to recursively compute the optimal (density) forecast of y_{T+h} conditional on information in period T , provided we are able to evaluate the conditional density of $v_{T+h} = \sum_{j=0}^{\infty} \beta_{T+h,j}w_{T+h+j}$. In practice, the latter and, hence, the forecasts are obtained as a by-product of the estimation algorithm (see Section 3 below). To compute v_{T+h} , following Lanne, Luoma, and Luoto (2012), and Lanne, Luoto, and Saikkonen (2012), we approximate it by $\sum_{j=0}^M \beta_{t,j}w_{t+j}$, where M is a positive integer. The approximation error can be made negligible by setting M sufficiently large.

As mentioned in the Introduction, the TVP-AR model also facilitates computing trend inflation that we denote by \bar{y}_t . We define it by the moving endpoint concept of Kozicki and Tinsley (1998) which is standard in the literature, i.e., $\bar{y}_t = \lim_{h \rightarrow \infty} E_t y_{t+h}$ (this equals the Beveridge and Nelson (1981) decomposition if there are no deterministic terms). Leading equation (5) by h periods and taking conditional expectations conditional on information at period t , we obtain

$$\begin{aligned} E_t (y_{t+h} - \mu_{t+h}) &= \phi_{t+h,1} E_t (y_{t+h-1} - \mu_{t+h-1}) + \cdots + \phi_{t+h,r} E_t (y_{t+h-r} - \mu_{t+h-r}) \\ &\quad + E_t \left(\sum_{j=0}^{\infty} \beta_{t+h,j} \epsilon_{t+h+j} \right). \end{aligned}$$

Because ϵ_{t+h} is virtually unpredictable by information at period t for $h \rightarrow \infty$, the last term converges to zero, and under the stationarity condition (2), this is the case also for the other terms on the right-hand side. Thus, a close approximation for $\bar{y}_t = \lim_{h \rightarrow \infty} E_t y_{t+h}$ is given by $\lim_{h \rightarrow \infty} E_t \mu_{t+h}$, which, following Cogley and Sargent (2005), we approximate by $[\phi_t(1)]^{-1} [\varphi_t(1)]^{-1} c_t$, calculated from the joint posterior distribution of the intercept and AR coefficients for $t = 1, \dots, T$.

2.2 Prior Distributions

As mentioned above, the model is estimated by Bayesian methods. To that end, we first specify the hierarchical prior distributions to be used for the time-varying parameters. All the priors are standard, and have previously been used in empirical macroeconomics by, inter alia, Cogley and Sargent (2001, 2005), Primiceri (2005), and Cogley and Sbordone (2008). Throughout, underlining a parameter refers to a prior hyperparameter set by the researcher.

We assume that the error term ϵ_t follows Student's t distribution with location parameter zero, scale parameter σ_t^2 , and λ degrees of freedom. To make the model operational, we employ the following parameterization (see Geweke (1993)):

$$\epsilon_t = \sigma_t \tilde{h}_t^{-1/2} \eta_t, \quad (7)$$

where η_t is a standard normal random variable, and $\lambda \tilde{h}_t$ is chi-square distributed with λ degrees of freedom ($\chi^2(\lambda)$). While a non-Gaussian error distribution may not be necessary for the identification of the AR coefficients $\boldsymbol{\phi}_t = (\phi_{t1}, \dots, \phi_{tr})'$ and $\boldsymbol{\varphi}_t = (\varphi_{t1}, \dots, \varphi_{ts})'$ in the TVP-AR model (unlike the noncausal AR model with constant parameters, see footnote 1), a Student's t error distribution is nevertheless useful for improved fit. Under the chosen parameterization, y_t following (1) is conditionally Gaussian conditional on σ_t^2 and \tilde{h}_t . As will be seen in Section 3, this property is critical in building a decent posterior sampler (see also Lanne, Luoma, and Luoto (2012)). Notice that the random vector $\tilde{\mathbf{h}}^T = (\tilde{h}_1, \dots, \tilde{h}_T)'$ can also be interpreted as a vector of parameters with hierarchical priors $\lambda \tilde{h}_t \sim \chi^2(\lambda)$ and $\lambda \sim \text{Exp}(\underline{\lambda})$, where $\underline{\lambda}$ is a hyperparameter.

Our hierarchical prior density for $\tilde{\sigma}_t \equiv \log \sigma_t$ can be defined through the random walk process

$$\tilde{\sigma}_t = \tilde{\sigma}_{t-1} + \xi_t, \quad (8)$$

where $\xi_t \sim N(0, \omega^2)$, and for the innovation variance ω^2 we assume an inverse-gamma prior density with shape and rate parameters \underline{a} and \underline{b} , respectively, i.e., $p(\omega^2) = i\text{Gamma}(\omega^2 | \underline{a}, \underline{b})$. The obvious drawback of this commonly used prior is that, because of its random walk structure, $\tilde{\sigma}_t$ hits any upper or lower bound with probability one.

Nevertheless, the random walk model tends to work well for $\tilde{\sigma}_t$ in practice as our results in Section 5 as well as those of Chan et al. (2012) and Primiceri (2005) suggest. The random walk assumption (8) has also the advantage that it reduces the number of estimated parameters (compared to the common AR(1) assumption; see, e.g., Fernández-Villaverde and Rubio-Ramírez (2007)).

We assume the following dynamics for the time-varying intercept and the AR coefficients $\boldsymbol{\theta}_t = (c_t, \boldsymbol{\phi}'_t)'$ and $\boldsymbol{\varphi}_t$:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\varepsilon}_{1t} \quad (9)$$

and

$$\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_{t-1} + \boldsymbol{\varepsilon}_{2t}, \quad (10)$$

where the innovation vectors $\boldsymbol{\varepsilon}_{1t}$ and $\boldsymbol{\varepsilon}_{2t}$ are assumed to be prior independent. In particular, $\boldsymbol{\varepsilon}_{1t}$ and $\boldsymbol{\varepsilon}_{2t}$ are multivariate normal with zero mean and covariance matrices \mathbf{Q}_1 and \mathbf{Q}_2 , respectively. Following the literature, we assume an inverse-Wishart prior for the innovation variance, i.e., $f(\mathbf{Q}_i) = iW(\mathbf{Q}_i | \underline{\nu}_i, \underline{\mathbf{Q}}_i)$, where the degrees of freedom parameters $\underline{\nu}_i$, and the scale matrices $\underline{\mathbf{Q}}_i$ are the prior hyperparameters assumed to be known by the researcher.

3 Simulation method

Lanne, Luoma, and Luoto (2012) considered Bayesian inference in the noncausal AR model with constant parameters, and in this section we discuss how their simulation method can be extended to estimate the parameters of the TVP-AR model. Unfortunately, allowing for time variation brings about complications. In particular, even though it is possible to write up the likelihood function of the TVP-AR model (1), it is by construction complex and cannot be successfully used in posterior inference. Therefore, we implement estimation by a Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. Here we sketch the algorithm, whose details can be found in Appendix A.

To simplify notation, we stack the row vectors $\boldsymbol{\theta}'_t$ and $\boldsymbol{\varphi}'_t$ for $t = 1, \dots, T$ into the $T \times (r + 1)$ and $T \times s$ matrices $\boldsymbol{\theta}^T$ and $\boldsymbol{\varphi}^T$, respectively, and define the $(T \times 1)$

vector $\boldsymbol{\sigma}^T = (\sigma_1, \dots, \sigma_T)'$. The proposed sampler exploits the full conditional posterior distributions of groups of unobservables, containing the latent variables $\boldsymbol{\theta}^T$, $\boldsymbol{\varphi}^T$, $\boldsymbol{\sigma}^T$, $\tilde{\mathbf{h}}^T$, y_{T+1}, \dots, y_{T+s} (five groups) as well as the prior hyper parameters ω , \mathbf{Q}_i ($i = 1, 2$), and λ (four groups). Simulating in turn from the unrestricted conditional posterior distributions of $\boldsymbol{\theta}^T$ and $\boldsymbol{\varphi}^T$ is straightforward, and it can be done by a standard simulation smoother for Gaussian state space models, such as those in Carter and Kohn (1994) or Durbin and Koopman (2002), the latter of which is used in this paper. This can be seen by expressing (1) in terms of u_t or v_t (see Lanne, Luoma, and Luoto (2012)), which results in linear but non-Gaussian observation equations for the AR parameters $\boldsymbol{\theta}_t$ and $\boldsymbol{\varphi}_t$. The Gaussian state space representation is then obtained by substituting (7) into these observation equations and using (9) and (10) as the state transition equations for $\boldsymbol{\theta}_t$ and $\boldsymbol{\varphi}_t$, respectively. We impose the stationarity restrictions (2) by discarding the draws from the unrestricted posterior distributions of $\boldsymbol{\theta}^T$ or $\boldsymbol{\varphi}^T$ that do not lie in the stationary region. This is similar to Cogley and Sargent (2005); see also the discussion in Koop and Potter (2011).

We simulate from the conditional posterior distribution of $\boldsymbol{\sigma}^T$ using the method of Kim et al. (1998), who propose a mixture of normals approximation of the logarithm of the squared errors that are $\log \chi^2(1)$ distributed. With this approximation, we can transform the model into a Gaussian one, obtaining a Gaussian observation equation for (8) using (7). Thus, the simulation smoother of Durbin and Koopman (2002) can be used for the state space model defined by the resulting observation equation and (8). The conditional posterior density of the vector of scale parameters $\tilde{\mathbf{h}}^T$ can be obtained by simulation in the usual way. In particular, its independent elements \tilde{h}_t ($t = 1, \dots, T$) can be drawn using a chi-square distribution, as explained in Geweke (2005).

The future observations y_{T+1}, \dots, y_{T+s} (or equivalently w_{T+1}, \dots, w_{T+M}) cannot be drawn in a straightforward manner. To that end, we follow Chib and Greenberg (1994) in using a component of the posterior density as a proposal density for a Metropolis-Hastings (MH) step. In particular, we use the hierarchical prior distribution of w_{T+1}, \dots, w_{T+M} to draw proposals for y_{T+1}, \dots, y_{T+s} which are then accepted with a probability based solely on (1). Thus, we do not need to evaluate the unknown

joint density of $(y_{T+1}, \dots, y_{T+s})$, but we can simulate from it by a straightforward extension of the forecasting procedures of Lanne, Luoma, and Luoto (2012), and Lanne, Luoto, and Saikkonen (2012) discussed in Section 2.1 above. In particular, a draw from the proposal distribution of y_{T+1}, \dots, y_{T+s} (or equivalently w_{T+1}, \dots, w_{T+M}) is obtained by simulating the random variables w_{T+1}, \dots, w_{T+M} in the last term of (6) from their hierarchical prior distribution and using these simulated values to compute y_{T+h} recursively for $h = 1, 2, \dots, s$ using (6). A Student's t distribution with location parameter c_T , scale parameter σ_T^2 , and λ degrees of freedom is used to simulate w_{T+1}, \dots, w_{T+M} , and the AR coefficients $\phi_{T+1}, \dots, \phi_{T+s}$ and $\varphi_{T+1}, \dots, \varphi_{T+s}$ are simulated from the random walk processes (9) and (10), respectively. This strategy seems to work well in practice, with acceptance rates in the MH step typically exceeding 50%.

As far as the hyper parameters ω and \mathbf{Q}_i ($i = 1, 2$) are concerned, they can be simulated directly using the inverse-Gamma and inverse-Wishart distributions, respectively. Following Geweke (2005), we simulate from the conditional posterior of the degree-of-freedom parameter λ using an independence-chain MH algorithm.

4 New Keynesian Phillips Curve

The TVP-AR model can be used to estimate the new Keynesian Phillips curve, which is a central building block of modern macroeconomic models that can be derived from several sets of microfoundations. According to its hybrid version often attributed to Galí and Gertler (1999),

$$y_t = \gamma_f E_t y_{t+1} + \gamma_b y_{t-1} + \gamma x_t, \quad (11)$$

inflation y_t depends on expected future inflation, lagged inflation and a measure of marginal cost x_t . Lanne and Luoto (2013) showed how the NKPC can be estimated based on a univariate noncausal constant-parameter AR model specified for inflation. A great advantage of their procedure is that it requires no proxy for the marginal cost that is difficult to measure. However, they implicitly assume a time series process with constant parameters for the marginal cost variable, which may be restrictive in that, among other things, it does not account for structural changes likely to have taken

place over multiple decades. The estimation of the NKPC based on the TVP-AR model (1), in contrast, facilitates taking such effects into account.

The relation of the parameters of the (hybrid) NKPC to those of the TVP-AR model can be seen as follows. As pointed out in Lanne and Luoto (2013), the hybrid NKPC has the following expression:

$$(1 - \phi B) (1 - \varphi B^{-1}) y_t = z_t, \quad (12)$$

where

$$\phi = \frac{1}{2} \left(\gamma_f^{-1} - \sqrt{\gamma_f^{-2} - 4\gamma_f^{-1}\gamma_b} \right) \quad \text{and} \quad \varphi^{-1} = \frac{1}{2} \left(\gamma_f^{-1} + \sqrt{\gamma_f^{-2} - 4\gamma_f^{-1}\gamma_b} \right),$$

and z_t is a linear combination of the expectation error and the marginal cost variable x_t . If z_t were independently and identically distributed, (12) would be the time invariant noncausal AR(1,1) model of Lanne and Saikkonen (2011). However, the marginal cost is likely to be autocorrelated, which is supported by the persistence of the theoretically implied variables driving inflation. To allow for autocorrelation in z_t , we assume that its dynamics can be successfully approximated by a (potentially noncausal) TVP-AR($r - 1, s - 1$) process:

$$a_t(B) b_t(B^{-1}) z_t = c_t + \epsilon_t, \quad (13)$$

where $a_t(B) = 1 - a_{t1}B - \dots - a_{t,r-1}B^{r-1}$, $b_t(B^{-1}) = 1 - b_{t1}B^{-1} - \dots - b_{t,s-1}B^{-s+1}$, and c_t and ϵ_t are defined as in (1). Substituting (13) into (12) yields

$$a_t(B) b_t(B^{-1}) (1 - \phi B) (1 - \varphi B^{-1}) y_t = c_t + \epsilon_t. \quad (14)$$

Defining $\phi_t(B) \equiv a_t(B) (1 - \phi B)$ and $\varphi_t(B^{-1}) \equiv b_t(B^{-1}) (1 - \varphi B^{-1})$, it can be seen that (14) is the TVP-AR(r, s) process with the restriction that two of the parameters are constant over time. Estimation of the TVP-AR model under these restrictions thus yields estimates of the parameters of the NKPC.

We use the abbreviation TVP-NKPC(r, s) for the restricted varying-parameter model defined by (14), which is more general than that used by Lanne and Luoto (2013) in that it allows for a time-varying error process. The prominent source of

such variation are potential structural breaks in the marginal cost variable caused by technological changes and shifts in policy regimes that we assume to be captured by the time-varying parameters of (13). Alternatively, or in addition to this, the effect of marginal cost on inflation may vary over time, i.e., γ in (11) may be time varying, giving rise to time variation in (13). The MCMC algorithm for the TVP-NKPC model is a straightforward modification of the algorithm discussed in Section 3 above; the details of the modifications required are given in Appendix B.

One advantage of our approach in the TVP framework is that it can capture the effect of potentially nonzero steady-state inflation. As recently pointed out by Sahuc (2006), if steady-state inflation is different from zero (in contrast to what is assumed in deriving (11)), the true NKPC contains an additional forward-looking term, which is then included in z_t and, thus, does not affect the estimation of coefficients γ_f and γ_b . As steady-state inflation very likely differs from zero, this is an obvious advantage although Sahuc (2006) also shows that the effect of assuming a zero-inflation steady state on the coefficients of the NKPC is minor when prices are highly indexed.

The shifting trend involved in (14) can be interpreted as evolving trend inflation in the same way as in the unrestricted TVP-AR model, and in this case the time variation in trend inflation can be thought of as being inherited from the (unobserved) marginal cost variable x_t . This is different from Cogley and Sbordone (2008) in that we assume a constant (potentially nonzero) steady state, while their approach is based on a shifting steady state of the Calvo pricing model with time-invariant structural parameters, leading to a NKPC with time-varying coefficients. It can be shown that replacing ϕ and φ in (14) with time-varying coefficients yields a TVP-AR(r,s) representation of Cogley and Sbordone's (2008) NKPC, and hence, comparison of the unrestricted TVP-AR model (1) and the restricted TVP-NKPC model (14) provides information on the source of time variation in trend inflation. Specifically, the significance of the restrictions underlying (14) can be interpreted in favor of the marginal cost variable x_t (potentially undergoing structural changes) as the major factor explaining the shifting trend inflation instead of an evolving steady state.

5 Empirical Results

We estimate models (1) and (14) for U.S. inflation computed as $y_t = 400 \times \log(P_t/P_{t-1})$, where P_t is the consumer price index (CPI) for all consumers. Our quarterly data set covers the period from 1947:1 to 2012:2, and the source of the data is the FRED database of the Federal Reserve Bank of St. Louis.

In estimation, we use the priors discussed in Section 2.2 above. The values of the prior hyperparameters of the TVP-AR(r, s) models are mainly taken from Chan et al. (2012), who also considered a similar quarterly inflation series. The prior mean $\underline{\lambda}$ is set at 5. The starting values of the time-varying parameters are set at zero and the variances associated with the logarithmic scale and intercept parameters and the AR coefficients at 5 and 1, respectively. The AR coefficients are assumed prior independent, and the roots of the polynomials with the initial AR coefficients $\varphi_0(z)$ and $\phi_0(z)$ are assumed to lie outside the unit circle. We assume that $\underline{\nu}_1 = \underline{\nu}_2 = 10$ and that the scale matrices $\underline{\mathbf{Q}}_1$ and $\underline{\mathbf{Q}}_2$ are diagonal. In particular, their diagonal elements are set such that the prior means of the innovation variances of the time-varying intercept and AR coefficients equal 0.141^2 and 0.0316^2 , respectively. For ω^2 , the shape and rate parameters \underline{a} and \underline{b} are set in such a way that the prior inverse Gamma density has 10 degrees of freedom, and the prior mean of the innovation variance ω^2 is 0.224^2 . For a detailed motivation of the values of these hyperparameters we refer to Chan et al. (2012). Our results also suggest that the estimates based on the chosen prior distribution of the degree-of-freedom parameter are identical to those based on less informative priors, indicating that estimation is not driven by the priors. However, it is our experience that faster convergence is achieved when the priors are not flat.

5.1 Unrestricted TVP-AR Model

Following Lanne and Saikkonen (2011), we start by specifying a Gaussian TVP-AR($l, 0$) model and check whether the residuals are serially uncorrelated and normally distributed. To that end, we calculate the Bayes p values $P(T(\mathbf{y}^T, \Theta) \leq T(\mathbf{y}^{pred}, \Theta) | \mathbf{y}^T)$ based on the Ljung-Box autocorrelation and Jarque-Bera normality tests (see, e.g.,

Lancaster (2004, 95)). Here \mathbf{y}^{pred} denotes the predicted data, Θ the matrix containing all the parameters of the TVP-AR(l) model, and $T(\cdot)$ the test statistic of interest. Small Bayes p values are interpreted in favor of autocorrelation or non-normality. Given the priors defined above, the Gaussian TVP-AR(4,0) model adequately captures the autocorrelation in the inflation series (for the Ljung-Box test statistic with five lags, the Bayes p values range from 16.5% to 30.9% with l less than four, increasing sharply to 59.2% when $l = 4$). The Bayes p values based on the Jarque-Bera test hover around 10%, giving strong evidence against the hypothesis of normally distributed errors. In particular, the residuals are leptokurtic, suggesting that a fat-tailed error distribution, such as a Student's t distribution with a small value of the degree-of-freedom parameter λ might be suitable.²

After specifying the adequate autoregressive order, the next step is determining the correct orders of the causal and noncausal lag polynomials, r and s , respectively. To that end, we compute the sum of the one-step-ahead log predictive likelihoods (PL) evaluated at the observed values over the period from 1960:1 to 2012:2, and select the model producing the greatest value. The one-step-ahead forecasts are computed as recursive pseudo out-of-sample forecasts based on reestimated models at each step, with the estimation sample always starting from 1947:1. Following Bauwens et al. (2011) and Clark and Doh (2011), we then compute the predictive likelihoods using kernel density estimation of the forecasted densities of the TVP-AR(r, s) models. We use the PL because it is straightforward to evaluate unlike the marginal likelihood. PL was also used by Geweke and Amisano (2010) and Chan et al. (2012), among others, and as emphasized by Geweke (1999, 2001) there is a very close connection between it and the log of the marginal likelihood. In particular, when the log predictive likelihoods are evaluated at the observed values over the entire sample period (1947:1-2012:2 here) these two measures are equivalent. Based on the values reported in Table 1, the TVP-AR(0,4) model is selected by a wide margin to all other fourth-order TVP-AR models, with the TVP-AR(1,3) model being the next-best alternative. The posterior mean of the degree-of-freedom parameter λ equals 5.038 (with standard

²The detailed results are not reported, but they are available upon request.

deviation 2.508), lending support to a fat-tailed error distribution.

The evolution of the posterior estimates of the time-varying parameters and trend inflation involved in the selected TVP-AR(0,4) model is depicted in Figure 1. In addition to the posterior median, the 16 and 84 per cent quantiles of the posterior densities are reported. There are substantial changes in the estimated values of the parameters over time. The estimates of φ_{t1} and φ_{t3} increase gradually from the mid-1960's until the mid-1970's and start decreasing again in the early 1980's. The value of φ_{t4} , on the other hand, stays fairly constant at a small value, while there is a gradual shift to a lower level of φ_{t2} between the mid-1960's and the early 1980's. Hence, in total, the dependence in the inflation process seems to have decreased considerably since the early 1980's. The logarithmic scale parameter $\tilde{\sigma}_t$ is very volatile with peaks in the early 1950's, the early 1980's and the late 2000's. In general, our findings are similar to those in the previous literature, including Cogley and Sbordone (2008), Clark and Doh (2011) and Chan et al. (2012). While in their causal (vector) AR models, decreasing dependence, measured by autocorrelation, necessarily implies smaller persistence, that need not be the case in the noncausal model, where the degree of persistence can depend on the initial level of inflation at each time point as well as the size of the shock being considered. However, according to the measure of persistence based on predictability recently suggested by Cogley et al. (2010), the changes in dependence implied by the TVP-AR model indeed seem to translate into changes in persistence.³ The trend inflation \bar{y}_t is accurately estimated, but without further restrictions, there are occasionally very high values and rapid changes, in line with the UC-SV model of Stock and Watson (2007) and the Trend-SV model of Chan et al. (2012).

5.2 New Keynesian Phillips Curve

As pointed out in Section 4, the NKPC can be estimated in the framework of the TVP-AR model by constraining two of the autoregressive parameters constant in time. However, the AR(0,4) model selected above is not suitable for this purpose,

³The results are not reported but they are available upon request.

as the lag polynomial $\phi_t(B)$ must be of order one or higher to facilitate the estimation of the coefficient of lagged inflation in the hybrid NKPC (11). Taken at face value, the purely noncausal TVP-AR model implies that inflation only depends on future expected inflation and the marginal cost, with no role left to past inflation. Therefore, we take the TVP-AR(1,3) model deemed the second best by the PL criterion as the starting point. This entails assuming that past inflation may play a role in the determination of current inflation. If this is not the case, it should be reflected by the estimates of the NKPC. We set the prior means $\underline{\varphi}$ and $\underline{\phi}$ of the time-invariant AR coefficients in (14) at 0.993 and 0.427, respectively. These values imply a relatively forward-looking NKPC, corresponding to prior means of γ_f and γ_b equal to 0.697 and 0.300, respectively. Our chosen value of $\underline{\varphi}$ reflects the assumption that the sum of γ_f and γ_b is close to unity, which, in terms of the deep parameters of the NKPC, corresponds to setting the discount factor in the new Keynesian model underlying the NKPC close but not equal to unity (for which reason we set γ_f at 0.697, slightly below 0.7). In order to give sufficient weight to data information, we specify relatively uninformative priors for the parameters φ and ϕ by setting the prior precision parameters \underline{h}_φ and \underline{h}_ϕ at 1. In other words, these priors are virtually flat in the region of the parameter space considered: $0 \leq \varphi < 1$ and $0 \leq \phi < 1$.⁴ The priors for the time-varying parameters are the same as in estimating the unrestricted TVP-AR model above.

The constancy of φ_t and ϕ_t , or equivalently of γ_f and γ_b , assumed in our approach to estimating the NKPC, is indeed strongly supported by the data, as discussed in detail at the end of Subsection 5.3. The posterior mean of the degree-of-freedom parameter equals 6.952 with standard deviation 3.470, again indicating a fat-tailed error distribution. The corresponding point estimates of the time-invariant coefficients φ and ϕ equal 0.709 and 0.339 with standard deviations 0.093 and 0.111, respectively. Figure 2 shows the evolution of the posterior estimates of the time-varying parameters

⁴The results are checked using $\underline{h}_\varphi = 10^2$. This implies a relatively diffuse and uninformative joint prior for γ_f and γ_b , except that it is informative with respect to the restriction $\gamma_f + \gamma_b \approx 1$. With this prior, the posterior means (standard deviations) of γ_f and γ_b are estimated to be 0.646 (0.056) and 0.237 (0.065), respectively. Hence, the results seem quite robust with respect to the priors.

and trend inflation. As in the case of the unrestricted AR(0,4) model above, the estimates imply declining persistence in the inflation process since the early 1980's. However, in contrast to those results, persistence seems to have been increasing slightly again since the late 1990's. The likeliest interpretation of the diminishing persistence is a decline in the persistence of the variable driving inflation. This can be seen by noticing that in the TVP-NKPC model, the time-varying AR parameters are related to the process of the linear combination of the expectation error and the marginal cost variable x_t (multiplied by its effect on inflation, γ , that may also be time varying, γ_t), presumably dominated by the latter. Following Lanne and Luoto (2013), we estimated the $\gamma_t x_t$ series implied by the TVP-NKPC model (not shown), and it indeed seems to exhibit less serial correlation in the 1980's and 1990's. Thus, taking the new Keynesian model at face value, the diminishing autocorrelation of inflation suggested by the results of the TVP-AR model above (and also reported in the previous literature, see e.g. Stock and Watson (2007)) seems mostly inherited from the driving variable. The evolution of the logarithmic scale parameter $\tilde{\sigma}_t$ is similar to that in the unrestricted model. Under the theory-induced restrictions, the trend inflation \bar{y}_t is very accurately estimated. Moreover, it evolves fairly smoothly, and the spikes implied by the unrestricted model are absent.

The joint posterior density of the NKPC parameters is depicted in Figure 3. The perspective plot shows that the distribution of γ_f and γ_b is unimodal with the mode at approximately $\gamma_f = 0.6$ and $\gamma_b = 0.25$. However, the probability of the region of the parameter space where γ_f is smaller and γ_b is relatively large, is also non-negligible, indicating the relative importance of past inflation. The posterior means of γ_f and γ_b equal 0.579 and 0.269 with standard deviations 0.083 and 0.077, respectively. Hence, the results are qualitatively well aligned with the previous literature in that expected future inflation is found more important in determining current inflation, but past inflation is also important. However, compared to Lanne and Luoto (2013), the estimate of γ_f is smaller in relation to that of γ_b . This suggests that not allowing for structural changes in the marginal cost or its effect on inflation tends to somewhat exaggerate the importance of inflation expectations. Nevertheless, the differences are minor.

5.3 Trend Inflation

To facilitate convenient comparison, the trend inflation estimates implied by the TVP-AR and TVP-NKPC models are reproduced in Figure 4. In the the upper panel also the observed inflation series is included, while, for more detailed inspection, in the middle panel, only the two estimated trend inflation series are depicted. It is clear that while the trend inflation estimate implied by the unrestricted AR(0,4) model tends to follow the observed inflation quite closely at times, the estimate based on the AR(1,3) model with the NKPC restrictions exhibits no such tendency. Moreover, as seen from the bottom panel of Figure 4, the estimates of the 16% and 84% posterior quantiles capture the mean of long-run survey expectations of inflation very well, which indeed should be the case if trend inflation is to be interpreted as the public's perception of the inflation target.⁵ This is in contrast to the recent results of Chan et al. (2012) based on a bounded time-varying-parameter model in which trend inflation is constrained to lie in an interval. At times, their posterior mean estimate tends to be much lower than the mean of survey expectations, hovering between two and three per cent. This discrepancy is particularly pronounced in the early 1980's. To a lesser extent, a similar remark applies to Cogley and Sbordone (2008) albeit they consider inflation measured by the GDP deflator instead of consumer price inflation.

As discussed at the end of Section 4, comparison of the unrestricted TVP-AR and the restricted TVP-NKPC models yields information on the source of time variation in trend inflation. To facilitate such a comparison, we estimated a TVP-AR(1,3) model corresponding to the TVP-NKPC(1,3) model reported above. The sums of log predictive likelihoods of these models equal -423.821 and -414.497 , respectively, lending decisive support (see Jeffreys (1961)) to γ_f and γ_b being time invariant in the NKPC (or equivalently to the restrictions $\varphi_t = \varphi$ and $\phi_t = \phi$ in (14)) and, thus, to variation in trend inflation being inherited from the marginal cost variable instead of being caused by time-varying steady state inflation, as assumed by Cogley

⁵We use the medians of the Blue Chip forecasts and the 10-year-ahead annual average inflation from the Survey of Professional Forecasters for the periods 1979:4-1991:4 and 1992:1-2012, respectively. The source of the data is the Federal Reserve Bank of Philadelphia.

and Sbordone (2008). The fact that their median estimate of trend inflation based on a shifting steady state is flat and at times much lower than the mean of survey expectations, reinforces this conclusion.

5.4 Inflation Forecasts

As discussed in Section 2.1, the estimation of the noncausal AR model yields predictive densities as a by-product. In order to gauge the forecast performance, we compute pseudo out-of-sample forecasts from a number of models for the period 1960:1 to 2012:2. The forecasts are computed recursively, at each step reestimating the models using an expanding data window starting at 1947:1. However, model selection is not performed recursively, but the models are kept fixed throughout. We consider the forecast horizons of one, four, and eight quarters, as is common in the inflation forecasting literature. Two evaluation criteria, the root mean squared forecast error (RMSFE) and the sum of log predictive likelihoods (PL) computed over the forecast period are reported. While the RMSFE summarizes the accuracy of point forecasts, the PL yields information on the forecasting performance of the entire predictive density. Geweke (1999, 2001), Geweke and Amisano (2010), and Chan et al. (2012), among others have also used the latter metric to evaluate the accuracy of density forecasts. We calculate the h -step-ahead PLs evaluated at the observed values from draws of forecasts, using kernel density estimation in the same way as described in Section 5.1.

The RMSFEs of the TVP-AR(0,4) model are reported in Table 2 along with the relative MSFE's of a number of alternative models, computed by dividing the respective MSFE by that of the benchmark TVP-AR(0,4) model. Thus, values greater than unity indicate superiority of the benchmark. Because of some previous results suggesting that inflation has become easier to forecast since the mid-1980's (see, e.g., Lanne and Luoto (2012) and the references therein), we also report results for two subsample periods, 1960:1–1983:4, and 1984:1–2012:2. The RMSFEs of the TVP-AR(0,4) model are indeed considerably lower in the latter subsample period, reconfirming those previous findings based on constant-parameter models. However, in all fore-

casting periods considered, the TVP-AR(0,4) model is superior to the corresponding constant-parameter noncausal and time-varying causal AR models (with the single exception of the two-year forecast of the AR(0,4) model in the latter subsample period). It also always beats the random walk forecast. In most cases, these differences in forecast accuracy are significant according to the test of Diebold and Mariano (1995) and West (1996). As far as the varying-parameter noncausal model incorporating the NKPC restrictions (TVP-NKPC) is concerned, the differences compared to the unrestricted model are minor, suggesting that for point forecasts, the restrictions are inconsequential.

In Table 3, we report the sums of the log predictive likelihoods of four different fourth-order AR models. Also by this criterion, the TVP-AR model is clearly superior to its constant-parameter counterpart, by even a wider margin than in terms of point forecasts. Among the TVP models, the unrestricted TVP-AR(0,4) model yields the greatest value at each forecast horizon, indicating its superiority over both the corresponding causal model as well as the noncausal model incorporating the NKPC restrictions. Somewhat surprisingly in view of the relative MSFEs in Table 2, in two out of three cases ($h = 4, 8$), the causal model slightly beats the TVP-NKPC model. Hence, for density forecasting, the NKPC restrictions seem to have an adverse effect despite not being important for point forecasts.

6 Conclusion

In this paper, we extend the noncausal AR model of Lanne and Saikkonen (2011) by allowing for time-varying parameters and apply the extended model to postwar U.S. consumer price inflation. Because of its complexity, the model is estimated by Bayesian methods, whose description may also be useful in potential future applications to other economic time series.

The new model turns out to be useful in modeling the U.S. inflation process. In accordance with the previous literature, there is strong evidence in favor of non-causality. As a matter of fact, our results based on the unrestricted time series model suggest that the observed autocorrelation in inflation follows from future inflation

expectations rather than dependence on past inflation. Moreover, inflation persistence seems to have varied over time, exhibiting a clear decline since the early 1980's and a slight increase again in the late 1990's. The proposed model also beats its constant-parameter and causal counterparts in forecast accuracy.

When imposing more structure in the form of the restrictions implied by the new Keynesian Phillips curve, we find that also past inflation has a role to play in the determination of U.S. inflation in addition to expectations. Specifically, our estimates of the new Keynesian Phillips curve reconfirm the previous findings that the determination of inflation is dominated by inflation expectations, while past inflation plays a minor but still important role. However, comparison to results based on the constant-parameter noncausal AR model reveals that not allowing for time variation slightly undermines the importance of past inflation. The trend inflation implied by the estimated NKPC evolves smoothly, unlike most estimates presented in the previous literature, and it is well aligned with survey expectations, as should indeed be the case if trend inflation is to be interpreted as the public's perception of the inflation target. There is also evidence in favor of the time variation in trend inflation following from the underlying marginal cost variable driving inflation rather than time-varying NKPC parameters.

Our analysis of the NKPC is based on the assumption of constant steady state inflation, and shifts in trend inflation are thought to be inherited from the marginal cost variable, which is also supported by the data. This is consistent with the idea that the public's perception of the inflation target is influenced by structural breaks in the marginal cost variable. However, the model put forth in this paper does not facilitate a detailed exploration of the origins of shifts in trend inflation. Therefore, in future work, it would be interesting to extend the noncausal TVP-AR model in this direction. One viable alternative would be the development of an algorithm for the joint estimation of the noncausal AR coefficients and trend inflation.

Appendix A: MCMC Algorithm

In this appendix, we describe the MCMC algorithm used for the estimation of the parameters of (1) and (14). The sampler involves sequential drawings from the full conditional posteriors given below. Our experience is that convergence occurs rapidly when the zeros of the polynomials $\phi_t(z) = 0$ and $\varphi_t(B^{-1}) = 0$ are not too close to the unit circle. In this case, further tuning of the sampler is not needed, and according to the results of Koop and Potter (2011), the sampler also provides an accurate approximation for the restricted posteriors of θ^T and φ^T .

Sampling AR parameters θ^T and φ^T . Let us first consider the conditional posterior densities of the AR parameters. Following Cogley and Sargent (2005), we impose stationarity restrictions by discarding the draws from the unrestricted posterior of θ^T (or φ^T) that do not lie in the stationary region. First, take φ^T as given to express (1) in terms of $u_t = \varphi_t(B^{-1})y_t$. Then, given $\sigma^T = (\sigma_1, \dots, \sigma_T)'$ and $\tilde{\mathbf{h}}^T = (\tilde{h}_1, \dots, \tilde{h}_T)'$, the observables u_t for $t = 1, \dots, T$, are normally distributed:

$$\left\{ u_t \mid \mathbf{U}_t, \boldsymbol{\theta}_t, \mathbf{y}^+, \boldsymbol{\varphi}^T, \boldsymbol{\sigma}^T, \tilde{\mathbf{h}}^T \right\} \sim N \left(\mathbf{U}_t \boldsymbol{\theta}_t, \sigma_t^2 \tilde{h}_t^{-1} \right), \quad (15)$$

where $\mathbf{U}_t = (1, \varphi_t(B^{-1})y_{t-1}, \dots, \varphi_t(B^{-1})y_{t-r})$ and $\mathbf{y}^+ = (y_{T+1}, \dots, y_{T+s})'$. From (3) it can be seen that the errors ϵ_t are independent of u_{t-1}, \dots, u_{t-r} . This, however, does not necessarily mean that they are independent of $\varphi_t(B^{-1})y_{t-1}, \dots, \varphi_t(B^{-1})y_{t-r}$. Thus, unfortunately, it cannot be guaranteed that the elements of \mathbf{U}_t in (15) are independent of ϵ_t . However, as long as the vector φ_t evolves smoothly, as is assumed (cf., (10)), the potential problems should be minor. If this is the case, locally we are close to the time-invariant case in which the errors ϵ_t are independent of $\varphi(B^{-1})y_{t-1} = u_{t-1}, \dots, \varphi(B^{-1})y_{t-r} = u_{t-r}$.

Based on (9), the parameter vector θ_t evolves according to

$$\left\{ \boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{Q}_1 \right\} \sim N \left(\boldsymbol{\theta}_{t-1}, \mathbf{Q}_1 \right). \quad (16)$$

Equations (15) and (16) define a conditionally linear and Gaussian state-space model. Thus, a standard algorithm such as that in Carter and Kohn (1994) or Durbin and Koopman (2002), can be readily used to draw $\boldsymbol{\theta}^T$. Furthermore, in the purely causal

case, \mathbf{U}_t in (15) reduces to $(1, y_{t-1}, \dots, y_{t-r})$ and θ^T can be drawn in the same way as explained above (recall also that $\theta_t = (c_t, \phi_t')'$). The prior distribution of the initial state θ_0 is assumed to be multivariate normal $N(\theta_{0|0}, \mathbf{P}_{0|0})$ (see, e.g., Durbin and Koopman (2002)).

Turning to φ^T , we take θ^T as given to express (1) in terms of $v_t = \phi_t(B)y_t$. Thus, the observables $\bar{v}_t = v_t - c_t$ for $t = 1, \dots, T$ are conditionally Gaussian:

$$\{\bar{v}_t \mid \mathbf{V}_t, \varphi_t, \mathbf{y}^+, \theta^T, \sigma^T, \tilde{\mathbf{h}}^T\} \sim N(\mathbf{V}_t \varphi_t, \sigma_t^2 \tilde{h}_t^{-1}), \quad (17)$$

where $\mathbf{V}_t = (\phi_t(B)y_{t+1}, \dots, \phi_t(B)y_{t+s})$. Again, it cannot be guaranteed that the elements of \mathbf{V}_t are independent of ϵ_t , even though v_{t+1}, \dots, v_{t+s} clearly are (cf., (4)). In practice, however, as long as the vector θ_t evolves smoothly (cf., 9), we expect the estimates of φ^T to be unbiased.

From (10) it can be seen that the evolution of the states φ_t for $t = 1, \dots, T$, is defined by

$$\{\varphi_t \mid \varphi_{t-1}, \mathbf{Q}_2\} \sim N(\varphi_{t-1}, \mathbf{Q}_2). \quad (18)$$

Again, exactly like in the case of θ^T , (17) and (18) define a conditionally linear and Gaussian state-space model, and the method of Durbin and Koopman (2002) is used to simulate random matrices from the full conditional posterior distribution of φ^T . In the purely noncausal TVP-AR(0, s) case, the method can be used by replacing φ_t with $(c_t, \varphi_t')'$, and using $\bar{v}_t = v_t$ and $\mathbf{V}_t = (1, v_{t+1}, \dots, v_{t+s})$ in (17). Obviously, (18) has to be modified accordingly. Finally, the prior distribution of the initial state φ_0 is assumed to be multivariate normal $N(\varphi_{0|0}, \mathbf{R}_{0|0})$.

Sampling stochastic volatilities σ^T . The stochastic volatilities are drawn using the method proposed by Kim et al. (1998). By squaring and taking logs of (7) we obtain

$$\epsilon_t^* = 2\tilde{\sigma}_t + \eta_t^*, \quad (19)$$

where $\epsilon_t^* = \log(\epsilon_t^2 \tilde{h}_t + 0.001)$ and the innovations $\eta_t^* = \log \eta_t^2$ are distributed as $\log \chi^2(1)$. The offset constant (0.001) is used to ensure the robustness of the estimation procedure. Conditional on θ^T and φ^T , the errors ϵ_t in (1) are observable. Therefore, (8) and (19) form a linear but non-Gaussian (because the innovations η_t^*

are distributed as $\log \chi^2(1)$ state-space model. To be able to use the standard state space methods, we transform the model into a Gaussian one. In particular, we follow Kim et al. (1998) and approximate the $\log \chi^2(1)$ distribution by a mixture of seven normal distributions, i.e., $\{\eta_t^* | s_t = j\} \sim N(m_j - 1.2704, \nu_j^2)$, $\Pr(s_t = j) = q_j$, $j = 1, \dots, 7$, where the parameters q_j , m_j , and ν_j are chosen to optimize the approximation, and are given in Kim et al. (1998). Notice that these parameter values are independent of the application.

Given the above approximation, we generate draws of $\boldsymbol{\sigma}^T$ using the simulation smoothing algorithm of Durbin and Koopman (2002), by setting (19) and (8) as the measurement and transition equations, respectively. The indicator variables $\mathbf{s}^T = (s_1, \dots, s_T)$, needed to calculate the mixture of normals, are drawn independently from the discrete distribution

$$\Pr(s_t = j | \mathbf{y}^T, \mathbf{y}^+, \boldsymbol{\theta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\sigma}^T, \tilde{\mathbf{h}}^T) \propto q_j \nu_j^{-1} \exp \left[-\frac{1}{2\nu_j^2} (\epsilon_t^* - 2\tilde{\sigma}_t - m_j + 1.2704)^2 \right].$$

The prior distribution for initial state is assumed normal $N(\tilde{\sigma}_{0|0}, \omega_{0|0}^2)$.

Sampling Precision Parameters $\tilde{\mathbf{h}}^T$. Following Geweke (1993), we obtain the kernel of the conditional density of $\tilde{\mathbf{h}}^T$, $p(\tilde{\mathbf{h}}^T | \mathbf{y}^T, \mathbf{y}^+, \boldsymbol{\delta}^T, \boldsymbol{\sigma}^T, \lambda)$, by presenting (1) in terms of (7) and multiplying the resulting density by the prior density $p(\tilde{\mathbf{h}}^T | \lambda)$ introduced in the in the Section 2.2. The kernel has the following expression:

$$p(\tilde{h}_t | \mathbf{y}^T, \mathbf{y}^+, \boldsymbol{\theta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\sigma}^T, \lambda) \propto \tilde{h}_t^{(\lambda-1)/2} \exp \left[-\frac{1}{2} [\lambda + \sigma_t^{-2} \epsilon_t^2] \tilde{h}_t \right],$$

where, conditional on $\boldsymbol{\theta}^T$ and $\boldsymbol{\varphi}^T$, ϵ_t is observable, and the precision parameters $\tilde{h}_1, \dots, \tilde{h}_T$ are conditionally independent. Thus, by the properties of the chi-squared density, $\tilde{h}_1, \dots, \tilde{h}_T$ can be sampled directly using

$$\left\{ [\lambda + \sigma_t^{-2} \epsilon_t^2] \tilde{h}_t | \mathbf{y}^T, \mathbf{y}^+, \boldsymbol{\theta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\sigma}^T, \lambda \right\} \sim \chi^2(\lambda + 1). \quad (20)$$

Sampling Post Sample Observations y_{T+1}, \dots, y_{T+s} . This step is based on the approximation $\phi_{T+h}(B) y_{T+h} \approx \sum_{j=0}^{M-h} \beta_{T+h,j} w_{T+h+j}$, discussed in (6), which is used recursively to map from w_{T+1}, \dots, w_{T+M} into y_{T+1}, \dots, y_{T+s} . In what follows, we

use these two parameterizations interchangeably and assume a Student-t hierarchical prior density for w_{T+j} ($j = 1, \dots, M$): $t(w_{T+j} | c_T, \sigma_T; \lambda)$, where c_T is a location parameter, σ_T^2 is a scale parameter, and λ is a degrees of freedom parameter. Then, based on equations (1) and (7), the conditional posterior density of the post-sample observations $\mathbf{y}^+ \equiv (y_{T+1}, \dots, y_{T+s})$ can be expressed as

$$p(\mathbf{y}^+ | \mathbf{y}^T, \boldsymbol{\theta}^T, \boldsymbol{\varphi}^T, \boldsymbol{\sigma}^T, \lambda) \propto \prod_{t=T-s+1}^T N\left(\varphi_t(B^{-1}) \phi_t(B) y_t \mid c_t, \sigma_t^2 \tilde{h}_t^{-1}\right) \times \prod_{t=T+1}^{T+M} t(w_t | c_T, \sigma_T; \lambda). \quad (21)$$

Since this distribution is non-standard, we follow Chib and Greenberg (1994), and augment the algorithm with a Metropolis-Hastings (MH) step. In other words, the prior density $\prod_{t=T+1}^{T+M} t(w_t | c_T, \sigma_T; \lambda)$ is used to draw proposals for y_{T+1}, \dots, y_{T+s} , which are then accepted with probability

$$\min \left\{ 1, \frac{\prod_{t=T-s+1}^T N\left(\varphi_t(B^{-1}) \phi_t(B) y_t^* \mid c_t, \sigma_t^2 \tilde{h}_t^{-1}\right)}{\prod_{t=T-s+1}^T N\left(\varphi_t(B^{-1}) \phi_t(B) y_t^{j-1} \mid c_t, \sigma_t^2 \tilde{h}_t^{-1}\right)} \right\},$$

where $y_{T+1}^*, \dots, y_{T+s}^*$ denote a candidate draw and $y_{T+1}^{j-1}, \dots, y_{T+s}^{j-1}$ the current draw ($j = 1, \dots, N$, where N is the number of simulation rounds). As discussed above, $\phi_{T+1}, \dots, \phi_{T+s}$ and $\varphi_{T+1}, \dots, \varphi_{T+s}$, needed to map w_{T+1}, \dots, w_{T+M} into y_{T+1}, \dots, y_{T+s} , are simulated from the random walk models ((9) and (10)), respectively.

Sampling Prior Hyperparameters \mathbf{Q}_i , ω , and λ . Because of the hierarchical prior structure in which \mathbf{Q}_i ($i = 1, 2$) affects the data only through $\boldsymbol{\theta}^T$ or $\boldsymbol{\varphi}^T$, the conditional posterior density of \mathbf{Q}_i has the expression

$$f(\mathbf{Q}_i | \mathbf{y}^T, \boldsymbol{\theta}^T, \boldsymbol{\varphi}^T) \propto iW(\mathbf{Q}_i | \bar{\nu}_i, \bar{\mathbf{Q}}_i), \quad (22)$$

where $\bar{\nu}_i = T + \underline{\nu}_i$, and

$$\begin{aligned} \bar{\mathbf{Q}}_1 &= \underline{\mathbf{Q}}_1 + \sum_{t=1}^T (\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}) (\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1})', \\ \bar{\mathbf{Q}}_2 &= \underline{\mathbf{Q}}_2 + \sum_{t=1}^T (\boldsymbol{\varphi}_t - \boldsymbol{\varphi}_{t-1}) (\boldsymbol{\varphi}_t - \boldsymbol{\varphi}_{t-1})', \end{aligned}$$

(see Cogley and Sargent (2005)). Thus, \mathbf{Q}_i can be drawn from the inverse-Wishart distribution: $\mathbf{Q}_i \sim iW(\bar{\nu}_i, \bar{\mathbf{Q}}_i)$.

Based on the assumptions concerning the error term of (8), ω^2 can be sampled directly from the inverse-gamma distribution:

$$\omega^2 \mid \mathbf{y}^T, \boldsymbol{\sigma}^T \sim iGamma(\bar{a}, \bar{b}), \quad (23)$$

where

$$\bar{a} = T/2 + \underline{a}, \text{ and } \bar{b} = \frac{1}{2} \sum_{t=1}^T (\tilde{\sigma}_t - \tilde{\sigma}_{t-1})^2 + \underline{b}.$$

Again, exactly like in the case of \mathbf{Q}_i (or ω^2), because of the hierarchical prior structure in which λ affects the data only through $\tilde{\mathbf{h}}^T$, the conditional posterior density of λ has the following expression:

$$\begin{aligned} p\left(\lambda \mid \mathbf{y}^T, \tilde{\mathbf{h}}^T, \mathbf{y}^+\right) &\propto [2^{\lambda/2} \Gamma(\lambda/2)]^{-T} \lambda^{\lambda T/2} \left(\prod_{t=1}^T \tilde{h}_t^{(\lambda-2)/2} \right) \\ &\exp \left[- \left(\frac{1}{\lambda} + \frac{1}{2} \sum_{t=1}^T \tilde{h}_t \right) \lambda \right] \\ &\prod_{t=T+1}^{T+M} f(w_t \mid c_T, \sigma_T; \lambda), \end{aligned} \quad (24)$$

(see Geweke (2005) and Lanne, Luoma, and Luoto (2012)). As a candidate distribution for λ we use a univariate normal distribution, with mean equal to the mode of (24) and precision parameter equal to the negative of the second derivative of the log posterior density, evaluated at the mode. The acceptance probability is calculated using (24) (see Geweke, 2005).

Appendix B: Sampling parameters ϕ and φ for the NKPC

In this appendix, we describe the estimation of the parameters of (14), which is done using the algorithm of Lanne, Luoma, and Luoto (2012) for $y_t^* = a_t(B) b_t(B^{-1}) y_t$. In particular, we let $u_t^* = (1 - \varphi B^{-1}) y_t^* = a_t(B) b_t(B^{-1}) u_t$ and $v_t^* = (1 - \phi B) y_t^* = a_t(B) b_t(B^{-1}) v_t$, and express (14) as $(1 - \phi B) u_t^* = (1 - \varphi B^{-1}) v_t^* = c_t + \epsilon_t$. This suggests that, assuming the truncated normal prior densities $TN(\underline{\phi}, \underline{h}_\phi^{-1}; 0 \leq \phi < 1)$ and $TN(\underline{\varphi}, \underline{h}_\varphi^{-1}; 0 \leq \varphi < 1)$ for ϕ and φ , respectively, their conditional posterior densities can be successfully approximated (see Lanne, Luoma, and Luoto (2012)) by

$$\left\{ \phi \mid \mathbf{y}^T, \varphi, \mathbf{a}^T, \mathbf{b}^T, \mathbf{y}^+, \boldsymbol{\sigma}^T, \tilde{\mathbf{h}}^T \right\} \sim TN\left(\bar{\phi}, \bar{h}_\phi^{-1}; 0 \leq \phi < 1\right),$$

and

$$\left\{ \varphi \mid \mathbf{y}^T, \phi, \mathbf{a}^T, \mathbf{b}^T, \mathbf{y}^+, \boldsymbol{\sigma}^T, \tilde{\mathbf{h}}^T \right\} \sim TN \left(\bar{\varphi}, \bar{h}_\varphi^{-1}; 0 \leq \varphi < 1 \right),$$

where

$$\begin{aligned} \bar{\phi} &= \bar{h}_\phi^{-1} \left[\underline{h}_\phi \phi + \sum_{t=1}^T \sigma_t^{-2} \tilde{h}_t [a_t(B) b_t(B^{-1}) u_{t-1}] (u_t^* - c_t) \right], \\ \bar{\varphi} &= \bar{h}_\varphi^{-1} \left[\underline{h}_\varphi \varphi + \sum_{t=1}^T \sigma_t^{-2} \tilde{h}_t [a_t(B) b_t(B^{-1}) v_{t+1}] (v_t^* - c_t) \right], \\ \bar{h}_\phi &= \left[\underline{h}_\phi + \sum_{t=1}^T \sigma_t^{-2} \tilde{h}_t [a_t(B) b_t(B^{-1}) u_{t-1}]^2 \right], \end{aligned}$$

and

$$\bar{h}_\varphi = \left[\underline{h}_\varphi + \sum_{t=1}^T \sigma_t^{-2} \tilde{h}_t [a_t(B) b_t(B^{-1}) v_{t+1}]^2 \right].$$

As far as the time-varying parameters $c_t, a_{t,1}, \dots, a_{t,r-1}$ and $b_{t,1}, \dots, b_{t,s-1}$ are concerned, they can be simulated using the above algorithm for the quantity $y_t^{**} = (1 - \phi B)(1 - \varphi B^{-1}) y_t$. Using $a_{t,1}, \dots, a_{t,r-1}, b_{t,1}, \dots, b_{t,s-1}, \phi$, and φ , we can then calculate the products $\phi_t(B) = a_t(B)(1 - \phi B)$ and $\varphi_t(B^{-1}) = b_t(B^{-1})(1 - \varphi B^{-1})$ to express (14) in terms of (1): $\phi_t(B) \varphi_t(B^{-1}) y_t = c_t + \epsilon_t$. Thus, the predictive density of y_{T+1}, \dots, y_{T+h} ($h \geq s$) can be evaluated using the algorithm described in the Appendix A.

Appendix C: Efficiency of MCMC Algorithm

Following Primiceri (2005) and Chan et al. (2012), we use the inefficiency factors (IF) to monitor the efficiency of the MCMC algorithm. The IF is defined as $1 + 2 \sum_{k=1}^K \rho_k$, where ρ_k is the k th sample autocorrelation of the chain, and K is set in such a way that the autocorrelation tapers off (we follow Primiceri (2005) and use 4 percent tapered windows). Almost independent draws from the posterior suggest that an algorithm is efficient. In this case the IF are around 1. Because of the very high dimension of the TVP models, the IF are typically reported to be much higher (see, e.g., Primiceri (2005), Chan and Strachan (2012), and Chan et al. (2012)), the averages of IF ranging from 2 to more than 100. We follow Chan et al. (2012) and report the 50th, 25th, and 75th percentiles of the IF for the parameters of the TVP-AR(0,4) and TVP-AR(1,3) models discussed in the text. In particular, the summary statistics are

provided for three sets of parameters, consisting of the intercepts and AR coefficients, the stochastic volatilities, and all the time-invariant parameters.

In the case of the TVP-AR(0,4) model, the median (and the 25th and 75th percentiles) of the IF is 21.6 (8.9, 41.1) for the intercepts and AR coefficients, 22.7 (18.6, 28.7) for the stochastic volatilities, and 81.6 (62.9, 117.2) for the time-invariant parameters. The corresponding statistics for the parameters of the TVP-AR(1,3) model incorporating the NKPC restrictions are 10.16 (5.4, 40.8), 6.7 (4.4, 9.9), and 45.4 (30.8, 66.2), respectively. Thus, the NKPC restrictions seem to improve the performance of the sampler.

References

- Bauwens L., G. Koop, D. Korobilis, and J.V.K. Rombouts (2011). A Comparison of Forecasting Procedures for Macroeconomic Series: the Contribution of Structural Break Models. Cahiers de recherche 1104, CIRPEE.
- Beveridge, S., and C.R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics* 7, 151–174.
- Carter, C. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541-553.
- Chan, J. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation* 1, 101–120.
- Chan, J., G. Koop, and S. M. Potter (2012). A New Model of Trend Inflation. *Journal of Business and Economic Statistics* (forthcoming).
- Chan, J. and R. Strachan (2012). “Estimation in non-linear non-Gaussian state-space models with precision-based methods,” Technical report, Research School of Economics, Australian National University, Canberra.

- Chib, S. and S. Ramamurthy (2011). DSGE Models with Student-t Errors. *Econometric Reviews* (forthcoming).
- Clark, T.E. and T. Doh (2011). A Bayesian evaluation of alternative models of trend inflation. Working Paper 1134, Federal Reserve Bank of Cleveland.
- Cogley, T., G.E. Primiceri, and T.J. Sargent (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics* 2, 43–69.
- Cogley, T. and T.J. Sargent, (2001). Evolving post World War II US inflation dynamics. *NBER Macroeconomics Annual* 16, 331–373.
- Cogley, T. and T. Sargent (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.. *Review of Economic Dynamics* 8, 262–302.
- Cogley, T. and A. Sbordone (2008). Trend inflation, indexation, and inflation persistence in the new Keynesian Phillips curve. *American Economic Review* 98, 2101–2126.
- Cúrdia, V., M. Del Negro, and D. L. Greenwald (2012). Rare Shocks, Great Recessions. *Federal Reserve Bank of New York Staff Reports*, no. 585.
- Diebold, F.X., and R.S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.
- Durbin, J. and S. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89, 603–616.
- Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007). Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies* 74, 1059–1087.
- Galí, J., and M. Gertler (1999). Inflation dynamics: A structural econometric approach. *Journal of Monetary Economics* 44, 195–222.

- Geweke, J. (1999). Using Simulation Methods for Bayesian Econometrics Models: Inference, Development and Communication. *Econometric Reviews* 18(1), 1–73.
- Geweke, J. (2001). Bayesian econometrics and forecasting. *Journal of Econometrics* 100: 11-15.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Wiley, Hoboken, New Jersey.
- Geweke, J. and G. Amisano (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26(2), 216–230.
- Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford Classic Texts in the Physical Sciences. Oxford University Press, Oxford.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies* 65, 361–93.
- Koop, G. and S. Potter (2011). Time varying VARs with inequality restrictions. *Journal of Economic Dynamics and Control* 35, 1126–1138.
- Kozicki, S., and P.A. Tinsley (1998). Moving endpoints and the internal consistency of agents' ex ante forecasts. *Computational Economics* 11, 21–40.
- Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics*. Blackwell, Oxford, UK.
- Lanne, M., A. Luoma, and J. Luoto (2012). Bayesian model selection and forecasting in noncausal autoregressive models. *Journal of Applied Econometrics* 27, 812–830.
- Lanne, M., and J. Luoto (2012). Has US inflation really become harder to forecast? *Economics Letters* 115, 383–386.

- Lanne, M., and J. Luoto (2013). Autoregression-based estimation of the new Keynesian Phillips curve. *Journal of Economic Dynamics and Control* 37, 561–570.
- Lanne, M., J. Luoto, and P. Saikkonen (2012). Optimal forecasting of nocausal autoregressive time series. *International Journal of Forecasting* 28, 623–631.
- Lanne, M., and P. Saikkonen (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics* 3 (3), Article 2.
- Lanne, M., and P. Saikkonen (2013). Noncausal vector autoregression. *Econometric Theory* (forthcoming).
- Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* 72, 821–852.
- Sahuc, J-G (2006). Partial Indexation, Trend Inflation, and the Hybrid Phillips Curve. *Economics Letters* 90. 42–50.
- Stock, J., and M. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking* 39, 3–33.
- West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

Table 1: The sums of log predictive likelihoods (PL) of the fourth-order TVP-AR models.

r	4	3	2	1	0
s	0	1	2	3	4
PL	-444.037	-436.177	-432.704	-423.821	-396.208

The entries are the sums over 1960:1 to 2012:2 of one-step ahead predictive likelihoods of the different AR-TVP(r, s) models. They are estimated from the simulated predictive densities using a kernel smoother.

Table 2: Pseudo out-of-sample forecasting results of quarterly inflation.

	1960:1–2012:2				1960:1–1983:4				1984:1–2012:2			
	$h = 1$	$h = 4$	$h = 8$	$h = 1$	$h = 4$	$h = 8$	$h = 1$	$h = 4$	$h = 8$	$h = 1$	$h = 4$	$h = 8$
TVP-AR(0,4) RMSFE	1.92	2.35	2.91	1.77	2.65	3.64	2.05	2.07	2.18			
Relative MSFEs												
TVP-NKPC	1.09	1.15*	0.98	1.11	1.16	0.94	1.08	1.14	1.08			
TVP-AR(4,0)	1.37***	1.15**	1.09	1.61***	1.22*	1.12	1.23	1.05**	1.01			
AR(0,4)	1.13	1.67***	1.51	1.29	1.92***	1.78*	1.03	1.33***	0.91			
Random Walk	1.24**	1.45***	1.38*	1.09	1.09	1.30**	1.33**	1.92***	1.56			

The first row reports the RMSFEs of the unrestricted TVP-AR(0,4) model for the one-quarter ($h = 1$), annual ($h = 4$) and two-year ($h = 8$) forecast horizons. The point forecast is the median of the posterior predictive distribution. The rest of the entries are relative MSFEs of alternative models. *, **, and *** denote significance of the Diebold-Mariano test statistic comparing each model to the TVP-AR(0,4) model at the 10%, 5%, and 1% significance levels, respectively.

Table 3: The sums of log predictive likelihoods of fourth-order AR models.

Model	$h = 1$	$h = 4$	$h = 8$
TVP-AR(0,4)	-396.21	-456.51	-479.71
TVP-NKPC	-414.52	-468.68	-494.50
TVP-AR(4,0)	-444.04	-461.15	-489.05
AR(0,4)	-445.26	-498.58	-520.73

The entries are the sums over 1960:1 to 2012:2 of predictive likelihoods a quarter ($h = 1$), a year ($h = 4$), and two years ($h = 8$) ahead. They are estimated from the simulated predictive densities using a kernel smoother.

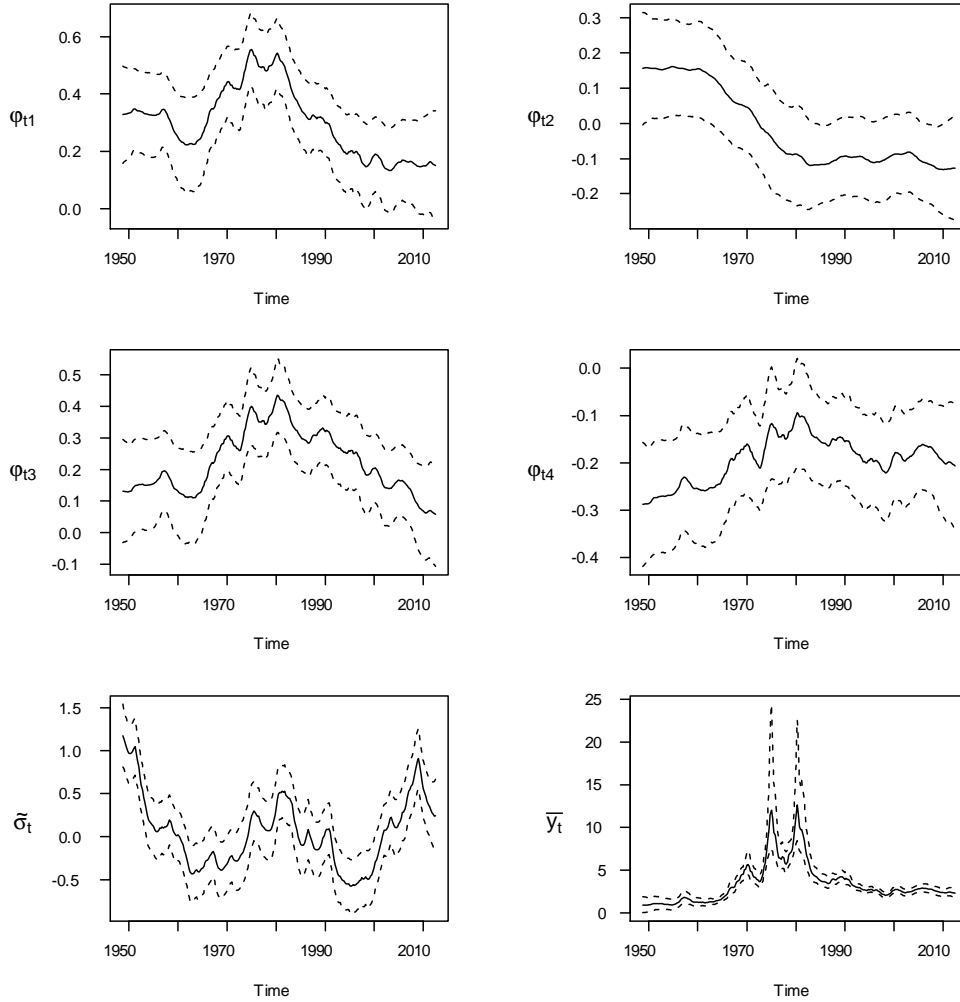


Figure 1: The evolution of the autoregressive (φ_{t1} , φ_{t2} , φ_{t3} , φ_{t4}) and scale ($\tilde{\sigma}_t$) parameters of the TVP-AR(0,4) model, and the implied trend inflation (\bar{y}_t). In each panel, the solid line is the posterior median, and the dashed lines depict the 16% and 84% quantiles of the posterior distribution.

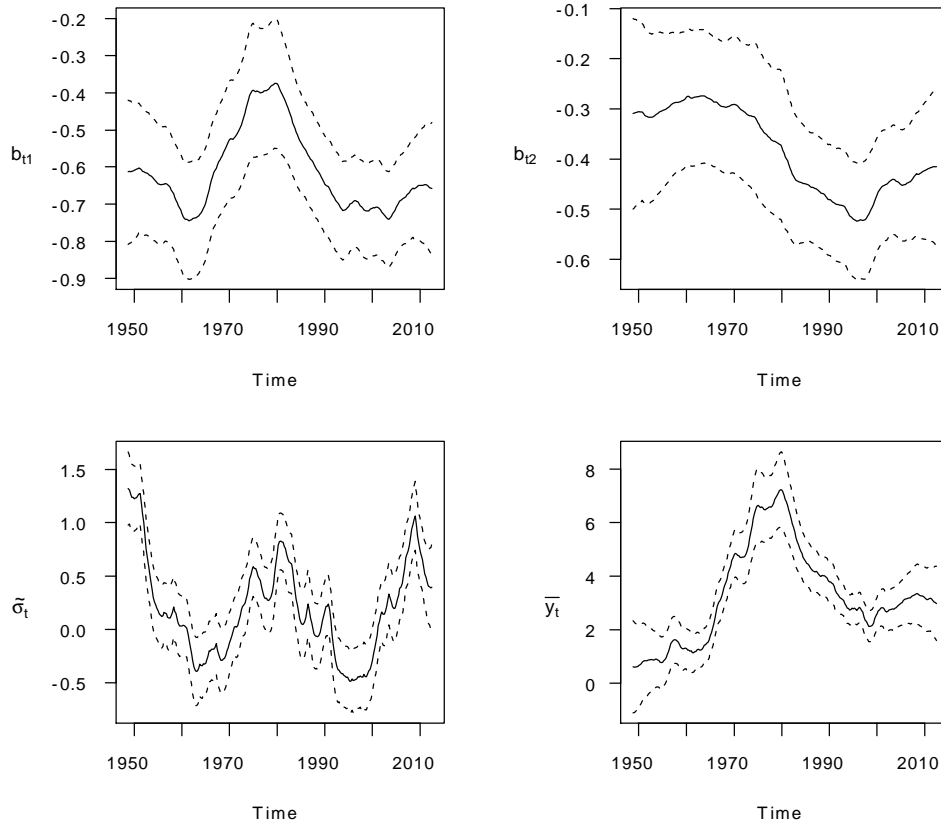


Figure 2: The evolution of the time-varying autoregressive (b_{t1} , b_{t2}) and scale ($\tilde{\sigma}_t$) parameters of the TVP-AR(1,3) model, and the implied trend inflation (\bar{y}_t). In each panel, the solid line is the posterior median, and the dashed lines depict the 16% and 84% quantiles of the posterior distribution.

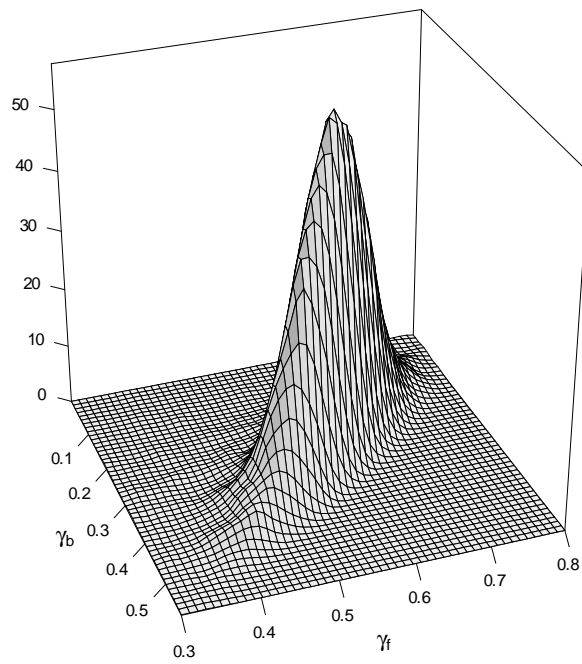


Figure 3: The joint posterior density of the parameters γ_f and γ_b of the new Keynesian Phillips curve (11) based on the TVP-NKPC(1,3) model.

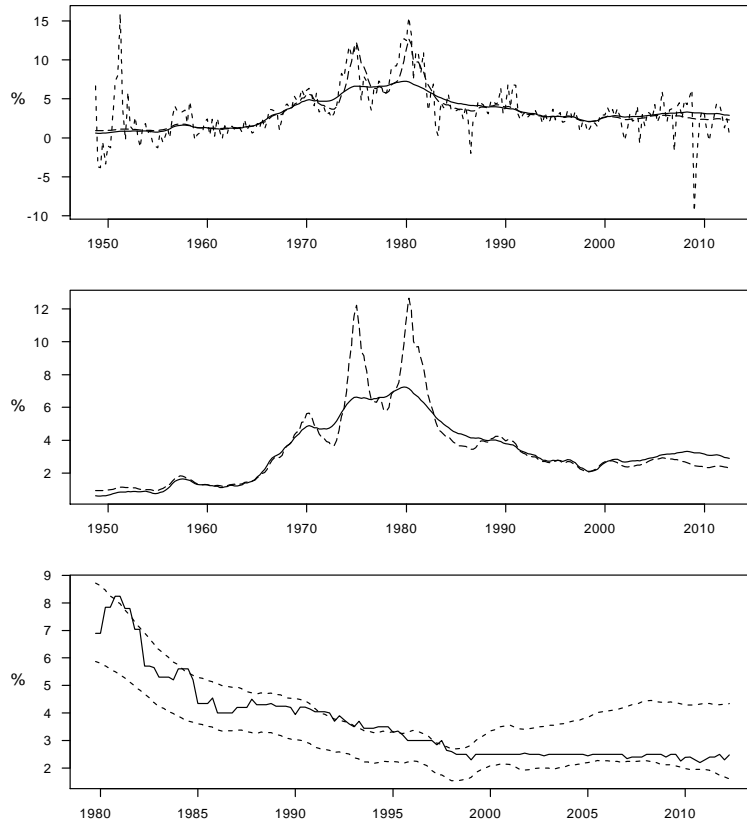


Figure 4: Trend inflation estimates. The upper and middle panels depict the trend inflation implied by the TVP-AR(0,4) (long dashes) and TVP-NKPC(1,3) (solid line) models, with observed inflation (dashes) included in the upper panel. The bottom panel shows the evolution of survey inflation expectations (solid line) and the 16% and 84% quantiles of the trend inflation implied by the TVP-NKPC(1,3) model (dashes).