Efficient estimation of time-varying parameter models with stochastic volatility

Turatti, Douglas Eduardo

DOI (link to publication from Publisher):
10.2139/ssrn.3236806

Publication date:
2021

Efficient estimation of time-varying parameter models with stochastic volatility

Douglas Eduardo Turatti *
Aalborg University

Abstract

This paper develops an efficient estimation procedure for time-varying parameter autoregressive models with stochastic volatility. Necessary restrictions are imposed on the time-varying autoregressive parameters, thus stability conditions are satisfied. We show that a conditional Gaussian likelihood representation is available with marginalization of linear latent states, thus only non-linear states need to be simulated. The sampling is based on a multivariate extension of the Numerically Accelerated Importance Sampling together with a Rao-Blackwellization step to construct a highly efficient maximum likelihood estimator. A simulation study highlights the precision of the procedure in the joint estimation of parameters and latent states. The models are applied to the analysis of inflation dynamics. Estimates of the time-varying parameters indicate the importance of the random innovations in explaining the inflation process, while the trend component is more stable than previously found in the literature. An out-of-sample forecasting exercise showed superior results with respect to several benchmark models, especially for long-term forecasting.

Keywords: Time-varying parameter models; stochastic volatility; time-varying parameter autoregressive models; Numerically Accelerated Importance Sampling; forecasting inflation.

1 Introduction

Parameter instability is known to be an important feature in economic time series. For instance, Stock and Watson (1996) have found substantial parameter instability in a large set of macroeconomic time series. In order to cope with this changing environment, a popular approach in the literature has been to allow the parameter vector to be time-varying. Furthermore, works from Primiceri (2005), and Stock and Watson (2007) highlight the importance of jointly considering heteroscedascity effects in the random innovations and time variation in the model parameters. As emphasized by Cogley and Sargent (2002) estimating time-varying parameter (TVP) models ignoring possible changes in the volatility is likely to generate fictitious dynamics in the coefficients. Thus, much research focuses on TVP models with heteroscedastic errors, typically modeled via a stochastic volatility specification.

However, parameter estimation and state filtering remain a major issue in the analysis of time-varying parameter models with stochastic volatility. Many studies have only focused on filtering via particle filters, which are inefficient and not suitable for parameter estimation due to the discontinuity introduced by the resampling step. To avoid the intricate estimation of these models, some authors have chosen to calibrate parameter values (e.g. Stock & Watson, 2007). This paper presents an efficient algorithm for parameter estimation and state filtering of heteroscedastic time-varying parameter models. Our method relies on a conditional Gaussian likelihood function representation with marginalization of latent linear states, and simulation of

*Aalborg University Business School, Aalborg University, Fibigerstraede 2, 9220 Aalborg, Denmark. Email: det@business.aau.dk

non-linear latent states via efficient importance sampling. It is important to mention that our approach delivers an efficient and continuous estimator for the log-likelihood function, which can be numerically maximized to find parameter estimates.

Statistical inference for time-varying parameter models with stochastic volatility is complex. Latent non-linear states lead to multidimensional filtering integrals with no closed form solution. Typical applications involve the use of Bayesian techniques such as Metropolis-Hastings. A common hindrance in non-linear time-varying parameter models is that their conditional posterior distributions are easily not available. Therefore, standard algoritms such as Durbin and Koopman (2002) do not lead to a valid posterior simulator. On the other hand, we base our methodology on a multivariate extension of the Numerically Accelerated Importance Sampling (NAIS) of Koopman, Lucas, and Scharth (2015), together with a Rao-Blackwellization step. We show that a conditionally linear and Gaussian state-space representation is possible given the state-vector. Thus, the likelihood conditional on this state-vector can be computed by the Kalman Filter. The procedure then simplifies to a sequence of Kalman Filters and ordinary least squares.

This paper extends time-varying parameter models by including autoregressive and stochastic volatility terms. This flexible model specification is able to account for time-variation in the conditional mean and variance in different ways. For instance, by allowing time-variation in the autocorrelation functions. After applying necessary restrictions, the model becomes a time-varying parameter model with non-linear latent states. Importantly, our modeling approach encompasses several relevant models discussed in the literature, for example the unobserved components with stochastic volatility. Therefore, our methodological contributions are easily applicable to several models of interest.

This paper's empirical contribution involves an application for modeling and forecasting inflation using time-varying parameter autoregressive models with stochastic volatility. An increasing number of studies have found that the inflation process is subject to time-variation in its trend and volatility (e.g. Chan, Koop, & Potter, 2013). During the 1960s and 1970s, the inflation trend and volatility increased substantially. In the aftermath of the Great Inflation of the 1970s, central banks of developed countries have made successful efforts not only to lower but also to stabilize inflation. These efforts have contributed to the Great Moderation of output volatility during the 1990s. More recently, inflation volatility peaks again as an outcome of the Global Financial Crisis. These facts highlight that parameter instability seems to be a feature of the inflation process and the forecasting model needs to be able to accommodate shifts in the mean and variance of the inflation process. The models are applied to forecast U.S. C.P.I. inflation and we perform an out-of-sample forecasting analysis in comparison to benchmark models.

The structure of the paper is as follows. Section 2 formulates the proposed model, necessary notation and restrictions. Section 3 presents our estimation procedure and our main methodological contributions, and section 4 investigates the finite sample properties of the importance sampling method in a simulation study. In section 5, we carry out our application on modeling and forecasting inflation.

2 Model specification

The basic framework we consider is the following time-varying parameter autoregressive model with stochastic volatility

$$y_t = z_t\phi_{z,t} + x_t\phi_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_t^2), \tag{1}$$

$$\sigma_t^2 = \exp(h_t), \tag{2}$$

where y_t is a scalar, $x_t = (y_{t-1}, ..., y_{t-p},)$ is the vector of observed lagged values of the dependent variable, $\phi_t = (\phi_{1,t}, ..., \phi_{p,t})'$ is the vector of latent time-varying autoregressive parameters. h_t is the latent log-volatility. z_t is a z-dimensional vector of observed exogenous regressors and $\phi_{z,t}$ is the associated z-dimensional time-varying parameters. The time-varying parameters and the stochastic volatility $(\phi_{z,t}, \phi_t, h_t)'$ are distributed according to a Gaussian autoregression process of order one

$$\begin{bmatrix} \phi_{z,t} \\ \phi_t \\ h_t \end{bmatrix} = d + T \begin{bmatrix} \phi_{z,t-1} \\ \phi_{t-1} \\ h_{t-1} \end{bmatrix} + \eta_t, \qquad \eta_t \sim N(0, Q). \tag{3}$$

The dynamics of the time-varying parameters are determined by the $l \times 1$ constant vector d, the $l \times l$ diagonal transition matrix T and the $l \times l$ covariance matrix Q, where $l = z + p + 1$. Note that the matrix Q is not assumed to be diagonal. All elements of the disturbance vectors η_t and ε_t are serially and mutually uncorrelated, that is $\mathbb{E}(\eta_t \varepsilon_t') = 0$ for all t. Our estimation procedure allows Q to be time-varying in a stochastic way when the state variable can be represented as conditionally linear and Gaussian variable.

Several models can be cast in the representation given by equations (1)-(3). For example, if we restrict $\phi_{1,t} = \cdots = \phi_{p,t} = 0 \ \forall t$, d as a null vector, T as an identity matrix, $z_t = 1$, the resulting specification belongs to the category of unobserved components model with stochastic volatility (e.g. Stock & Watson, 2007). Moreover, time-varying parameter models with stochastic volatility are also special cases. Finally, models with time-invariant parameters and stochastic volatility are also special cases. Our estimation procedure is able to handle these models with only small changes in the baseline algorithm.

However, time-varying parameter models depend on necessary restrictions to be imposed on some coefficients. For instance, as it is known, an AR model must have its roots within the unit circle at all times. Without further reparametrization it is not possible to guarantee that stability restrictions are going to be satisfied. Explosive paths may induce large errors which contaminate the explanatory and forecasting capabilities. Indeed, in the absence of restrictions, the estimated paths of the autoregressive parameters can fall into explosive regions generating filtered states and forecasts with large mean or variances.

2.1 Model restrictions

In this section, we discuss the restrictions imposed on the time-varying autoregressive parameters to guarantee that y_t has roots inside the unit circle at all times. It is worth mentioning that roots within the unit circle at all times is a necessary condition for a time-varying autoregressive model to be locally stationary (see Dahlhaus, 2012 for the conditions on local stationarity). Moreover, we also present how to introduce time-variation in the autoregressive parameters connecting them to a latent state.

To satisfy stability restrictions, a monotonic transformation has to be enforced on the necessary time-varying parameters. Let α_t be the unrestricted time-varying autoregressive parameters, we impose the link function $\Phi(.)$ to ensure that the resulting vector of parameters ϕ_t satisfy stability restrictions at all times

$$\phi_t = \Phi(\alpha_t). \tag{4}$$

The link function $\Phi(.)$ has to be continuous and strictly monotonic to avoid identifiability issues.

We continue by defining the restriction function that satisfies these conditions. Let $\rho_t = (\rho_{1,t}, \ldots, \rho_{p,t})'$ be the vector of partial autocorrelations for a given vector of AR coefficients ϕ_t, and $\nu_t = (\nu_{1,t} \ldots \nu_{p,t})'$ be the vector of associated roots. For the AR(1) case, stability can easily be imposed by restricting the partial autocorrelation, $\rho_{1,t}$, to be within the unit circle for every time t. In the general case, a well-defined time-varying AR(p) model can be imposed by

restricting all the partial autocorrelations $\rho_{i,t}$ $\forall i = 1, \dots p$, to be within the unit circle for each time t. This means that all characteristic roots (eigenvalues) are smaller than one in absolute value at each point in time.

Definition 2.1. The roots are within the unit circle for every time t if $\phi_t \in S^p$ $\forall t$, where S^p is the hyperplane in which all the roots ν_i are within the unit circle, i.e. $|\nu_{i,t}| < 1$ $\forall t$. Moreover, $\phi_t \in S^p$ $\forall t$ and $|\nu_{i,t}| < 1$ $\forall t$ if and only if $|\rho_{i,t}| < 1$ $\forall t$.

Definition 2.1 applies the results of Barndorff-Nielsen and Schou (1973) and Monahan (1984) to autoregressive models with time-varying parameters. To restrict the partial autocorrelations a monotonic fuction defined in $(-1, 1)$ is required. In this article we employ the the hyperbolic tangent function

$$\rho_t = \tanh(\alpha_t). \tag{5}$$

The hyperbolic tangent function can be understood as a rescaling of the logistic function, such that its outputs range from -1 to 1. Finally, a transformation $\psi(.)$ is necessary, which maps the partial autocorrelations, ρ_t, into autoregressive coefficients, ϕ_t. We then implement the Durbin-Levinson algorithm to map ρ_t into ϕ_t for every point in time.

Definition 2.2. The required transformation that maps the PACs into the ARs at every point in time is given by the following algorithm,

$$\begin{aligned} \phi_t^{i,k} &= \phi_t^{i,k-1} - \rho_{k,t}\phi_t^{k-1,i-1}, \\ i &= 1, \dots, k-1, \quad k = 2, \dots, p, \end{aligned}$$

where $\phi_t^{k,k} = \rho_{k,t}$, and $\phi_t^{1,1} = \rho_{1,t}$.

The transformation $\psi(.)$ is defined by the last column of the $\phi_t^{i,k}$ matrix. For example, when $p = 2$ this transformation can be easily written as

$$\phi_{1,t} = \rho_{1,t}(1 - \rho_{2,t}), \tag{6}$$

$$\phi_{2,t} = \rho_{2,t}. \tag{7}$$

It is possible to show that this correspondence is bijective and continuously differentiable (see Barndorff-Nielsen & Schou, 1973). Thus, the composite link function $\Phi(\alpha_t) = \psi\big(\tanh(\alpha_t)\big)$ maps the unrestricted time-varying parameters $\alpha_t \in (-\infty, \infty)$ into non-explosive autoregressive coefficients $\phi_t \in S^p$ for every time t. The composite link function $\Phi(\alpha_t)$ is also bijective and continuous, thus avoinding identifiability issues.

Time-varying autoregressive processes of order p require p partial autocorrelations that will be converted into restricted time-varying parameters via the Durbin-Levinson recursions. We propose to introduce time-variation in the partial autocorrelation vector via a scalar latent state. The partial autocorrelation vector is then defined as

$$\rho_t = \tanh(\Lambda\ \alpha_t), \tag{8}$$

where ρ_t is a $p \times 1$ vector, α_t is a scalar latent state following a state equation as (3), and Λ is a $p \times 1$ vector of static parameters. For identification purposes we impose the restriction $\Lambda_1 = 1$, i.e. the first element of the vector Λ equals one. This specification can result in more precise estimates, as usually time-varying parameters tend to be highly correlated and their covariance matrix is near-singular, see for example (Chan, Eisenstat, & Strachan, 2020). By concentrating the sources of time-variation of ρ_t in the scalar latent state α_t, we expect to obtain more precises parameter estimates while maintaining model flexibility.

3 Estimation procedure

This section develops an efficient algorithm for parameter estimation and state filtering of time-varying parameter models with non-linear latent states. The estimation procedure is a combination of the Kalman Filter with importance sampling techniques. We show that given the non-linear state vector, the model can be written via a conditionally linear and Gaussian state space representation. Hence, the measurement density can be conveniently computed by the prediction error decomposition step of the Kalman Filter. We also show that the importance sampler admits a linear representation, thus the sampling and the evaluation of the transition density is again based on the Kalman Filter and related techniques. An efficient log-likelihood estimator is also presented, which can deliver maximum likelihood estimates of the parameters.

3.1 Conditionally linear and Gaussian state-space representation

The likelihood function for model (1)-(3) is given by the integral over the joint density of observables y_t and latent states $\theta_t = (\phi_{z,t}, \alpha_t, h_t)'$

$$L(\gamma, y) = \int \prod_{t=1}^{T} p(y_t, \theta_t;\, \gamma) d\theta_1 \dots d\theta_T, \tag{9}$$

$$L(\gamma, y) = \int \prod_{t=1}^{T} p(y_t|\theta_t; \gamma) p(\theta_t|\theta_{t-1}; \gamma) d\theta_1 \dots d\theta_T, \tag{10}$$

where γ is the vector of static parameters. The likelihood function equation (10) is a $(z+2) \times T$ analytically intractable integral. The integrals in the expression for the likelihood function have to be approximated numerically via simulation methods. We apply importance sampling methods to evaluate (10).

However, note that it is possible to exploit the structure of the state-space model to take advantage of its conditional linearity, thus reducing the dimension of the likelihood function and increasing efficiency by integrating out the linear time-varying parameters via the Kalman Filter. For a given non-linear state vector $\tilde{\theta}_t = (\alpha_t, h_t)$, the model (1)-(3) becomes a traditional linear and Gaussian state-space model. Therefore, it is possible to evaluate the measurement density given the non-linear time-varying parameters by the Kalman Filter (KF). This procedure is known as Rao-Blackwellization, as it is an implication of the Rao-Blackwell Theorem. Given the 2-dimensional vector $\tilde{\theta}_t$, the linear states $\phi_{z,t}$ can be integrated out of (10) yielding

$$L(\gamma, y) = \int \prod_{t=1}^{T} p^*(y_t|\tilde{\theta}_t; \gamma)\, p(\tilde{\theta}_t|\tilde{\theta}_{t-1}; \gamma)\, d\tilde{\theta}_1 \dots d\tilde{\theta}_T, \tag{11}$$

and

$$p^*(y_t|\tilde{\theta}_t; \gamma) \equiv p(\upsilon_{t|t-1}|\tilde{\theta}_t; \gamma) \sim N(0, F_t), \tag{12}$$

where $p^*(y_t|\tilde{\theta}_t, \gamma)$ is the prediction error density, $\upsilon_{t|t-1}$ is the prediction error and F_t its associated variance delivered by Kalman filter given the non-linear time-varying parameters $\tilde{\theta}_t$. It is worth mentioning that a dimension reduction is obtained while increasing computational and statistical efficiency. The intuition behind this result is that we replace the importance sampling estimator for $\phi_{z,t}$ by the KF estimator. As the KF is an analytical filter its Monte Carlo variance is zero, thus always smaller than any importance sampling filter.

Alternatively, if the non-linear state vector $\tilde{\theta}_t$ is kept fixed, the resulting conditional state-space model is given by

$$y_t - \Phi\big(\Lambda\alpha_t\big) = y_t^* = z_t\phi_{z,t} + \varepsilon_t, \qquad \varepsilon_t \sim N\big(0, \exp(h_t)\big) \tag{13}$$

with the same transition equation. Given $\tilde{\theta}_t$, the measurement density $p^*(y_t|\tilde{\theta}_t;\gamma)$ is obtained by applying the Kalman Filter to the model (13). Note that the resulting conditional state-space representation is a regression model with linear time-varying parameters, thus its likelihood and filtering functions are easy to obtain by the KF. From the prediction error decomposition it is possible to compute $p^*(y_t|\tilde{\theta}_t;\gamma)$ as

$$p^*(y_t|\tilde{\theta}_t;\gamma) \sim N\big(y_t^* - z_t E[\phi_{z,t|t-1}], F_t\big), \tag{14}$$

where $y_t^* - z_t E[\phi_{z,t|t-1}]$ and F_t are the prediction error and its associated variance respectively. The updating and the smoothing steps of the Kalman Filter yield estimates of $E[\phi_{z,t|t}]$ and $E[\phi_{z,t|T}]$ respectively. A minor modification of the Kalman Filter is necessary when the matrix Q is not diagonal, see Schon, Gustafsson, and Nordlund (2005) for a discusion on marginalization in the particle filtering context.

However, it is still necessary to develop a procedure to sample trajectories of the non-linear state vector $\tilde{\theta}_t$. Therefore, we construct an efficient Gaussian importance sampler based on the Numerically Accelerated Importance Sampling method of Koopman et al. (2015) and the efficiency criteria of Richard and Zhang (2007).

3.2 Importance sampler

We obtain the importance sampler $g(\tilde{\theta}|y;\varphi)$ based on the linear Gaussian joint density $g(y,\tilde{\theta};\varphi)$. This density can be decomposed as $g(y,\tilde{\theta};\varphi) = g(y|\tilde{\theta};\varphi)g(\tilde{\theta};\varphi)$ where

$$g(y|\tilde{\theta};\varphi) = \prod_{t=1}^{T} g(y_t|\tilde{\theta}_t;\varphi_t), \tag{15}$$

$$g(\tilde{\theta};\varphi) = p(\tilde{\theta};\gamma) = \prod_{t=1}^{T} p(\tilde{\theta}_t|\tilde{\theta}_{t-1};\gamma). \tag{16}$$

Applying the Bayes' theorem, the Gaussian multivariate importance sampler $g(\tilde{\theta}|y;\varphi)$ is written as

$$g(\tilde{\theta}|y;\varphi) = \frac{g(y|\tilde{\theta};\varphi)g(\tilde{\theta};\varphi)}{g(y;\varphi)}. \tag{17}$$

For the sake of simplicity we have ommited the time subscripts. If the latent states follow Gaussian distributions, $g(\tilde{\theta};\varphi)$, $g(\tilde{\theta}|y;\varphi)$, $g(y;\varphi)$ are all Gaussian densities. $g(y;\varphi)$ is a normalizing constant and φ is the vector of parameters of the importance density $g(\tilde{\theta}|y;\varphi)$. Given the linear and Gaussian assumption for the transition equations, it follows that $g(\tilde{\theta};\varphi) \equiv p(\tilde{\theta};\gamma)$.

The important step in the procedure is the representation of $g(y_t|\tilde{\theta}_t;\varphi_t)$ as a linear and Gaussian state-space density

$$g(y_t|\tilde{\theta}_t;\varphi_t) = \exp(a_t + b_t'\tilde{\theta}_t - \frac{1}{2}\tilde{\theta}_t' C_t \tilde{\theta}_t), \tag{18}$$

where a_t is an integrating constant, and the importance parameters $b_t \in \mathbb{R}^2$ and $C_t \in \mathbb{R}^{2\times 2}$ for $t = 1, \dots T$ are a function of y_t and γ. The importance sampler (17) and the observation density (18) are only functions of the auxiliary parameters φ_t, which depend on the data y_t. Therefore, given $\{b_t\}_{t=1}^T$ and $\{C_t\}_{t=1}^T$, it is possible to construct the importance density.

Following Koopman et al. (2015), the proposal densities $g(y_t|\tilde{\theta}_t;\varphi_t)$ and $g(\tilde{\theta}_t|y_t;\varphi_t)$ can be conveniently computed via the constructed variable $\bar{y}_t = C_t^{-1} b_t$ with measurement equation

$$\bar{y}_t = \tilde{\theta}_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, C_t^{-1}), \tag{19}$$

where the transition equation remains the same as the original model. Therefore, it can be shown that

$$\log g(\bar{y}_t|\tilde{\theta}_t;\varphi_t) = -\log(2\pi) + \frac{1}{2}\log|C_t| - \frac{1}{2}\{(C_t^{-1}b_t - \tilde{\theta}_t)'C_t(C_t^{-1}b_t - \tilde{\theta}_t)\}, \quad (20)$$

$$= a_t + b_t'\tilde{\theta}_t - \frac{1}{2}\tilde{\theta}_t'C_t\tilde{\theta}_t, \quad (21)$$

where $a_t = (\log|C_t| - 2\log 2\pi - b_t'\bar{y}_t)/2$ is an integrating constant and collects terms not related to $\tilde{\theta}_t$. We can conclude that $g(\tilde{\theta}|y;\varphi) = g(\tilde{\theta}|\bar{y};\varphi)$. Hence, the linear state-space representation (19) implies that we can use the Kalman Filter and related techniques to evaluate and sample from $g(\tilde{\theta}|y;\varphi)$ for a given set of importance parameters $\varphi = \left(\{b_t\}_{t=1}^T, \{C_t\}_{t=1}^T\right)$. Given this linear representation, the smoothing density $g(\tilde{\theta}_t|y_t;\varphi_t)$ can be written as

$$g(\tilde{\theta}_t|y_t;\varphi_t) = \frac{1}{\sqrt{(2\pi)|V_{t|T}|}}\exp\left(-\frac{1}{2}(\tilde{\theta}_t - \mu_{t|T})'(V_{t|T})^{-1}(\tilde{\theta}_t - \mu_{t|T})\right), \quad (22)$$

where $\mu_{t|T}$ is the expected value and $V_{t|T}$ is the covariance matrix of the non-linear vector $\tilde{\theta}_t$ conditional on information of the whole sample, and both are obtained by the Kalman Smoother applied to the importance sampler model (19) given a set of importance parameters $\varphi = \left(\{b_t\}_{t=1}^T, \{C_t\}_{t=1}^T\right)$.[1]

The linear representation for the importance sampling density allows us to use standard linear techniques, such as the simulation smoother of Durbin and Koopman (2002), to sample non-linear states. This contrasts with other available methods for time-varying parameter models where the sampling must take into account constraints, usually by rejection sampling. This may, however, lead to heavy inefficiencies, see for example Koop and Potter (2011). It is important to mention that $g(\tilde{\theta}|y;\varphi)$ has a linear state-space representation as long as the transition equations are linear and Gaussian.

Using this importance sampler, it is possible to evaluate the likelihood function, given some regularity conditions (see Geweke, 1989),

$$L(\gamma, y) = \int \frac{p^*(y|\,\tilde{\theta};\gamma)\,p(\tilde{\theta};\gamma)\,g(y,\tilde{\theta};\varphi)}{g(y,\tilde{\theta};\,\varphi)}d\tilde{\theta}. \quad (23)$$

Using the Bayes' Theorem, we can apply $g(y,\tilde{\theta};\varphi) = g(\tilde{\theta}|y;\varphi)g(y;\varphi)$ and $g(y,\tilde{\theta};\varphi) = g(y|\tilde{\theta};\varphi)g(\tilde{\theta};\varphi)$ in the numerator and denominator of (23) to find

$$L(\gamma, y) = g(y;\varphi)\int \frac{p^*(y|\,\tilde{\theta};\gamma)\,p(\tilde{\theta};\gamma)\,g(\tilde{\theta}|y;\varphi)}{g(y|\tilde{\theta};\varphi)g(\tilde{\theta};\varphi)}d\tilde{\theta}, \quad (24)$$

$$L(\gamma, y) = g(y;\varphi)\int \omega(\tilde{\theta}, y;\varphi)g(\tilde{\theta}|y;\varphi)d\tilde{\theta}, \quad (25)$$

as $g(\tilde{\theta};\varphi) \equiv p(\tilde{\theta};\gamma)$. $g(y;\varphi)$ is the likelihood of the importance density model. The importance weights are defined as

$$\omega(\tilde{\theta}, y;\varphi) = \frac{p^*(y|\,\tilde{\theta};\gamma)}{g(y|\,\tilde{\theta};\varphi)}, \quad (26)$$

evaluated at S independent trajectories of the non-linear time-varying parameters $\tilde{\theta}$. The likelihood estimator is then obtained by sampling from the density $g(\tilde{\theta}|y;\varphi)$ and evaluating (26)

$$\hat{L}(\gamma, y) = g(y;\varphi) \times \left[\frac{1}{S}\sum_{i=1}^{S}\omega(\tilde{\theta}^{(i)}, y;\varphi)\right] \quad (27)$$

[1] An alternative sampling scheme for computing the moments $\mu_{t|T}$ and $V_{t|T}$ is described in the appendix. The method is more robust to the case when somes matrices C_t are not positive definite.

Note that the likelihood estimator is reduced to compute an average of the ratio of two densities easily computed by the Kalman Filter. In constrast to particle filters, this likelihood estimator is a continuous function of the parameters of the model, thus it can be numerically maximized to find simulated maximum likelihood estimates. To complete the procedure, a criteria to obtain the importance parameters is necessary. Naive choices of φ ignore valuable information about the underlying latent process carried by the observables available at time t, and can be highly inefficient (Richard & Zhang, 2007). We apply the efficiency criteria of Richard and Zhang (2007), which select φ optimally, resulting in a fully adapted likelihood estimator. In other words, we sample the latent states so the likelihood estimator has the lowest possible variance (mean squared error).

3.3 Numerically Accelerated Importance Sampling

The criteria to find the importance parameters $\varphi = \left(\{\varphi_t\}_{t=1}^T\right)$ is to minimize the variance of $\hat{L}(\theta, y)$ in its full support. Richard and Zhang (2007) proposed that the appropriate Gaussian density can be found at every time t by solving the low dimensional integral

$$\varphi_t = \text{argmin} \int \lambda^2(y_t, \tilde{\theta}_t; \varphi_t)\omega(\tilde{\theta}_t, y_t; \varphi_t)g(\tilde{\theta}_t|y_t; \varphi_t)\ d\tilde{\theta}_t, \tag{28}$$

where $\lambda(.)$ is defined as

$$\lambda(y_t, \tilde{\theta}_t; \varphi_t) = \log\ p^*(y_t|\tilde{\theta}_t; \gamma) - \log\ g(y_t|\tilde{\theta}_t; \varphi_t). \tag{29}$$

According to Richard and Zhang (2007), this operational criteria approximates the problem of minimizing the MC variance (mean squared error) of the importance weights $\omega(\tilde{\theta}_t, y_t; \varphi_t)$ in its full support.

However, a main complication arises as the importance sampler $g(\tilde{\theta}_t|y_t; \varphi_t)$ itself depends upon the importance parameters φ_t. As in Richard and Zhang (2007), this can be resolved by a standard fixed point argument using intermediate importance samplers until convergence of $\left(\{\varphi_t\}_{t=1}^T\right)$. Therefore, the minimization problem (28) can be rewritten using the intermediate sampler $g(\tilde{\theta}_t|y_t; \varphi_t^{(k)})$

$$\varphi_t^{(k+1)} = \text{argmin} \int \lambda^2(y_t, \tilde{\theta}_t; \varphi_t^{(k)})\omega(\tilde{\theta}_t, y_t; \varphi_t^{(k)})g(\tilde{\theta}_t|y_t; \varphi_t^{(k)})\ d\tilde{\theta}_t. \tag{30}$$

To initiate the procedure, an initial condition $\varphi^{(0)}$ is required, which will be sequentially updated through iterations on (30) until a convergence criteria is satisfied.

The minimization (30) can be written as a weighted least squares problem, which has two possible solutions. Richard and Zhang (2007) applied Monte Carlo simulations and weighted least squares to find the auxiliary parameters, which approximate the minimization (30). Using M for draws $\tilde{\theta}_t^{(k)}$ at the k-iteration from the importance sampler, we have the following minimization problem

$$\log\ p^*(y_t, \tilde{\theta}_t^{(k)}; \gamma) = \text{constant} + \kappa'\ \tilde{\theta}_t^{(k)} - \frac{1}{2}\xi'\text{vech}(\tilde{\theta}_t^{(k)}\ \tilde{\theta}_t^{(k)}) + \text{error}, \tag{31}$$

with solution

$$\hat{\varphi_t}^{(k+1)} = (X_t^T W_t X_t)^{-1} X_t^T\ W_t\ \log\ p^*(y_t|\tilde{\theta}_t^{(k)}; \gamma), \tag{32}$$

where $X_t = [\mathbf{1}\ \tilde{\theta}_t^{(k)}\ \text{vech}(\tilde{\theta}_t^{(k)}\ \tilde{\theta}_t^{(k)})]$, and vech(.) stacks the elements of a symmetric matrix into a vector. W_t is a diagonal matrix of weights, $\omega(\tilde{\theta}_t^{(k)}, y_t; \varphi_t^{(k)})$. The least square estimates for κ

and ξ are the new estimates of $b_t^{(k+1)}$ and $C_t^{(k+1)}$ respectively for each time $t = 1 \dots T$. This method has been denominated Efficient Importance Sampling (EIS).

Alternatively, Koopman et al. (2015) proposed numerical integration by Gauss-Hermite (GH) methods to find estimates for φ_t, which are computationally faster and numerically more accurate when $\tilde{\theta}_t$ is of low dimension at time t. This variant is denominated Numerically Accelerated Importance Sampling (NAIS). Let $\{z_i\}_{i=1}^{M^2}$ be the set of abscissae after all combinations of M predefined GH nodes and $h(z_i)$ be the respective Gauss-Hermite weights. The weights are given by $\omega(\tilde{\theta}_t^{(k)}, y_t; \varphi_t^{(k)})h(z_i)\exp(\frac{1}{2}z_i^2)$ and $\tilde{\theta}_t^{(k)} = \mu_{t|T}^{(k)} + V_{t|T}^{0.5} z_i$, for $i = 1, \dots M^2$, where $V_t^{0.5}$ is the root matrix computed via the Cholesky decomposition of $V_{t|T}^{(k)}$. The solution for every time t is

$$\hat{\varphi_t}^{(k+1)} = (X_t^T W_t X_t)^{-1} X_t^T \; W_t \; \log \, p^*(y_t|\tilde{\theta}_t^{(k)}; \gamma), \tag{33}$$

where $X_t = [\mathbf{1} \;\; \tilde{\theta}_t^{(k)} \;\; \text{vech}(\tilde{\theta}_t^{(k)} \;\; \tilde{\theta}_t^{(k)})]$, and vech(.) stacks the elements of a symmetric matrix into a vector. W_t is a diagonal matrix of weights. The iteration must be initialized with starting values for $\varphi_t = (b_t, C_t)$ for each time $t = 1 \dots T$. Generally, a vector of zeros for b_t and a proportion of the identity matrix for C_t will suffice. The iterations can continue until convergence is achieved. However, convergence of the algorithm is not crucial as only the initial iterations typically generate substantial reductions in the variance of the likelihood estimate (DeJong, Dharmarajan, Liesenfeld, Moura, & Richard, 2013).

The set of static parameters can thus be estimated by numerically maximizing the likelihood function (27). An estimator for the time-varying parameters can be easily constructed based on the importance sampling methodology. For example, the mean of the smoothed density of $\exp(h_t)$ can be estimated by the importance sampling procedure as

$$\widehat{\exp(h_t)} = \frac{\sum_{i=1}^{S} \exp(h_t^{(i)})\omega_i(.)}{\sum_{i=1}^{S} \omega_i(.)}. \tag{34}$$

As the nominator and the denominator are evaluated using the same draws of the importance sampler, the estimator (34) can be written using normalized weights

$$\widehat{\exp(h_t)} = \sum_{i=1}^{S} \exp(h_t^{(i)})\hat{\omega}_i(.), \qquad \hat{\omega}_i(.) = \frac{\omega_i(.)}{\sum_{k=1}^{S} \omega_k(.)}. \tag{35}$$

Analogously, an importance sampling estimator can be constructed for the latent states $\phi_{z,t}$ and ϕ_t as

$$\hat{\phi}_{z,t} = \sum_{i=1}^{S} \phi_{z,t}^{(i)} \hat{\omega}_i(.) \tag{36}$$

$$\hat{\phi}_t = \widehat{\Phi(\alpha_t)} = \sum_{i=1}^{S} \Phi(\Lambda \alpha_t^{(i)}) \hat{\omega}_i(.). \tag{37}$$

The Monte Carlo maximum likelihood method previously developed is based on the importance sampling technique, thus when $T \to \infty$ and $S \to \infty$ the likelihood estimator converges to the true likelihood value (see Geweke, 1989 for conditions of consistency of importance sampling). Therefore, given usual assumptions on maximum likelihood estimation, asymptotic properties apply as well (see Hamilton, 1994 for linear state-space models and the Kalman Filter). However, the likelihood function is subject to simulation error in common samples sizes. Thus, the next section investigates the finite-sample properties of the maximum likelihood via a simulation study.

4 Simulation study

In this section, we perform a simulation experiment to investigate the finite-sample properties of the importance sampling procedure on an empirically relevant set of parameters. In this simulation study, we focus on the following model

$$y_t = \phi_{0,t} + \tanh(\alpha_t) y_{t-1} + \exp(h_t/2)\varepsilon_t, \qquad \varepsilon_t \sim N(0,1). \tag{38}$$

The dynamics of the time-varying parameters are determined by the subsequent autoregressive processes

$$\begin{aligned} \phi_{0,t} &= \beta_0 \phi_{0,t-1} + \sigma_0 \eta_{0,t}, \qquad \eta_{0,t} \sim N(0,1), & (39)\\ \alpha_t &= \beta_1 \alpha_{t-1} + \sigma_1 \eta_{1,t}, \qquad \eta_{1,t} \sim N(0,1), & (40)\\ h_t &= \delta + \beta_h h_{t-1} + \sigma_h \eta_{h,t}, \qquad \eta_{h,t} \sim N(0,1), & (41) \end{aligned}$$

where $\beta_0 = \beta_1 = 0.98$. This choice for the state equation implies we can analyze stationary time-varying parameters, however, taking into account that their behaviour may resemble a unit root process in finite samples. Hence, parameter estimation will be more difficult when compared to the usual specification of random walks, as there is an additional parameter in the state equation. Hence, we investigate the identifiability of β_0 and β_1. The remaining vector of statistic parameters is set at $(\sigma_0, \sigma_1, \delta, \beta_h, \sigma_h) = (0.288, 0.054, 0.047, 0.874, 0.469)$, a choice which is guided by the parameter estimates found in the empirical application to inflation discussed in the next section.

We apply the Rao-Blackwellized NAIS (RB-NAIS) procedure to compute the auxiliary parameters with 20 GH nodes (i.e. 400 combined nodes), and 500 independent trajectories are used to draw latent states. We present mean, sample standard deviations, mean of statistical (asymptotical) standard errors for 500 estimates generated with sample sizes $T = 250$, $T = 500$, and $T = 1000$. We also discuss the sampling distribution of the parameters by looking at the 5%, 50%, and 95% quantiles. We present the RMSE of each parameter, and the total RMSE defined as

$$\text{Total RMSE} = \sqrt{\sum_{i=1}^{7} \text{MSE}_i}. \tag{42}$$

Equation (42) measures the average distance from the vector of estimates to the real vector of parameters. Additionally, some diagnostic tests are applied on expected standardized prediction errors defined as

$$\hat{\upsilon}_{t|t-1} = \sum_{i=1}^{S} \left(\frac{\upsilon_{t|t-1}^{(i)}}{\sqrt{F_t^{(i)}}} \right) \hat{\omega}_i(.), \tag{43}$$

where $\upsilon_{t|t-1}$ is the prediction error and F_t its associated variance delivered by prediction step of Kalman filter given the non-linear state vector. Expected standardized prediction errors should follow a standard Normal distribution, be uncorrelated and homoscedastic. We will apply the Kolmogorov–Smirnov (KS) Normality test, the Ljung–Box (LB) Q test with 15 lags, and the ARCH-LM test with 10 lags. As the null hypotheses are true, we expect the tests to reject the null in only 5% of the cases asymptotically. However, the actual size of the tests will probably be higher in finite samples, and the closer they are to 5%, the more precise the tests are in finite samples.

Classical parameter estimation for non-linear state-space models models is difficult due to the intractable form of the likelihood function. The main issue for maximum likelihood estimation in

particle filtering is the resampling step, which prevents the use of optimization algorithms. Nevertheless, we compare the RB-NAIS maximum likelihood estimates with the local approximation importance sampling method of Shephard and Pitt (1997) and Durbin and Koopman (1997), which we refer to as SPDK. This is based on a second-order Taylor expansion of log $p^*(y_t|\tilde{\theta}_t;\gamma)$ around the conditional mode of $p(\tilde{\theta}_t|y_t;\gamma)$ which can be computed iteratively using the Kalman Filter applied to the importance sampling model (19). This method yields continuous but not efficient approximations to the likelihood function meaning that its variance will likely be large. More details on this method can be found in Durbin and Koopman (2012). The local approximation method has been used for stochastic volatility models producing accurate estimates, see for example Jungbacker and Koopman (2007). However, our model has complex features including linear and non-linear time-varying parameters, which can be challenging for a method based on a local approximation.

The estimates for the NAIS method are presented in table 1. We note that for small sample sizes ($T = 250$), all the parameters have large sample standard deviations, but for $T = 500$ and $T = 1000$ they become much smaller. As a consequence, the RMSE also significantly decreases for every parameter and in total, indicating that the average distance between the estimator and real parameter reduces with larger sample sizes. The average of asymptotical standard errors tend to become smaller when the sample size increases. This means that the curvature of the log-likelihood function increases with more information, and it becomes easier to identify parameters. Quantiles also suggest that the sampling distribution of the parameters is getting closer to the true value as the sample size increases. Analysis of prediction errors shows that the size of the asymptotic tests are slightly larger than the expected 5% for the KS and the ARCH-LM. This means that the asymptotic distribution approximation is not precise at the sample size $T = 1000$. On the other hand, as indicated by the LB test, prediction errors seem to be uncorrelated even in small samples.

It is important to highlight that the autoregressive parameters of non-linear states, β_h and especially β_1, are poorly estimated in small samples, displaying large standard errors and broad sampling distribution. However, they improve subtantially when $T = 1000$. Hence, these parameters have identification issues in very small samples, and larger time series are necessary to identify them.

The estimates for the SPDK method are presented in table 2. We observe that all parameter estimates have larger RMSE, standard errors, and are more dispersed than the RB-NAIS for the same sample size. Moreover, some parameters have broad sampling distribution even when $T = 1000$, especially the autoregressive parameters of latent states. These results are due to the fact that the SPDK method generates likelihood estimates with larger variances. Koopman et al. (2015) show that the SPDK method yields log-likelihood estimates of SV models with variances 30 to 40 times larger than NAIS. We suspect that this difference could be larger for time-varying parameter models due to their higher complexity.

Importance sampling provides consistent estimates of moments of latent variables when $T \to \infty$ and $S \to \infty$. Thus, it is relevant to verify how close the estimated and the real states are in finite-samples. Especially as most in-sample and out-of-sample analyses rely on estimated (i.e. smoothed) state variables. An example is out-of-sample forecasting, in which we need a reliable estimator for the conditional mean to obtain accurate forecasts. We evaluate how close the smoothed estimates of the latent states are to the simulated time-varying parameters by applying the RMSE across all simulated data, for example

$$\text{RMSE} = \sqrt{\frac{1}{500T}\sum_{i=1}^{500T}\left(\exp(h_i) - \widehat{\exp(h_i)}\right)^2} \tag{44}$$

Table 1: Parameter estimates RB-NAIS

Parameter	$T = 250$	$T = 500$	$T = 1000$
$\beta_0 = 0.98$			
Estimates	0.964 (0.028) [0.014]	0.968 (0.017) [0.010]	0.971 (0.011) [0.007]
Quantiles	0.900, 0.973, 0.992	0.934, 0.972, 0.988	0.950, 0.973, 0.985
RMSE	0.032	0.021	0.014
$\sigma_0 = 0.288$			
Estimates	0.354 (0.132) [0.106]	0.328 (0.087) [0.106]	0.321 (0.052) [0.043]
Quantiles	0.178, 0.333, 0.594	0.201, 0.323, 0.473	0.235, 0.320, 0.407
RMSE	0.147	0.096	0.061
$\beta_1 = 0.98$			
Estimates	0.760 (0.296) [0.138]	0.848 (0.235) [0.076]	0.949 (0.091) [0.018]
Quantiles	0.048, 0.920, 0.987	0.247, 0.952, 0.990	0.773, 0.974, 0.991
RMSE	0.369	0.269	0.096
$\sigma_1 = 0.054$			
Estimates	0.096 (0.055) [0.038]	0.080 (0.041) [0.026]	0.058 (0.023) [0.015]
Quantiles	0.035, 0.083, 0.210	0.033, 0.069, 0.161	0.029, 0.052, 0.101
RMSE	0.070	0.048	0.023
$\delta = 0.047$			
Estimates	0.049 (0.076) [0.051]	0.051 (0.052) [0.029]	0.047 (0.022) [0.018]
Quantiles	-0.053, 0.036, 0.186	0.003, 0.041, 0.123	0.015, 0.043, 0.088
RMSE	0.077	0.051	0.022
$\beta_h = 0.874$			
Estimates	0.769 (0.190) [0.058]	0.846 (0.094) [0.030]	0.876 (0.041) [0.018]
Quantiles	0.343, 0.836, 0.931	0.697, 0.870, 0.930	0.806, 0.881, 0.928
RMSE	0.217	0.098	0.041
$\sigma_h = 0.469$			
Estimates	0.567 (0.205) [0.124]	0.473 (0.126) [0.088]	0.425 (0.075) [0.058]
Quantiles	0.293, 0.517, 0.959	0.307, 0.455, 0.702	0.314, 0.418, 0.552
RMSE	0.227	0.127	0.087
Total RMSE	0.518	0.336	0.154
KS	0.066	0.068	0.076
LB	0.050	0.064	0.028
ARCH	0.128	0.114	0.138

The table presents sample average of estimated parameters from 500 estimations. Values in parentheses are sample standard deviations of the estimates. Values in brackets are the average of statistical standard errors. We present 5%, 50%, and 95% quantiles respectively. 500 trajectories are used to evaluate the log-likelihood function. The first 100 observations were excluded for each simulated series to reduce the effects of the initial conditions.

where $\exp(h_i)$ and $\widehat{\exp(h_i)}$ are the vectorized variance and its vectorized estimate respectively. Analogously, the RMSE can be computed for the remaining time-varying parameters. As importance sampling provides a consistent estimate of the conditional mean of the latent state distribution at time t, we expect the RMSE to decrease when sample size increases. This means that the estimate of the conditional mean is improving with more observations. Furthermore, as the conditional mean provides the best forecast in terms of mean squared error, a lower RMSE implies more accurate estimates of the latent states. Results are presented in table (3). The RMSE decrease with sample size, indicating that estimates of the conditional mean of time-varying parameters are improving. Moreover, NAIS provides more accurate estimates for all time-varying parameters and sample sizes.

Table 2: Parameter estimates SPDK

Parameter	$T = 250$	$T = 500$	$T = 1000$
$\beta_0 = 0.98$			
Estimates	0.923 (0.083) [0.021]	0.934 (0.052) [0.016]	0.940 (0.034) [0.011]
Quantiles	0.781, 0.946, 0.989	0.841, 0.947, 0.982	0.889, 0.947, 0.980
RMSE	0.100	0.069	0.050
$\sigma_0 = 0.288$			
Estimates	0.560 (0.238) [1.343]	0.567 (0.196) [0.843]	0.578 (0.144) [0.570]
Quantiles	0.245, 0.519, 1.056	0.293, 0.543, 0.953	0.351, 0.579, 0.847
RMSE	0.361	0.341	0.326
$\beta_1 = 0.98$			
Estimates	0.673 (0.374) [0.195]	0.786 (0.326) [0.113]	0.810 (0.317) [0.081]
Quantiles	0.007, 0.912, 0.991	0.035, 0.962, 0.994	0.044, 0.980, 0.994
RMSE	0.483	0.379	0.377
$\sigma_1 = 0.054$			
Estimates	0.096 (0.101) [0.096]	0.070 (0.076) [0.056]	0.060 (0.065) [0.030]
Quantiles	0.001, 0.058, 0.324	0.007, 0.043, 0.256	0.001, 0.034, 0.219
RMSE	0.109	0.078	0.066
$\delta = 0.047$			
Estimates	0.065 (0.160) [0.055]	0.066 (0.133) [0.039]	0.069 (0.128) [0.032]
Quantiles	-0.143, 0.024, 0.4916	-0.015, 0.026, 0.477	-0.012, 0.027, 0.496
RMSE	0.161	0.134	0.141
$\beta_h = 0.874$			
Estimates	0.702 (0.287) [0.046]	0.755 (0.259) [0.031]	0.757 (0.250) [0.022]
Quantiles	0.101, 0.843, 0.946	0.103, 0.869, 0.948	0.110, 0.868, 0.935
RMSE	0.334	0.285	0.298
$\sigma_h = 0.469$			
Estimates	0.544 (0.363) [0.116]	0.457 (0.313) [0.079]	0.434 (0.299) [0.056]
Quantiles	0.217, 0.401, 1.350	0.191, 0.356, 1.340	0.195, 0.339, 1.349
RMSE	0.227	0.313	0.340
Total RMSE	0.814	0.684	0.652
KS	0.084	0.086	0.172
LB	0.396	0.554	0.766
ARCH	0.438	0.752	0.970

The table refers to the local approximation estimation technique. 500 estimations were obtained. Values in parentheses are sample standard deviations of the estimates. Values in brackets are the average of statistical standard errors. We present 5%, 50%, and 95% quantiles respectively. 500 trajectories are used to evaluate the log-likelihood function. The first 100 observations were excluded for each simulated series to reduce the effects of the initial conditions.

5 Empirical application

Inflation forecasting is an important task for many central banks. As discussed in Clarida, Gali, and Gertler (1998), since the adoption of inflation-targeting regimes, central bankers based their monetary policy decisions on inflation and output gap forecasts. Furthermore, it is now widely accepted that decision making has become more complex in high and persistent inflation scenarios, as inflation may cloud public confidence as well as economic agents' assessments of future economic activity (Golob, 1994). Therefore, it is not surprising that a lot of effort has been devoted to the development of models that can accurately explain the dynamics and volatility of inflation rates.

The adoption of inflation-targeting regimes together with the changes in the inflation dynamics has led to a great interest in modeling and forecasting trend inflation. Stock and Watson

Table 3: RMSE between simulated and estimated time-varying parameters

RMSE	$T = 250$	$T = 500$	$T = 1000$
NAIS			
$\phi_{0,t}$	0.601	0.548	0.522
$\tanh(\alpha_t)$	0.222	0.205	0.188
$\exp(h_t)$	3.295	2.750	2.595
SPDK			
$\phi_{0,t}$	0.690	0.646	0.651
$\tanh(\alpha_t)$	0.258	0.238	0.236
$\exp(h_t)$	3.673	3.509	3.142

The table presents the RMSE between simulated time-varying parameters and the estimated conditional mean of time-varying parameters.

(2007) argue that the benchmark model for inflation forecasting seems to change over time and it has been hard to use a single model in different periods of time. For instance, Atkeson and Ohanian (2001) have shown that Phillips-curve type of models have performed well until the 1980s, but cannot display the same accuracy in recent decades, where simple statistical models, like the first order autoregressive model, have been dominating forecasts.

More recent evidence suggests the importance of the trend component in forecasting inflation. For example, Faust and Wright (2013) have argued that the precision of the forecasting depends on the estimate of the underlying trend inflation especially. In addition, forecasting trend inflation has become more relevant after the 1980s with the adoption of inflation-targeting regimes by many central banks. Meanwhile, some central banks do not have an official target, but still pursue stabilizing long-term inflation expectations. Although several methodologies are available, several studies have explored univariate models for trend inflation forecasting, usually outperforming other approaches such as survey-based forecasts (see Clark & Doh, 2014).

The empirical evidence emphasizes the importance of considering stochastic volatility and time-varying parameters to account for sudden changes in the exogenous shocks and variations in the propagation mechanism of these shocks. Consequently, recent studies have developed time-varying parameter models with stochastic volatility to model inflation (e.g. Clark & Doh, 2014). In this empirical section we propose time-varying autoregressive models with stochastic volatility (TVP-AR-SV hereafter) to forecast inflation. This model has three main features: time-varying trend, time-varying stochastic volatility, time-varying bounded persistence.

The time-varying parameter model we consider is given by

$$\pi_t = \phi_{0,t} + x_t \Phi(\Lambda\ \alpha_t) + \exp(h_t/2)\varepsilon_t, \qquad \varepsilon_t \sim N(0,1), \tag{45}$$

where x_t is the $1 \times p$ vector of observed lagged values of inflation, Λ is a $p \times 1$ vector of parameters, and α_t is a scalar latent variable. $\Phi(.)$ maps $\Lambda\alpha_t$ into a $p \times 1$ vector of time-varying autoregressive parameters, $\phi_{i,t}$ for $i = 1, \ldots, p$, with roots inside the unit circle at all times. The intercept, $\phi_{0,t}$, is a linear time-varying parameter defined by a random walk process

$$\phi_{0,t} = \phi_{0,t-1} + \sigma_0 \eta_{0,t}, \qquad \eta_{0,t} \sim N(0,1). \tag{46}$$

The random walk has the important implication that the trend inflation has a unit root. The scalar latent state α_t driving the partial autocorrelations also follows a random walk

$$\alpha_t = \alpha_{t-1} + \sigma_1 \eta_{1,t}, \qquad \eta_{1,t} \sim N(0,1). \tag{47}$$

The stochastic volatility is modeled via an autoregressive process of first order

$$h_t = \delta + \beta h_{t-1} + \sigma_h \eta_{h,t}, \qquad \eta_{h,t} \sim N(0,1). \tag{48}$$

The model capture changes in the inflation trend via the time-varying parameters and sudden shocks in the variance process through the stochastic volatility. At each time t the trend inflation can be approximated by the long-term forecast function

$$\lim_{h\to\infty} E[y_{t+h}|y_t] = \frac{\phi_{0,t}}{1-\sum_{i=1}^{p}\phi_{i,t}}. \tag{49}$$

Inflation forecasts are expected to converge to (49) at every time t when the effects of the transitory innovations vanish. Note that the trend inflation is a well-defined process due to the restrictions on the time-varying autogressive coefficients, $\phi_{i,t}$. Therefore, the non-linear constrains are necessary due to the central bank activity in controlling the trend inflation. We follow the literature on forecasting inflation and impose that the trend component has a unit root.

5.1 Estimation results

The models were estimated using annualized seasonally adjusted quarterly U.S. C.P.I. inflation from the first quarter of 1960 to the last quarter of 2019. More specifically, letting P_t be the quarterly CPI figures, we define $\pi_t = 400\log(P_t/P_{t-1})$ as CPI inflation. The data were obtained in the OECD Database. For the autoregressive orders, we select $p = 0, 1, 2, 4$. The case $p = 0$ is a local level model with stochastic volatility (LL-SV), and was selected to compare the effects of including autoregressive terms. The maximum likelihood estimation and state smoothing were performed using the Rao-Blackwellized NAIS with 20 Gauss-Hermite nodes, and 500 draws of the importance sampler were used to evaluate the log-likelihood function. The estimates are presented in table (4). Most parameters are statistically significant. It is interesting to note that the estimate of the volatility of the intercept (σ_0) becomes smaller when we introduce more autoregressive terms in the model. Thus, there is a trade-off between time-variation in the intercept and in the autoregressive terms. However, including autoregressive parameters leads to an improvement in the log-likelihood value, the Akaike information criteria, and the Ljung-Box Q-test. In terms of residual diagnostics, we observe that the null hypotheses of Normal distribution and homoscedasticity are not rejected. However only for the TVP-AR(4)-SV model, the null hypothesis of the Ljung-Box test is not rejected.

After estimating the parameters, they were used to obtain smoothed estimates of the time-varying parameters. Figure 1 presents the estimated long-run inflation across specifications. The most important difference among models is that the inclusion of time-varying autoregressive terms yields a more smoothed trend. On the other hand, the long-run inflation delivered by the LL-SV is much closer to the actual inflation. Thus, models with autoregressive parameters indicate that agents appear to stick to their long-run reference of inflation when making their forecasts. In other words, they do no expect that rises in inflation are necessarily completely incorporated in the inflation trend, but can be transitory effects. It is interesting that after the 1980s, the trend inflation seems to be stable. This can be explained by the public belief that the Federal Reserve has increased its commitment to price stability, resulting in low and stable trend inflation with inflation expectations anchored in the inflation trend.

The volatility results in Figure 2 reveal that inflation was very volatile during the Great Inflation of the 1970s. On the other hand, estimates for the 1990s show a reduction in the volatility, which is consistent with price stability over the period. Recently, volatility peaked again in the aftermath of the Global Financial Crisis, reaching higher levels than those observed

Table 4: Estimation results

Parameter	LL-SV	TVP-AR(1)-SV	TVP-AR(2)-SV	TVP-AR(4)-SV
σ_0	0.838	0.288	0.209	0.083
	(0.088)	(0.062)	(0.127)	(0.116)
σ_1	NA	0.054	0.051	0.049
		(0.010)	(0.010)	(0.004)
δ	−0.057	0.047	0.048	0.061
	(0.083)	(0.029)	(0.047)	(0.031)
β	0.835	0.874	0.883	0.865
	(0.073)	(0.040)	(0.046)	(0.038)
σ_h	0.791	0.469	0.451	0.467
	(0.143)	(0.096)	(0.215)	(0.074)
Λ_2	NA	NA	0.042	−0.01
			(0.096)	(0.040)
Λ_3	NA	NA	NA	0.404
				(0.059)
Λ_4	NA	NA	NA	−0.248
				(0.012)
LogL	−454.596	−448.448	−445.374	−432.270
AIC	901.193	886.897	880.748	854.540
KS	0.049	0.051	0.045	0.051
	(0.576)	(0.547)	(0.689)	(0.534)
LB	44.782	38.942	39.131	22.529
	(0.000)	(0.000)	(0.000)	(0.094)
ARCH-LM	13.328	10.403	10.271	10.059
	(0.205)	(0.405)	(0.417)	(0.435)

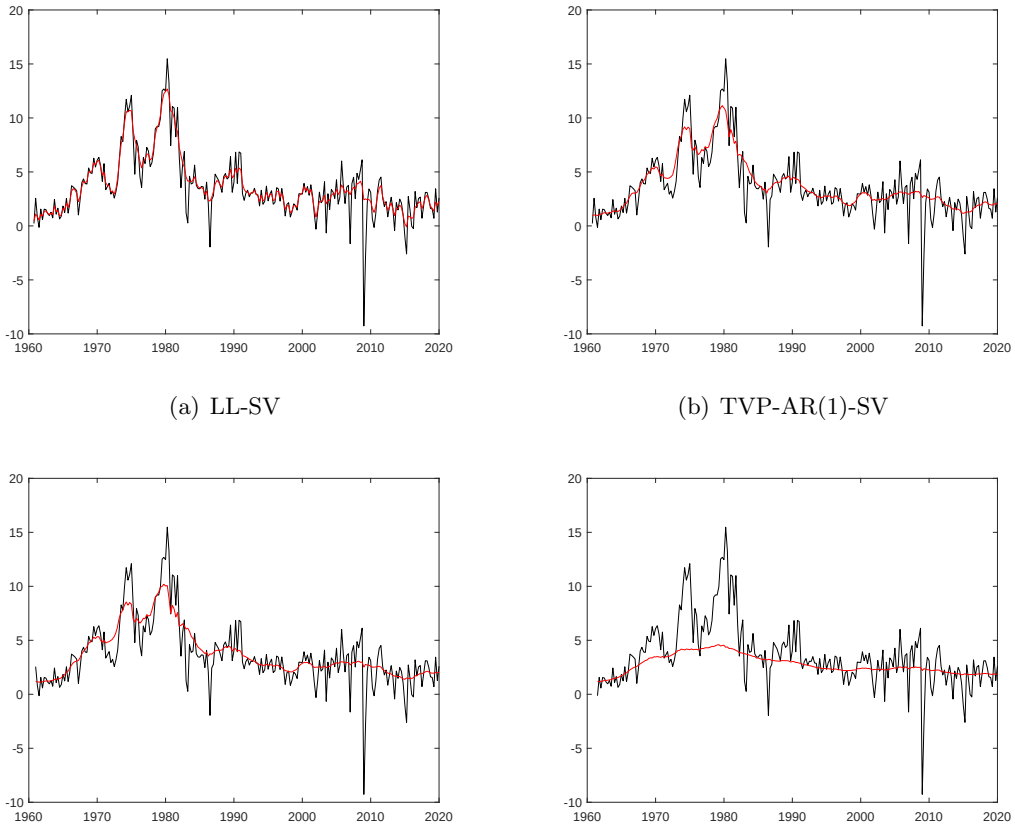
The table reports the parameter estimates of the time-varying autoregressive model with stochastic volatility for U.S. inflation. The statistical standard errors are in parentheses. LogL is the log-likelihood value. AIC is the Akaike Information Criterion. For diagnostics tests, we report test statistic and p-values in parentheses. KS is the the Kolmogorov-Smirnov normality test on expected standardized prediction errors. LB is the Ljung-Box Q-test. 15 lags were used in the LB test. ARCH-LM is the Lagrange multiplier test for autocorrelation in the expected standardized squared prediction errors. 10 lags were used in the ARCH-LM test.

during the 1970s, however without changes in the inflation trend. The volatility estimates are similar across models however the LL-SV seems to generate slightly larger estimates.

Figure 3 displays the sum of the autoregressive parameters as the measure of overall persistence in the model. Similarly to Cogley and Sargent (2002), our models suggest that the persistence of inflation increased during the 1960s and reached the maximum during the 1970s. A gradual decline was observed on the 1980s and afterwards. The estimated paths are similar across models however the TVP-AR(4)-SV seems to generate higher levels of persistence.

5.2 Forecasting evaluation

The proposed models were used to forecast U.S. inflation and their predictive power was measured by a pseudo out-of-sample forecasting exercise in comparison to univariate benchmark models. We consider a range of univariate models covering the features of inflation. Simple ARMA models: the random walk (RW), AR models, and the model with lowest Akaike information criterion in the full sample: ARMA(1,3). Models with time-varying trend: the local-level (LL) and time-varying autoregressive models (TVP-AR). Models with time-varying trend



(a) LL-SV

(b) TVP-AR(1)-SV

(c) TVP-AR(2)-SV

(d) TVP-AR(4)-SV

Figure 1: Inflation and its estimated trend. The dark line is the actual inflation. The red line is the trend inflation.

and volatility: Local-level with stochastic volatility (LL-SV), time-varying autoregressive models with stochastic volatility (TVP-AR-SV), and unobserved components stochastic volatility model (UCSV) of Stock and Watson (2007).

1. RW

$$\pi_t = \pi_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2). \tag{50}$$

2. AR(p) with $p = 1, 2, 4$

$$\pi_t = \phi_0 + \sum_{i=1}^{p} \phi_i \pi_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2). \tag{51}$$

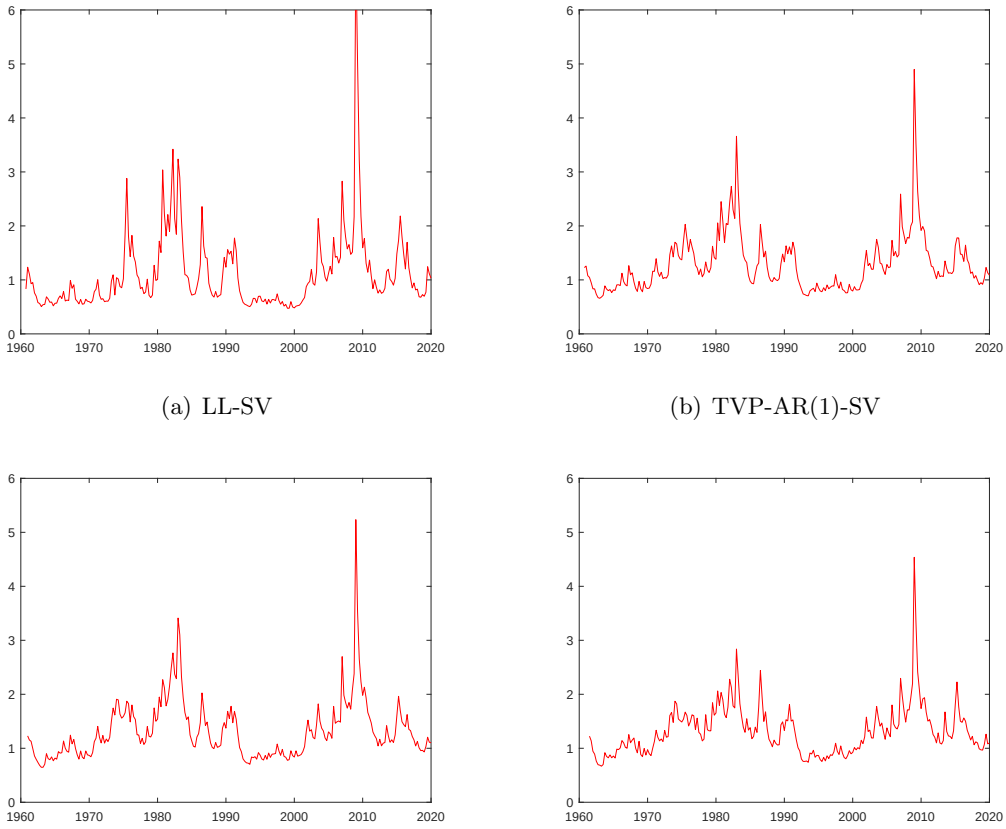
3. ARMA(1,3)

$$\pi_t = \phi_0 + \phi_1 \pi_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}, \qquad \varepsilon_t \sim WN(0, \sigma^2). \tag{52}$$

4. LL: Local level

$$\pi_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{53}$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2). \tag{54}$$



(a) LL-SV

(b) TVP-AR(1)-SV

(c) TVP-AR(2)-SV

(d) TVP-AR(4)-SV

Figure 2: Inflation volatility. The red line is the volatility estimate.

5. TVP-AR(p) with $p = 1, 2, 4$: Time-varying autoregressive models

$$\begin{aligned} \pi_t &= \phi_{0,t} + x_t \Phi(\Lambda\ \alpha_t) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2), && (55)\\ \phi_{0,t} &= \phi_{0,t-1} + \eta_{0,t}, \qquad \eta_{0,t} \sim N(0, \sigma_0^2), && (56)\\ \alpha_t &= \alpha_{t-1} + \eta_{1,t}, \qquad \eta_{1,t} \sim N(0, \sigma_1^2). && (57) \end{aligned}$$

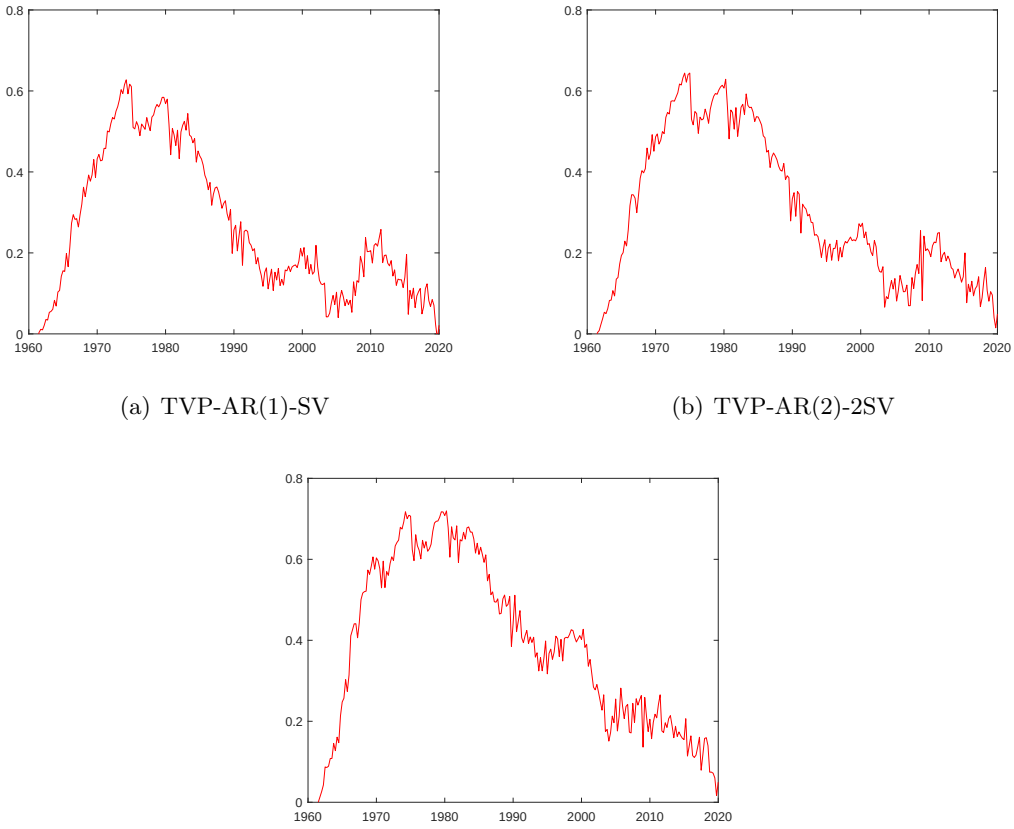
6. LL-SV: Local level with stochastic volatility

$$\begin{aligned} \pi_t &= \phi_{0,t} + \exp(h_t/2)\varepsilon_t, \qquad \varepsilon_t \sim N(0, 1), && (58)\\ \phi_{0,t} &= \phi_{0,t-1} + \eta_{0,t}, \qquad \eta_{0,t} \sim N(0, \sigma_0^2), && (59)\\ h_t &= \delta + \beta h_{t-1} + \eta_{h,t}, \qquad \eta_{h,t} \sim N(0, \sigma_h^2). && (60) \end{aligned}$$

7. UCSV: Unobserved components stochastic volatility

$$\begin{aligned} \pi_t &= \mu_t + \exp(h_{1,t}/2)\varepsilon_t, \qquad \varepsilon_t \sim N(0, 1), && (61)\\ \mu_t &= \mu_{t-1} + \exp(h_{2,t}/2)\eta_t, \qquad \eta_t \sim N(0, 1), && (62)\\ h_{1,t} &= h_{1,t-1} + \nu_{1,t}, \qquad \nu_{1,t} \sim N(0, \gamma), && (63)\\ h_{2,t} &= h_{2,t-1} + \nu_{2,t}, \qquad \nu_{2,t} \sim N(0, \gamma). && (64) \end{aligned}$$

8. TVP-AR(p)-SV with $p = 1, 2, 4$: Time-Varying AR models with stochastic volatility.



(a) TVP-AR(1)-SV

(b) TVP-AR(2)-2SV

(c) TVP-AR(4)-SV

Figure 3: Inflation persistence. The red line is the estimated sum of autoregressive parameters.

The TVP-AR(p) and the UCSV were also estimated by the Rao-Blackwellized importance sampling procedure using 20 GH nodes and 500 draws of the importance sampler. The estimation is done through an expanding window starting in the first quarter of 1985. The initial date includes part of the Great Inflation of the 1970s and was selected to test the forecasting ability of the model under different inflation conditions. This means that the forecasting sample covers the end of the Great Inflation, the entire Great Moderation of the 1990s and Global Financial Crisis. Forecasts are computed for 1, 2, 4, 8, 12, and 16 steps ahead, and are evaluated on the basis of the root mean square error (RMSE) and the mean absolute error defined as

$$\text{RMSE} = \sqrt{\frac{1}{(T-n)}\sum_{t=n}^{T} e_{t,h}^2}, \qquad \text{MAE} = \frac{1}{(T-n)}\sum_{t=n}^{T} |e_{t,h}|, \tag{65}$$

where n is the starting point of the forecasts, and $e_{t,h} = y_{t+h} - \hat{y}_{t+h|t}$ is the $t+h$ forecast error. The variable $\hat{y}_{t+h|t}$ is the importance sampling estimate for $E[y_{t+h}|y_t]$, and can be computed by

$$\hat{y}_{t+h|t} = \sum_{i=1}^{S} \left(\phi_{0,t+h|t}^{(i)} + x_{t+h|t}^{(i)} \Phi(\Lambda\ \alpha_{t+h|t}^{(i)}) \right) \hat{\omega}_i(.), \tag{66}$$

where $x_{t+h|t}^{(i)} = (y_{t+h-1|t}^{(i)}, \ldots, y_{t+h-p|t}^{(i)})$. Equation (66) estimates the expected value of the predictive density for $y_{t+h|t}$ based on importance sampling draws. This estimator can be further

simplified to filtered estimates of the latent states when no autoregressive terms are present. To construct the forecasts, the time-varying parameters are estimated until time t and updated forward in the forecast horizon based on their estimated law of motion using information up to time t.

The results of the forecasts are presented in Table 5 relative to the UCSV model as main benchmark, i.e. the relative RMSE of model i is defined as $\text{RMSE}_i/\text{RMSE}_{\text{UCSV}}$. Values smaller than one imply that the model i has smaller RMSE than the UCSV. We note that all models with time-varying coefficients forecast inflation well and outperform most benchmark specifications, especially in medium and longer horizons. For one-step ahead forecasts our models predict inflation as good as the UCSV. However for longer horizons, the UCSV is easily surpassed by our time-varying parameter models. These results suggest that models based on time-varying parameters are better estimators of the inflation trend than models based on unobserved components as the trend embedded in an inflation model plays a key role in longer-term inflation forecasts (Clark & Doh, 2014). Hence, they are able to measure the public's perception of the inflation goal of the central bank in a more accurate way. It is noteworthy that the inclusion of stochastic volatility reduces the forecasting errors for all horizons, but especially for short and medium step-ahead. Hence, it is the combination of time-varying autoregressive parameters and stochastic volatility which yields the best inflation forecasts.

The corrected Diebold-Mariano test proposed by Harvey, Leybourne, and Newbold (1997) is applied to compare the predictive performance of the forecasts. Let $L(.)$ be the mean squared error loss function, the loss differential between the benchmark model j and model i is

$$d_t = L(e_{t,h}^{(j)}) - L(e_{t,h}^{(i)}), \tag{67}$$

where $e_{t,h}^{(j)}$ are the forecasting errors of the benchmark models (1) to (7) previously presented, and $e_{t,h}^{(i)}$ are the forecasting errors of the TVP-AR-SV models. We test the null hypothesis

$$H_0 : E(d_t) = 0 \quad \forall t, \tag{68}$$

with alternative

$$H_1 : E(d_t) > 0 \quad \forall t. \tag{69}$$

Under the null the models have equally accurate forecasts, and under the alternative our models have more accurate forecasts. Note that the test provides only an approximation as the asymptotic normal distribution only holds for forecasts obtained by non-nested models.

Table (6) presents the results. We note that the differences on the forecasting accuracy are often statistically significant, especially for $h = 4$, $h = 12$, and 16 step-ahead forecasts. The TVP-AR-SV models easily outperform simple autoregressive and local level models in most forecasting horizons, but especially in the medium and long-run. It is interesting to highlight that complex models such as the UCSV were outperformed in medium and long forecasting horizons. This suggests that only unobserved components and stochastic volatility are not enough to provide the most accurate forecasts of the trend inflation. These results are consistent with our previous estimation using the whole dataset.

6 Conclusions

This paper develops efficient importance sampling methods for time-varying parameter autoregressive models with stochastic volatility. Stability and identifiability of time-varying autoregressive models are discussed. The estimation procedure is built on the importance sampling of Koopman et al. (2015) and Richard and Zhang (2007) together with a Rao-Blackwelization

Table 5: Forecast results

	RMSE						MAE					
	$h=1$	$h=2$	$h=4$	$h=8$	$h=12$	$h=16$	$h=1$	$h=2$	$h=4$	$h=8$	$h=12$	$h=16$
RW	1.132	1.257	1.270	1.234	1.232	1.236	1.146	1.246	1.242	1.212	1.155	1.202
AR(1)	1.054	1.111	1.109	1.115	1.155	1.179	1.074	1.124	1.148	1.135	1.300	1.329
AR(2)	1.038	1.097	1.091	1.096	1.140	1.181	1.035	1.105	1.113	1.116	1.259	1.328
AR(4)	1.036	1.090	1.086	1.077	1.104	1.160	1.040	1.085	1.087	1.092	1.206	1.298
ARMA(1,3)	1.021	1.081	1.093	1.055	1.060	1.112	0.999	1.066	1.094	1.062	0.990	1.218
LL	1.107	1.222	1.233	1.203	1.163	1.202	1.119	1.212	1.213	1.185	1.142	1.179
TVP-AR(1)	1.112	1.016	1.009	0.996	0.995	1.005	1.033	1.009	1.013	0.987	0.969	0.997
TVP-AR(2)	1.155	1.035	1.002	0.990	0.983	0.999	1.040	1.025	1.009	0.978	0.966	0.989
TVP-AR(4)	1.129	1.041	1.082	1.051	1.152	1.017	1.082	1.050	1.083	1.002	1.037	1.024
LL-SV	1.015	1.024	1.026	1.026	1.007	1.027	1.003	1.022	1.035	1.021	1.037	1.039
UCSV	1	1	1	1	1	1	1	1	1	1	1	1
TVP-AR(1)-SV	1.008	0.973	0.977	0.978	0.968	0.978	0.989	0.965	0.996	0.958	0.985	0.979
TVP-AR(2)-SV	1.003	0.964	0.966	0.970	0.963	0.964	0.993	0.962	0.993	0.956	0.980	0.965
TVP-AR(4)-SV	0.989	0.963	0.983	0.936	0.927	0.897	0.990	0.956	0.994	0.910	0.935	0.890

The table presents root mean square errors and mean absolute errors relative to the UCSV. Values smaller than one imply that the model has smaller RMSE than the UCSV. Forecasts are computed using an expanding window initially starting in 1985Q1, yielding 139, 138, 136, 132, 127, and 124 out-of-sample forecasts for h=1, 2, 4, 8, 12, and 16 respectively.

step. We show that the model has a linear and Gaussian representation conditional on the non-linear latent states. The method yields a numerically and computationally efficient estimation procedure, as only non-linear states need to be simulated. The likelihood estimator is continuous and efficient due to it having the lowest variance.

A simulation study for an empirically relevant set of parameters shows that the average of parameter estimates tend to be closer to the real parameters when the sample size increases. Furthermore, standard deviations and the RMSE also significantly decrease. The sampling distribution of the parameters seems to be getting closer to the true value as the sample size increases. The RB-NAIS method outperforms an alternative maximum likelihood procedure in all criteria. This suggests that the procedure is reliable and yields estimates with good finite-sample properties for the proposed model.

The model is then applied to forecast U.S. C.P.I. inflation. Smoothed estimates of the time-varying parameters show that the inflation trend has been stable since the 1980s, indicating the public belief in the Federal Reserve's commitment to price stability in the long-run. An out-of-sample forecasting exercise highlights the predictive power of the proposed model, outperforming simpler and more complex univariate benchmark specifications, especially in medium and longer horizons. This indicates that the models provide more accurate trend inflation estimates, as the success of a forecasting model in the long-run depends on its implied trend.

References

Atkeson, A., & Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, *25*(1), 2-11.

Barndorff-Nielsen, O., & Schou, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *Journal of Multivariate Analysis*, *3*, 408 - 419.

Chan, J. C. C., Eisenstat, E., & Strachan, R. W. (2020). Reducing the state space dimension in a large TVP-VAR. *Journal of Econometrics*, *218*, 105–118.

Chan, J. C. C., Koop, G., & Potter, S. M. (2013). A new model of trend inflation. *Journal of Business & Economic Statistics*, *31*(1), 94-106.

Table 6: Diebold-Mariano test results

Benchmark	$h=1$	$h=2$	$h=4$	$h=8$	$h=12$	$h=16$
			TVP-AR(1)-SV			
RW	**2.456**	**2.066**	**2.355**	1.405	0.837	**1.941**
AR(1)	0.917	**2.545**	**3.379**	**1.963**	**2.359**	**2.622**
AR(2)	0.479	**2.146**	**3.058**	**1.810**	**2.395**	**2.947**
AR(4)	0.378	1.600	**3.213**	1.610	**2.072**	**3.066**
ARMA(1,3)	0.193	1.668	**2.399**	1.309	0.422	**2.634**
LL	**1.875**	**2.012**	**2.369**	1.346	**1.926**	**1.934**
TVP-AR(1)	1.190	1.445	1.342	0.473	0.784	1.050
TVP-AR(2)	1.211	1.427	**1.660**	0.348	0.550	0.548
TVP-AR(4)	1.466	**1.656**	**1.834**	0.862	1.099	1.027
LL-SV	0.252	**2.227**	**2.444**	**2.819**	**1.799**	**1.900**
UCSV	-0.274	1.584	1.140	0.638	1.150	0.678
			TVP-AR(2)-SV			
RW	**2.454**	**2.059**	**2.323**	0.638	0.840	**1.937**
AR(1)	1.002	**2.462**	**3.289**	**2.109**	**2.398**	**2.852**
AR(2)	0.559	**2.113**	**2.933**	**1.952**	**2.416**	**3.189**
AR(4)	0.446	1.647	**2.983**	**1.721**	**2.049**	**3.307**
ARMA(1,3)	0.265	**1.711**	**2.312**	1.437	0.437	**2.869**
LL	**1.918**	**2.007**	**2.330**	1.380	**1.901**	**1.928**
TVP-AR(1)	1.200	1.497	**1.757**	0.726	0.904	1.216
TVP-AR(2)	1.217	1.483	**1.978**	0.624	0.665	0.779
TVP-AR(4)	1.461	1.640	**1.938**	0.961	1.128	1.170
LL-SV	0.390	**1.885**	**2.168**	**3.697**	**1.838**	**1.904**
UCSV	-0.114	1.677	1.561	1.005	1.316	0.938
			TVP-AR(4)-SV			
RW	**2.483**	**2.140**	**2.138**	1.565	0.918	**2.222**
AR(1)	1.511	**2.559**	**2.493**	**2.562**	**2.744**	**5.521**
AR(2)	0.981	**2.234**	**2.160**	**2.432**	**2.814**	**6.266**
AR(4)	0.888	**1.811**	**2.325**	**2.238**	**2.607**	**7.430**
ARMA(1,3)	0.706	**1.907**	**1.855**	**1.979**	0.563	**8.548**
LL	**2.177**	**2.095**	**2.111**	1.526	**2.134**	**2.239**
TVP-AR(1)	1.046	**1.757**	1.057	**1.789**	**2.337**	**1.704**
TVP-AR(2)	1.112	**1.818**	0.709	**1.665**	**2.362**	1.449
TVP-AR(4)	1.289	**1.762**	**1.925**	1.366	1.332	1.593
LL-SV	0.752	1.572	1.077	**3.298**	**2.846**	**2.512**
UCSV	0.413	1.460	0.593	**1.757**	**2.450**	**1.693**

A positive t-statistic indicates that the benchmark models produced larger average loss than the TVP-AR-SV models, while a negative sign indicates the opposite. In bold, we have the t-statistic greater than 1.65, which indicates a rejection of the null of equal predictive accuracy at the 0.05 level for one sided tests.

Clarida, R., Gali, J., & Gertler, M. (1998). *Monetary policy rules and macroeconomic stability: Evidence and some theory* (Working Paper No. 6442). National Bureau of Economic Research.

Clark, T. E., & Doh, T. (2014). A Bayesian evaluation of alternative models of trend inflation. *International Journal of Forecasing*, *30*, 426-448.

Cogley, T., & Sargent, T. J. (2002). Evolving post World War II U.S. in inflation. *NBER Macroeconomics Annual*, *16*, 331-388.

Dahlhaus, R. (2012). Locally stationary processes. In T. S. Rao, S. S. Rao, & C. R. Rao (Eds.), *Handbook of statistics 30: Time series analysis: Methods and applications* (p. 351-413). North Holland.

DeJong, D. N., Dharmarajan, H., Liesenfeld, R., Moura, G. V., & Richard, J.-F. (2013). Efficient

likelihood evaluation of state-space representations. *Review of Economic Studies*, *80*, 538-567.

Durbin, J., & Koopman, S. J. (1997). Monte carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, *84*(3), 669–684.

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state-space time series analysis. *Biometrika*, *89*, 603-616.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Faust, J., & Wright, J. H. (2013). Forecasting inflation. In G. Elliot & A. Timmerman (Eds.), *Handbook of economic forecasting* (Vol. 2, p. 2-56). North Holland.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, *577*, 1317-1339.

Golob, J. (1994). Does inflation uncertainty increase with inflation? *Economic Review*, *3*, 27-38.

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

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281-291.

Jung, R. C., Liesenfeld, R., & Richard, J.-F. (2011). Dynamic factor models for multivariate count data: An application to stock-market trading activity. *Journal of Business & Economic Statistics*, *29*, 73-85.

Jungbacker, B., & Koopman, S. J. (2007). Monte carlo estimation for nonlinear non-gaussian state space models. *Biometrika*, *94*(4), 827-839.

Koop, G., & Potter, S. M. (2011). Time varying VARs with inequality restrictions. *Journal of Economic Dynamics and Control*, *35*, 1126 - 1138.

Koopman, S. J., Lucas, A., & Scharth, M. (2015). Numerically accelerated importance sampling for nonlinear non-Gaussian state-space models. *Journal of Business and Economic Statistics*, *33*, 114-127.

Monahan, J. F. (1984). A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, *71*, 403-404.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, *72*, 821–852.

Richard, J.-F., & Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, *141*, 1385-1411.

Schon, T., Gustafsson, F., & Nordlund, P.-J. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, *53*(7), 2279-2289.

Shephard, N., & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, *84*(3), 653–667.

Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, *14*(1), 11-30.

Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*, 3-33.

A Backward-Forward Sampling Algorithm

This appendix discusses the Backward-Forward (BF) sampling algorithm to sample from a multivariate linear and Gaussian proposal density, which can be regarded as an alternative to the Kalman Filter. The algorithm is able to sample from the proposal density even when some

matrices C_t are close to singular. Moreover, infrequent problematic cases can be corrected by finding the closest symmetric positive definite matrix to C_t. The procedure is based on Jung, Liesenfeld, and Richard (2011), and modified by Koopman et al. (2015) to yield a more efficient sampling scheme. More details on the construction of the sampler can be found in Koopman et al. (2015). Initially, we rewrite the state equations for the non-linear latent states $\tilde{\theta}_t$

$$\tilde{\theta}_t = d + T\tilde{\theta}_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \Omega), \tag{70}$$

where d is the 2×1 vector of constants and T is the 2×2 matrix of autoregressive parameters.

The procedure is based on the following decomposition of the importance density for the states $\tilde{\theta}_t$

$$\begin{aligned} g(\tilde{\theta}|y;\varphi) &= \prod_{t=1}^{T} g(\tilde{\theta}_t|\tilde{\theta}_{t-1}, y_t; \varphi_t) && (71)\\ &= \prod_{t=1}^{T} g(y_t|\tilde{\theta}_t;\varphi_t) g(\tilde{\theta}_t|\tilde{\theta}_{t-1};\varphi_t)\chi(\tilde{\theta}_{t-1}, b_t, C_t)\zeta_t(\tilde{\theta}_t, b_{t+1}, C_{t+1}), && (72)\end{aligned}$$

where $\chi(\tilde{\theta}_{t-1}, b_t, C_t)$ is the integrating constant transferred from $t+1$, and $\zeta_t(\tilde{\theta}_t, b_{t+1}, C_{t+1})$ is a Gaussian kernel intended to approximate the proposal and the target integrand. Given the linearity of the representation $g(\tilde{\theta}|y;\varphi)$ the Gaussian kernel is proportional to the integrating constant

$$\zeta_t(\tilde{\theta}_t, b_{t+1}, C_{t+1}) \propto 1/\chi_{t+1}(\tilde{\theta}_t, b_t, C_t). \tag{73}$$

Thus the resulting sampling density is proportional to

$$g(\tilde{\theta}_t|\tilde{\theta}_{t-1}, y_t; \varphi_t) \propto g(y_t|\tilde{\theta}_t;\varphi_t) g(\tilde{\theta}_t|\tilde{\theta}_{t-1};\varphi_t). \tag{74}$$

The integrating constant is defined as

$$\chi_t(\tilde{\theta}_{t-1}, b_t, C_t) = \exp\left(r_t - q_t'\tilde{\theta}_{t-1} + \frac{1}{2}\tilde{\theta}_{t-1}' P_t \tilde{\theta}_{t-1}\right), \tag{75}$$

where r_t is a scalar, q_t is a $m \times 1$ vector, and P_t is a $m \times m$ matrix, with m being the dimension of the state vector $\tilde{\theta}_t$. The next period Gaussian kernel is, therefore, given by

$$\zeta_{t+1}(\tilde{\theta}_t, b_{t+1}, C_{t+1}) = \exp\left(q_{t+1}'\tilde{\theta}_t - \frac{1}{2}\tilde{\theta}_t' P_{t+1}\tilde{\theta}_t\right). \tag{76}$$

It follows then that the Gaussian sampler $g(\tilde{\theta}_t|\tilde{\theta}_{t-1}, y_t, ;\varphi_t)$ is proportional to

$$g(\tilde{\theta}_t|\tilde{\theta}_{t-1}, y_t; \varphi_t) \propto g(y_t|\tilde{\theta}_t;\varphi_t)\zeta_{t+1}(\tilde{\theta}_t, b_{t+1}, C_{t-1}) g(\tilde{\theta}_t|\tilde{\theta}_{t-1};\varphi_t). \tag{77}$$

It can be shown that the mean and variance of $g(\tilde{\theta}_t|\tilde{\theta}_{t-1}, y)$ are

$$\begin{aligned} \mu_t &= \Sigma_t(b_t + \Omega^{-1}(d + T\tilde{\theta}_{t-1}) + q_{t+1}), && (78)\\ \Sigma_t &= (\Omega_t^{-1} + C_t + P_{t+1})^{-1}. && (79)\end{aligned}$$

Given these expressions for the mean and variance of the importance density it is possible to rewrite the sampling of $\tilde{\theta}_t$ as a linear representation

$$\tilde{\theta}_t = d_t^* + T_t^*\tilde{\theta}_{t-1} + \eta_t^*, \qquad \eta_t^* \sim N(0, \Sigma_t), \tag{80}$$

where $d_t^* = \Sigma_t(b_t + \Omega^{-1}d + q_{t+1})$ and $T_t^* = \Sigma_t\Omega^{-1}T$. The initial condition for the recursion is $\tilde{\theta}_1 \sim N(d_0^*, \Sigma_1)$.

We continue by showing how to express the integrating constant, and the value of $r(.)$ required to compute the linear state-space representation for sampling the state vector. The function χ_t is given by the integrating constant of the Gaussian kernel $g(\tilde{\theta}_t|\tilde{\theta}_{t-1};\varphi_t)$ over the state variable $\tilde{\theta}_{t-1}$

$$\chi_t(\tilde{\theta}_{t-1}, b_t, C_t) = \sqrt{\frac{|\Omega|}{|\Sigma_t|}}\exp\left(\frac{1}{2}(d + T\tilde{\theta}_{t-1})'\Omega^{-1}(d + T\tilde{\theta}_{t-1})\right)\exp\left(-\frac{1}{2}\mu_t'\Sigma_t^{-1}\mu_t\right). \tag{81}$$

We can substitute the expression for the mean of the importance (78) into (81). After some manipulation this yields

$$\begin{aligned}
P_t &= T'\Omega^{-1}T - T'\Omega\Sigma_t\Omega^{-1}T, && (82)\\
q_t &= T\Omega^{-1}\Sigma_t(b_t + \Omega^{-1}d + q_{t+1}) - T'\Omega^{-1}d, && (83)\\
r_t &= \frac{1}{2}\log\left(\frac{|\Omega|}{\Sigma_t}\right) + \frac{1}{2}d'\Omega^{-1}d && (84)\\
&- \frac{1}{2}(b_t + \Omega^{-1}d + q_{t-1})'\Sigma_t(b_t + \Omega^{-1}d + q_{t+1}), && (85)
\end{aligned}$$

where the covariance matrix Σ_t was defined previously. We can initiate the recursions with $\chi_T(.) = 1$, thus P_{T+1} can be initialized with a matrix of zeros and q_{T+1} with a vector of zeros. The procedure starts with the backward filter, where we evaluate Σ_t, P_t, q_t, r_t. From these values it is obtained d_t^*, T_t^*. In the forward pass we compute the mean and the variance of the state vector

$$\begin{aligned}
E[\tilde{\theta}_t] &= d_t^* + T_t^* E[\tilde{\theta}_{t-1}], && (86)\\
V(\tilde{\theta}_t) &= T_t^* V(\tilde{\theta}_{t-1})T_t^{*'} + \Sigma_t, && (87)
\end{aligned}$$

where $E[\tilde{\theta}_t]$ and $V(\tilde{\theta}_t)$ can be used in the implementation of the Efficient Importance sampling and the Numerically Accelerated version of Koopman et al. (2015). Finally, the likelihood of the importance model is computed as

$$g(y;\varphi) = \prod_{i=1}^{T}\exp(-r_i), \tag{88}$$

which is necessary in the likelihood estimator equation (27) in the main text.