Monte Carlo Analysis of Time-Varying Parameter Models with Stochastic Volatility

PhD dissertation

Douglas Eduardo Turatti

Aarhus BSS, Aarhus University
Department of Economics and Business Economics

2018
MONTE CARLO ANALYSIS OF TIME-VARYING PARAMETER MODELS WITH STOCHASTIC VOLATILITY

By Douglas Eduardo Turatti

A PhD thesis submitted to
School of Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of
the PhD degree in
Economics and Business Economics

May 2018

CREATES
Center for Research in Econometric Analysis of Time Series
This dissertation is the result of my PhD studies at the Department of Economics and Business Economics at Aarhus University and was written between the period of March 2015 to March 2018. I am grateful to the Department of Economics and Business Economics as well as CREATES - Center for Research in Econometric Analysis of Time Series, for providing a unique research environment and financial support.

Several people deserve my gratitude. I would like to thank my main supervisor, Kim Christensen for his support during the PhD. I also would like to thank my co-supervisor Eric Hillebrand, and the Center Administrator Solveig Nygaard Sørensen. A special thanks to my master's supervisor Guilherme Valle Moura, for having supported my PhD applications and for having presented me many important econometric methods in which this thesis heavily relies on.

From April 2017 to July 2017 I had the great pleasure of visiting Prof. Siem Jan Koopman at the Department of Econometrics and Operations Research at VU Amsterdam. I had the opportunity to have several meeting with Siem Jan, who never lacked enthusiasm and ideas for my research projects.

I would like to thank all my friends in Denmark and in Brazil, especially my friend Fernando for many interesting conversations about research and econometrics. A special thank you to my parents, Izaura and Egino, and my sister Paula for their encouragement and support despite the distance. I would like to thank my girlfriend Ditte for her love during these years.

Douglas Eduardo Turatti
Aarhus, March 2018
The pre-defence meeting was held on the 24th April 2018, in Aarhus. I am grateful to the members of the assessment committee consisting of Charles S. Bos - VU Amsterdam -, Tommaso Proietti - University of Rome Tor Vergata -, and Martin M. Andreasen - Aarhus University -, for their careful reading of the dissertation and their many interesting comments and suggestions. Some of the suggestions have been incorporated into the present version of the dissertation while others remain for future research.

Douglas Eduardo Turatti
Aarhus, May 2018
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>vii</td>
</tr>
<tr>
<td>Danish summary</td>
<td>ix</td>
</tr>
<tr>
<td>1 Time-Varying Autoregressive Models with Common Factors: An Application to Inflation Forecasting</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Time-varying autoregressive models with common factors and stochastic volatility</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Model Restrictions</td>
<td>7</td>
</tr>
<tr>
<td>1.4 Estimation procedure</td>
<td>8</td>
</tr>
<tr>
<td>1.5 Monte Carlo Study</td>
<td>13</td>
</tr>
<tr>
<td>1.6 Modeling and Forecasting Inflation via Time-Varying Autoregressive Models</td>
<td>19</td>
</tr>
<tr>
<td>1.7 Conclusions</td>
<td>29</td>
</tr>
<tr>
<td>1.8 References</td>
<td>31</td>
</tr>
<tr>
<td>1.9 Appendix</td>
<td>33</td>
</tr>
<tr>
<td>1.10 Supplementary Appendix</td>
<td>36</td>
</tr>
<tr>
<td>2 Classical Analysis of Time-Varying Factor-Augmented Vector Autoregressions</td>
<td>43</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>44</td>
</tr>
<tr>
<td>2.2 Time-varying factor-augmented vector autoregressive models</td>
<td>45</td>
</tr>
<tr>
<td>2.3 Estimation and Filtering</td>
<td>47</td>
</tr>
<tr>
<td>2.4 Simulation Study</td>
<td>54</td>
</tr>
<tr>
<td>2.5 Empirical Application: Macroeconomic Forecasting with Large Datasets</td>
<td>57</td>
</tr>
<tr>
<td>2.6 Conclusions</td>
<td>61</td>
</tr>
<tr>
<td>2.7 References</td>
<td>62</td>
</tr>
<tr>
<td>2.8 Appendix</td>
<td>64</td>
</tr>
<tr>
<td>3 Univariate Maximum Likelihood Estimation of a TVP-VAR</td>
<td>77</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>78</td>
</tr>
<tr>
<td>3.2 Time-Varying VAR Models with Stochastic Volatility</td>
<td>79</td>
</tr>
</tbody>
</table>
This dissertation is composed of three self-contained, independent chapters. The modeling framework in all three chapters is that of conditionally linear state-space models with stochastic volatility. And the statistical methodology is the combination of simulation methods via importance sampling and analytical filters via the Kalman filter to approximate the analytical intractable likelihood function. In the first chapter we study statistical inference on time-varying autoregressive models with a focus on inflation forecasting. The second chapter introduces a time-varying vector autoregressive model augmented with dynamic factors, and proposes an estimation procedure. Finally, the third chapter proposes a new estimation procedure for time-varying vector autoregressive models with time-varying variances and covariances.

More in detail, all chapters are associated with non-linear state-space models, particularly time-varying parameter models with stochastic volatility. We discuss classical inference based on a combination of simulation methods and analytical filters. We show that the time-varying parameters can be treated analytically by the Kalman filter, and the volatility states are simulated by the importance sampling technique. Our base method combines the Numerically Accelerated Importance Sampling of Koopman et al. (2015) with the Rao-Blackwellized Efficient Importance Sampling of Moura and Turatti (2014). This new procedure has several desirable properties: The linear time-varying parameters are integrated out of the likelihood function by the Kalman filter, greatly increasing numerical efficiency; the parameters of the importance sampler are efficiently selected by Gauss-Hermite quadrature methods; and the sampling from the proposal density can be carried out by standard state-space methods.

In Chapter 1 we consider time-varying autoregressive model with stochastic volatility. In addition, a novel specification for this class of models is proposed based on common factors, greatly enhancing estimation. In contrast to the widely used Bayesian methods, we develop a classical maximum likelihood estimator. The estimation procedure is based on a multivariate extension of the numerically accelerated importance sampling together with a Rao-Blackwellization step to construct a highly efficient estimation procedure. We applied the model to U.S. C.P.I. inflation and we show it is able to accommodate the observed dynamics in inflation. An out-of-sample forecasting exercise showed superior results with respect to benchmark models.
In Chapter 2 we analyze estimation and forecasting with time-varying factor-augmented vector autoregressions (TVP-FAVAR) with stochastic volatility. We propose an estimation procedure in two steps. In the first one principal components are used as an estimator of the latent factors. In the second we show that the model can be estimated equation-by-equation conditional on the independence of the importance samplers. To estimate this sequence of regressions we propose the numerically accelerated importance sampling together with a Rao-Blackwellization step. In a simulation study we show that the estimation procedure delivers good estimates in finite samples and principal components are able to properly estimate the factor structure. An empirical application to a large dataset of U.S. macroeconomic variables shows that the TVP-FAVAR improves short-term forecasts of inflation measures.

In chapter 3, we propose an estimation procedure for time-varying vector autoregressive (TVP-VAR) models with stochastic volatility when the covariance matrix is subject to changes in the variances and covariances. Our estimation procedure takes advantage of the triangular form of the Cholesky decomposition and rewrite the TVP-VAR as a sequence of conditionally independent autoregressions. It is worth mentioning that the estimation procedure does not suffer from the curse of dimensionality. A Monte Carlo study shows that the equation-by-equation has good asymptotic and finite-sample properties. An empirical application to U.S. data generates results commonly found in the macroeconomics literature.

References


Danish summary


med hensyn til benchmark modeller.

I Kapitel 2 analyserer vi estimering og prognose med time-varying factor-augmented vector autoregressions (TVP-FAVAR) med stokastisk volatilitet. Vi foreslår en procedu- re for estimering, der har to trin. I det første bliver primære komponenter brugt som estimator for de latente faktorer. I det andet trin viser vi, at modellen kan estimeres ligning-for-ligning, på betingelse af importance samplernes uafhængighed. For at estimere denne sekvens af regression foreslår vi numerisk accelereret importance sampling sammen med et Rao-Blackwellization trin. I et simulationsstudie viser vi, at proceduren for estimering leverer gode estimator i finitte prøver og at primære kom- ponenter kan estimere faktorstrukturen ordentligt. En empirisk applikation til et stort dataset af amerikanske makroøkonomiske variabler viser at TVP-FAVAR forbedrer kortsigtede prognoser af inflationsindeks.

TIME-VARYING AUTOREGRESSIVE MODELS WITH COMMON FACTORS: AN APPLICATION TO INFLATION FORECASTING

Douglas Eduardo Turatti
Aarhus University and CREATES

Abstract

This paper considers time-varying autoregressive models with stochastic volatility. In contrast to the widely used Bayesian approaches, we develop a classical maximum likelihood estimator. The estimation procedure is based on a multivariate extension of the numerically accelerated importance sampling together with a Rao-Blackwellization step to construct a highly efficient estimation procedure. In addition, a novel specification for non-linear time-varying parameter models is developed summarizing the time-variation in the coefficients through a common factor structure, greatly enhancing estimation and retaining model flexibility. An extensive Monte Carlo study shows that the importance sampling procedure can identify model parameters and its likelihood estimator compares favorably to the particle filter. We applied the proposed specifications to U.S. C.P.I. inflation and we show it is able to accommodate the observed dynamics in inflation. 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 benchmark models, both in point and density forecasting. In addition,
well-calibrated forecasts can only be obtained when time-varying parameters and stochastic volatility are included.
1.1 Introduction

Structural instability is known to be an important feature in empirical time series (see, Stock and Watson, 1996; Cogley and Sargent, 2002). 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 Cogley and Sargent (2005), Primiceri (2005), and Stock and Watson (2007) highlight the importance of jointly considering heteroscedasticity 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, recently, a growing literature has focused on TVP models with heteroscedastic errors, typically modeled via a stochastic volatility specification.

However, despite the fact that complex multivariate models have been proposed, in practice simpler time-varying models based on the autoregressive moving average (ARMA) methodology outperform forecasts (see for example Stock and Watson, 2007). Several studies have emerged combining ARMA terms for the conditional mean and stochastic volatility for the conditional variance. Moreover, multivariate extensions built on a time-varying vector autoregression form have been receiving much attention in the empirical macroeconomics literature (e.g. Primiceri, 2005). Therefore, time-varying ARMA models with stochastic volatility and extensions seem to be a state of the art technique for forecasting and policy analysis, especially in the macroeconomics literature (e.g. D’Agostino et al., 2013).

In this paper we propose time-varying ARMA models with stochastic volatility for inflation forecasting. The presence of time-varying parameters and stochastic volatility allows for structural instability and regime changes in the inflation behavior. We let the parameters to evolve over time in a stochastic way, meaning they depend on idiosyncratic innovations. The aim of the present paper is twofold. The first one is to present our specification for time-varying autoregressive models and to develop a classical maximum likelihood framework for such class of models. The second objective is to assess the ability of the proposed model to explain U.S. CPI inflation and to verify its forecasting performance in comparison to benchmark models.

Although flexible and theoretically appealing, the estimation of TVP ARMA models with stochastic volatility is not straightforward. Latent time-varying autoregressive coefficients require non-linear stability restrictions, which lead to an analytically intractable likelihood function. To the best of our knowledge, only Bayesian inference is available in the literature, frequently with high inefficiency factors. One of the major drawbacks to adopting Bayesian methods is that the required restrictions are usually satisfied via rejection sampling, leading to a high number of draws being discarded (see Koop and Potter, 2011; Chan et al., 2013). On the other hand, simulated maximum likelihood methods via importance sampling are fast and efficient procedures for
low dimensional nonlinear state models (see Koopman et al., 2015 for the efficiency results of simulated maximum likelihood methods). Differently from Markov Chain Monte Carlo methods, importance sampling estimates all parameters simultaneously, and the signal extraction is a natural byproduct. Therefore an improved estimation framework for time-varying ARMA models can be developed based on a classical maximum likelihood alternative.

This paper presents a novel specification for time-varying parameter models using a common factor structure for a large number of time-varying coefficients. The rationale is that a small number of components vary over time, inducing instability in all autoregressive coefficients, where the correlation structure between TVPs is mainly explained by the common component. The estimation procedure is a non-trivial combination of Kalman Filtering and Smoothing (KFS) methods and importance sampling techniques. We base our methodology on a multivariate extension of the numerically accelerated importance sampling (NAIS) of Koopman et al. (2015) combined with the Rao-Blackwellized Efficient Importance Sampling proposed by Moura and Turatti (2014) to incorporate linear time-varying states. In addition, we show that the importance density can also be efficiently evaluated by standard state-space methods. Therefore, the resulting procedure simplifies to a sequence of Kalman filters and least squares problems, yielding a computationally simple and numerically efficient methodology. In this paper, we mainly investigate univariate time-varying AR models with stochastic volatility, but the procedure can be applicable to other cases of interest in the literature, for example models with unobserved components (e.g. Stock and Watson, 2007), moving average terms (e.g. Chan, 2013) and multivariate extensions (e.g. Primiceri, 2005).

This paper’s second contribution involves an empirical application to modeling and forecasting inflation with time-varying autoregressive models. An increasing number of studies have found that the inflation process is subject to changes in the trend, persistence and the volatility. During the 1960s and 1970s, inflation volatility and persistence increased substantially (see Cecchetti et al., 2007). 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 (Chan, 2017). 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 volatility terms. The time-varying AR with stochastic volatility is able to accommodate these observed dynamics in the inflation process.

A forecasting study is performed to determine the forecasting ability of our time-varying autoregressive process. The model is applied to forecast C.P.I. inflation and we perform a point and density out-of-sample forecasting analysis in comparison to the univariate benchmark models. To preview some of our results, our time-varying
1.2. Time-varying autoregressive models with common factors and stochastic volatility

AR is shown to outperform all benchmark models both in point and density forecasting analysis, especially in medium and longer forecasting horizons. In contrast to model based on unobserved components, the model with time-varying parameters highlights the importance of random shocks in explaining the inflation process. For example, they indicate that energy price shocks may only have a transitory effect on inflation.

The structure of the paper is as follows. Section 2 formulates the time-varying autoregressive model with common factors and necessary notation. Section 3 provides a description of the necessary restriction to be imposed on the parameter space to ensure that the model and its likelihood function are well-defined. Section 4 presents our estimation procedure and our main methodological contributions, and section 5 investigates the finite sample properties of our simulated maximum likelihood estimator. In section 6 we carry out our application to modeling and forecasting US quarterly inflation.

1.2 Time-varying autoregressive models with common factors and stochastic volatility

The general framework we consider is the following time-varying parameter model with stochastic volatility,

\[ y_t = x_t' \phi_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad (1.1) \]

\[ \sigma_t^2 = \exp(h_t), \quad (1.2) \]

where \( y_t \) is a scalar, \( x_t = (C_t, y_{t-1}, \ldots, y_{t-p})' \) is the vector of observed exogenous variables, \( \phi_t = (\phi_{C,t}, \phi_{1,t}, \ldots, \phi_{p,t})' \) is the vector of latent time-varying parameters, and \( h_t \) is the log-volatility. \( C_t \) is a \( c \)-dimensional vector of exogenous regressors and \( \phi_{C,t} \) is the associated \( c \)-dimensional time-varying parameters. Note that we allow for \( p \)-autoregressive terms. Given the state vector \( \alpha_t = (\phi_t, h_t)' \), the time-varying parameters and the stochastic volatility are distributed according to a Gaussian vector autoregression process of order one,

\[ \alpha_{t+1} = d + T \alpha_t + \eta_t, \quad \eta_t \sim N(0, Q), \quad (1.3) \]

and \( \alpha_{t+1} \) is the \( l \times 1 \) state vector, where \( l = p + c + 1 \). The dynamics of the state vector are determined by the \( l \times 1 \) constant vector \( d \), the \( l \times l \) diagonal transition matrix \( T \) and the \( l \times l \) diagonal covariance matrix \( Q \). All elements of the disturbance vectors \( \eta_t \) and \( \varepsilon_t \) are serially and mutually uncorrelated, that is \( \mathbb{E}(\eta_t \varepsilon_{t}') = 0 \) for all \( t \). For the sake of simplicity the system parameters \( d, T, \) and \( Q \) are assumed to be time-invariant. Extensions for time variation in state equation are straightforward with the inclusion of additional state variables.

The model described by equations (1.1)–(1.3) may lead to a large number of non-linear time-varying parameters that need to be estimated, especially for high
CHAPTER 1. **Time-Varying Autoregressive Models with Common Factors: An Application to Inflation Forecasting**

autoregressive orders in which the necessary restrictions to ensure non-explosive paths for \( \phi_t \) increase the numerical complexity of the estimation procedure. In addition, it is likely that the time variability in the time-varying parameters presents commonalities. For example, Cogley and Sargent (2005) have found that the main time instability in their VAR model is related to only a few components. Therefore, we propose to assume that the dynamics of the time-varying autoregressive parameters are driving by a factor structure. The idea is to define the \( \phi_t \) vector as,

\[
\phi_t = \Phi^c + \Phi^f f_t \tag{1.4}
\]

\[
f_t = \varphi f_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Omega), \tag{1.5}
\]

where \( f_t \) is a \( r \times 1 \) vector, \( \phi_t \) and \( \Phi^c \) are \( p \times 1 \) vectors, \( \Phi^f \) is a \( p \times r \) vector, and \( \Omega \) is a \( r \times r \) covariance matrix of the factors. \( \varphi \) is the \( r \times r \) transition matrix of the factors, and it is imposed to be stationary. It is assumed that the factors \( f_t \) are of lower dimension than \( \phi_t \) and the factor loading matrix \( \Phi^f \) exploits the commonalities in the stochastic factors \( f_t \) to generate the time-varying parameters. The intuition is that only a small number of components vary over time, inducing instability in some coefficients while other parameters may remain unchanged. For identification purposes we impose the following restriction on the loading matrix \( \Phi^f \),

\[\text{(A) To identify both } f_t \text{ and the factor loading matrix } \Phi^f, \text{ } r \text{ rows of } \Phi^f \text{ must be fixed to form an identity matrix.}\]

We have chosen to partially restrict the loading matrix and leave the factor process unrestricted. This set of restrictions is sufficient to identify both the loading matrix and the factors (see Bai and Wang, 2015).

Equations (1.1)-(1.3) describe the specification of the time-varying parameter model with stochastic volatility. While this extension of the basic stochastic volatility model is readily motivated for macroeconomic variables, its estimation is challenging and prone to numerical difficulties due to the presence of stochastic volatility and conditional linear states. Given that the model in (1.1)-(1.3) delivers a nonlinear state-space representation, the likelihood function cannot be expressed in the form of the prediction error decomposition computed by the Kalman Filter. To tackle this nonlinear state-space estimation, we propose an estimation procedure based on an extension of the NAIS method of Koopman et al. (2015) together with the Rao-Blackwell step proposed by Moura and Turatti (2014) to efficiently construct an importance sampling approximation to the likelihood function. The resulting estimation procedure is applicable to a range of models with linear and non-linear time-varying parameters.

Additionally, the evaluation of the likelihood function depends upon necessary restrictions to be imposed on the time-varying parameters. 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. The next section shows a necessary parametrization to ensure a well-defined AR process for $y_t$.

### 1.3 Model Restrictions

Time-varying parameter models usually require restrictions to be imposed on the parameter space. Particularly, time-varying autoregressive terms must have their roots within the unit circle for all time $t$. To satisfy these restrictions all the time, a monotonic transformation has to be enforced on the necessary time-varying parameters. The restrictions considered here were analyzed in Prado and West (2010), and a Bayesian Markov Chain Monte Carlo framework was developed by Koop and Potter (2011). Let $\alpha_{i,t}$ be the unrestricted state vector, we impose the link function $\psi(.)$ to ensure that the resulting vector of parameters $\phi_t$ satisfy stability restrictions at all times,

$$
\phi_{i,t} = \psi(\alpha_{i,t}), \quad \text{for } i = 1, \ldots, p. \tag{1.6}
$$

The link function $\psi(.)$ has to be time-invariant, continuous, and strictly monotonic to properly define a smooth stochastic process for $\phi_t$.

We continue by defining the link function that satisfies these conditions. First, we introduce some notation. Let $\rho_t = (\rho_{1,t}, \ldots, \rho_{p,t})'$ be the vector of partial autocorrelation for a given vector of AR coefficients $\phi_t$, and $z_t = (z_{1,t}, \ldots, z_{p,t})'$ 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 AR(p) model can be imposed on equation (1.1) by restricting all the partial autocorrelations $\rho_{i,t}$, $\forall i = 1, \ldots, p$, to be within the unit circle for each time $t$.  

**Definition 1.** Eq. (1.1) has roots within the unit circle if $\phi_t \in S^p \forall t$, where $S^p$ is the hyperplane in which all the roots $z_i$ are within the unit circle, i.e. $|z_{i,t}| < 1 \forall t$. Moreover, $\phi_t \in S^p \forall t$ and $|z_{i,t}| < 1 \forall t$ if and only if $|\rho_{i,t}| < 1 \forall t$.

Definition 1 extends the results of Barndorff-Nielsen and Schou (1973) and Monahan (1984) to models with time-varying parameters. Therefore, we can dismiss explosive paths for $y_t$ by modeling its partial autocorrelations and restricting their space to be within the unit circle. To achieve this task, a monotonic function $\Upsilon(.)$ defined in $(-1, 1)$ is required. Several functions are available, but in this article we employ the Fisher transformation, $\rho_{i,t} = \tanh(\alpha_{i,t})$. Therefore, our nonlinear state variable, $\alpha_{i,t}$ yields the partial autocorrelations of $y_t$.

Finally, a transformation $\psi(.)$ is necessary, which maps the partial autocorrelations, $\rho_{i,t}$, into autoregressive coefficients, $\phi_{i,t}$. Hence, we implement the Durbin-Levinson algorithm to map $\rho_{i,t}$ into $\phi_{i,t}$ for every point in time.

**Definition 2.** The required transformation that maps the PACs into the ARs is given
by the following algorithm,
\[
\phi_t^{i,k} = \phi_t^{i,k-1} - \rho_{k,t} \phi_t^{k-1,i-1},
\]
\[
i = 1, \ldots, k-1, \quad k = 2, \ldots, p
\]
where \(\phi_t^{k,k} = \rho_{k,t}\) and \(\phi_t^{1,1} = \rho_{1,t}\). The transformation \(\psi(.)\) is defined by the last recursion of the Durbin-Levinson algorithm. Thus, the composite function \(g(.) = \psi(Y(.))\) maps the unrestricted time-varying parameters \(\alpha_{i,t} \in (-\infty, \infty)\) into non-explosive autoregressive coefficients \(\phi_{i,t} \in \mathbb{S}^p\) for every time \(t\).

### 1.4 Estimation procedure

This section develops a numerical and computationally efficient procedure to evaluate the log-likelihood function for time-varying AR models with stochastic volatility. The estimation procedure is a combination of Kalman Filtering and Smoothing methods and importance sampling techniques. Moreover, variance reduction through Rao-Blackwellization methods are employed to increase the numerical efficiency of the algorithm even further. Finally, unlike particle filters, the log-likelihood estimate is a continuous function of the parameter space, therefore, gradient-based optimization methods can be used to find its maximum (see Creal, 2012; for problems related to particle filters). We show that a major part of the likelihood evaluation procedure can be done by fast numerical integration rather than Monte Carlo integration, and the sampling from the proposal density can be based on standard state-space methods such as the Kalman filter and smoother. The method simplifies to a sequence of low-dimensional weighted least square problems.

The likelihood function for the model discussed in the paper is given by the integral over the joint density of observables \(y_t\) and latent states \(\alpha_t = (\phi_{C,t}, \alpha_{1,t}, \ldots, \alpha_{p,t}, h_t)^\prime\),

\[
L(\gamma, y) = \int p(y_t, \alpha_t; \gamma) d\alpha_t,
\]
\[
L(\gamma, y) = \int \prod_{t=p+1}^{T} p(y_t|\alpha_t; \gamma) p(\alpha_t|\alpha_{t-1}; \gamma) d\alpha_{p+1} \ldots d\alpha_T,
\]
where \(\gamma\) is the parameter vector. The likelihood function equation (1.7) is a \(l \times T\) analytically intractable integral. Therefore, the integrals in the expression for the likelihood function have to be approximated numerically via simulation methods. We aim to evaluate the likelihood function by importance sampling techniques.

However, note that it is possible to exploit the structure of the state space model to take advantage of the conditional linearity of it, thus reducing the dimension of the likelihood function and increasing efficiency by integrating out the unrestricted time-varying parameters via the KFS. For a given set of the autoregressive coefficients and the stochastic volatility, the model is a traditional linear and Gaussian
state space model. Therefore, it is possible to analytically integrate out the conditionally linear parameters contained in $\phi_{C,t}$ using the Kalman filter, thus reducing the Monte Carlo (MC) variation of the likelihood estimator. This procedure is known as Rao-Blackwellization (marginalization), as it is an implication of the Rao-Blackwell Theorem. Therefore, given the $p + 1$-dimensional nonlinear state vector $\tilde{\alpha}_t = (a_1, \ldots, a_{p,t}, h_t)'$, the linear states $\phi_{C,t}$ can be integrated out of (1.8) yielding,

$$L(\gamma, y) = \int \prod_{t=p+1}^{T} p^*(y_t|\tilde{\alpha}_t; \gamma) p(\tilde{\alpha}_t|\tilde{\alpha}_{t-1}; \gamma) d\tilde{\alpha}_{p+1} \ldots d\tilde{\alpha}_T$$

(1.9)

and,

$$p^*(y_t|\tilde{\alpha}_t; \gamma) = p(v_t|\tilde{\alpha}_t; \gamma) \equiv N(0, F_t),$$

(1.10)

where $p^*(y_t|\tilde{\alpha}_t, \gamma)$ is the prediction error density, $v_t$ and $F_t$ are the prediction error and its associated variance delivered by Kalman filter given the nonlinear state vector $\tilde{\alpha}_t$. It is worth mentioning that $(p+1) < l$, thus a dimension reduction can be obtained while increasing computational and statistical efficiency. The intuition behind this result is that we replace the importance sampling (IS) estimator for $\phi_{C,t}$ by its KFS estimator. As the KFS is an analytical filter its MC variance is zero, thus always smaller than the IS filter. For the models in this paper $v_t$ and $F_t$ can be expressed as,

$$v_{t|t-1} = y_t - x_t'\phi_{t|t-1},$$

$$F_t = C_t'P_{t|t-1}C_t + \sigma_t^2,$$

(1.11)

where $P_{t|t-1}$ is the predicted variance of the linear time-varying parameters $\phi_{C,t}$. Note that the prediction error $v_{t|t-1}$ is always a stationary variable when the model is correct specified. Therefore, a unit root may be contained in the time-varying parameters and the asymptotical properties of the likelihood function still holds. However, it is still necessary to approximate the lower dimensional likelihood integral (1.9), therefore we construct a Gaussian importance sampler based on the approximating linear model of Durbin and Koopman (1997) and the NAIS method of Koopman et al. (2015).

The Gaussian multivariate importance sampler $g(\tilde{\alpha}|y; \varphi)$ can be represented as,

$$g(\tilde{\alpha}|y; \varphi) = \frac{g(y|\tilde{\alpha}; \varphi)g(\tilde{\alpha}; \varphi)}{g(y; \varphi)},$$

(1.13)

where $g(\tilde{\alpha}; \varphi)$, $g(y; \varphi)$ are all Gaussian densities. $g(y; \varphi)$ is a normalizing constant and $\varphi$ is the vector of parameters of the importance density $g(\tilde{\alpha}|y; \varphi)$. It is implied that $g(\tilde{\alpha}; \varphi) \equiv p(\tilde{\alpha}; \gamma)$. Using this importance sampler it is possible to rewrite the likelihood function given some regularity conditions (see Geweke, 1989).

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

(1.14)
where the transition equation remains the same as the original model. Therefore, it can be shown that,

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

(1.15)

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

(1.16)

where $g(y; \varphi)$ is the likelihood of the importance density model and the importance weights are defined as,

$$\omega(\gamma, y) = \frac{p^*(y|\tilde{\alpha})}{g(y|\tilde{\alpha}, \varphi)},$$

(1.17)

evaluated at $S$ independent trajectories of the non-linear signal $\tilde{\alpha}$. The likelihood estimator is given by,

$$\hat{L}(\gamma, y) = g(y; \varphi) \times \tilde{\omega}(\gamma, y), \quad \tilde{\omega}(\gamma, y) = \frac{1}{S} \sum_{i=1}^{S} \omega(\tilde{\alpha}^{(i)}, y).$$

(1.18)

Finally, naive choices of the importance parameters $\varphi$ ignore valuable information about the underlying latent process carried by the observables available at time $t$, and can be highly inefficient (Richard and Zhang, 2007). We apply the NAIS method of Koopman et al. (2015), which selects $\varphi$ optimally yielding a fully adapted likelihood estimator, hence minimizing its MC variance.

### 1.4.1 Multivariate Numerically Accelerated Importance Sampling

The success of an importance sampling technique relies on matching the proposal density and the target integrand. The key element in the NAIS procedure is the representation of $g(y_t|\tilde{\alpha}_t, \varphi)$ as a linear and Gaussian state space density,

$$g(y_t|\tilde{\alpha}_t) = \exp(a_t + b'_t \tilde{\alpha}_t - \frac{1}{2} \tilde{\alpha}'_t C_t \tilde{\alpha}_t),$$

(1.19)

where $a_t$ is an integrating constant, and the importance parameters $b_t \in \mathbb{R}^{p+1}$ and $C_t \in \mathbb{R}^{p+1 \times p+1}$ for $t = 1, \ldots, T$ are a function of $y_t$ and $\gamma$. Therefore, we have to select $\{b_t\}_{t=1}^T$ and $\{C_t\}_{t=1}^T$ to construct the importance density. Most choices yield consistent estimators (see Geweke, 1989 for conditions of consistency). Here we aim to select $\varphi = \{\{b_t\}_{t=1}^T, \{C_t\}_{t=1}^T\}$ to deliver an estimator that minimizes the MC variance of $\hat{L}(\gamma, y)$.

For a given set of importance parameters, $\varphi_t = (b_t, C_t)$, the proposal density $g(y_t|\tilde{\alpha}_t, \varphi_t)$ can be computed via the artificial observations $y^*_t = C_t^{-1} b_t$ with measurement equation,

$$y^*_t = \tilde{\alpha}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, C_t^{-1}),$$

(1.20)

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

$$\log g(y^*_t|\tilde{\alpha}_t, \varphi_t) = \frac{-(p+1)}{2} \log(2\pi) + \frac{1}{2} \log|C_t| - \frac{1}{2} [(C_t^{-1} b_t - \tilde{\alpha}_t)' C_t^{-1} b_t - \tilde{\alpha}_t]\|,$$

$$= a_t + b'_t \tilde{\alpha}_t - \frac{1}{2} \tilde{\alpha}'_t C_t \tilde{\alpha}_t,$$

(1.22)
where \( a_t \) is an integrating constant and collects terms not related to \( \tilde{\alpha}_t \). The linear state-space representation (1.20) implies that we can use the Kalman filter and smoothing techniques to evaluate and sample from \( g(\tilde{\alpha}_t | y_t, \varphi_t) \), thus resulting in a computationally efficient and simple algorithm. This contrasts to Richard and Zhang (2007)’s efficient importance sampling in which the sampler has to be decomposed into a kernel and an integrating constant that must be followed and transferred in time. On the other hand, NAIS’ artificial representation of the importance sampler means that the integrating constant does not play an active role in the simulation, simplifying the algorithm. Additionally, the smoothing density \( g(\tilde{\alpha}_t | y_t, \varphi_t) \) is available analytically and is given by,

\[
g(\tilde{\alpha}_t | y_t, \varphi_t) = \frac{1}{\sqrt{(2\pi)^{p+1}|V_t|}} \exp\left(-\frac{1}{2} (\tilde{\alpha}_t - \hat{\alpha}_t)'(V_t)^{-1}(\tilde{\alpha}_t - \hat{\alpha}_t)\right),
\]

(1.23)

where \( \hat{\alpha}_t \) and \( V_t \) are the mean and the variance of the non-linear signal vector \( \tilde{\alpha}_t \), and are obtained by the modified backward-forward (BF) sampling algorithm of Koopman et al. (2015) or the KFS applied to the linear state-space (1.20) for a given set of importance parameters \( \varphi = ([b_t]_{t=1}^{T}, [C_t]_{t=1}^{T}) \).

Moreover, the importance parameters \( \varphi = ([\varphi_t]_{t=1}^{T}) \) are efficiently selected, approximating the importance sampler and the likelihood integrand, thus minimizing 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,

\[
\varphi_t = \arg\min_{\varphi_t} \int \lambda^2(y_t, \tilde{\alpha}_t|\varphi_t) \omega(\tilde{\alpha}_t, y_t|\varphi_t) g(\tilde{\alpha}_t | y_t, \varphi_t) \, d\tilde{\alpha}_t,
\]

(1.24)

where \( \lambda(.) \) is defined as,

\[
\lambda(y_t, \tilde{\alpha}_t|\varphi_t) = \log p^*(y_t|\tilde{\alpha}_t) - \log g(y_t^*|\tilde{\alpha}_t, \varphi_t).
\]

(1.25)

According to Richard and Zhang (2007) this operational criteria approximates the problem of minimizing the MC sampling variance of the importance weights \( \omega(\gamma, y) \) in its full support.\(^1\) We note that whereas the likelihood integrand (1.9) is very high-dimensional in most cases of interest, the integral (1.24) is of (much) lower dimension, therefore an almost exact solution can be obtained by numerical methods.\(^2\) In constrast to Richard and Zhang (2007)’s Monte Carlo approximation for integral (1.24), we follow Koopman et al. (2015) and numerically implement Gauss-Hermite (GH) quadrature methods to solve (1.24) and obtain the importance parameters \( \varphi_t = (b_t, C_t) \).

However, note that a main complication arises as the importance sampler \( g(\tilde{\alpha}_t | y_t, \varphi_t) \) itself depends upon the importance parameters \( \varphi_t \). As in Richard and Zhang (2007)

\(^1\)See Richard and Zhang (2007), for more details on the EIS criteria.

\(^2\)The dimension of (1.24) is the dimension of \( \tilde{\alpha}_t \), which is of much lower dimension compared to the likelihood function.
this can be resolved by a standard fixed point algorithm using intermediate importance samplers until convergence of \( \varphi = (|\varphi_t|)^T \). Therefore, the minimization problem (1.24) can be rewritten using the intermediate sampler \( g(\tilde{a}_t|y_t, \varphi_t^{(k)}) \),

\[
\varphi_t^{(k+1)} = \text{argmin}_{\varphi_t^{(k)}} \int \lambda^2(y_t, \tilde{a}_t|\varphi_t^{(k)}) \omega(\tilde{a}_t, y_t|\varphi_t^{(k)}) g(\tilde{a}_t|y_t, \varphi_t^{(k)}) \, d\tilde{a}_t.
\]  

(1.26)

To initiate the NAIS procedure an initial condition \( \varphi^{(0)} \) is required, which will be sequentially updated through iterations on (1.26) until a convergence criteria is satisfied.

Let \( \{z_i\}_{i=1}^{M_p+1} \) be the set of abscissae after all combinations of \( M \) predefined GH nodes and \( h(z_i) \) be the respective Gauss-Hermite weights. Thus, the minimization (1.24) is implemented as,

\[
\text{argmin}_{\varphi_t^{(k+1)}} \sum_{i=1}^{M_p+1} w_{i,t} \Phi(y_t, Z_{i,t}^{(k)}, \varphi_t^{(k)}),
\]

(1.27)

where,

\[
\Phi(y_t, \tilde{a}_t, \varphi_t^{(k)}) = \lambda^2(y_t, \tilde{a}_t|\varphi_t^{(k)}) \omega(\tilde{a}_t, y_t|\varphi_t^{(k)}).
\]

(1.28)

The weights are given by \( w_{i,t} = h(z_i) \exp(1/2 \, z_i^2) \) and \( Z_{i,t}^{(k)} = \hat{a}_t^{(k)} + V_{t,0.5} Z_i, \) for \( i = 1, \ldots, M_p+1 \), where \( V_{t,0.5} \) is the root matrix computed via the Cholesky decomposition of \( V_{t}^{(k)} \). For this implementation, the distribution of the combined nodes is given by \( g(Z_{i,t}^{(k)}|y_t, \varphi_t^{(k)}) \propto \exp(-1/2 \, z_i^t Z_i) \), as they are associated with the multivariate standard normal distribution.

Finally, the minimization (1.26) can be written as a weighted least squares problem applied to \( M_{p+1} \) observations for the auxiliary regression,

\[
\log p^*(y_t|Z_{i,t}^{(k)}) = \text{constant} + \kappa^T Z_{i,t}^{(k)} - \frac{1}{2} \xi^T \text{vech}(Z_{i,t}^{(k)} Z_{i,t}^{(k)}) + \text{error},
\]

(1.29)

where \( \text{vech}(.) \) stacks the elements of a symmetric matrix into a vector. The weighted least square estimates for \( \kappa \) and \( \xi \) are the new estimates of \( b_{t}^{(k+1)} \) and \( C_{t}^{(k+1)} \) respectively for each time \( t = 1 \ldots T \). The weights for the auxiliary regression are given by \( w_{i,t} = w(Z_{i,t}^{(k)}|y_t, \varphi_t^{(k)}), g(Z_{i,t}^{(k)}|y_t, \varphi_t^{(k)}) \). Using the new estimates \( b_{t}^{(k+1)} \) and \( C_{t}^{(k+1)} \), we can obtain the smoothed means \( \hat{a}_t^{(k+1)} \) and variance \( V_t^{(k+1)} \) for \( g(\tilde{a}_t|y_t, \varphi_t^{(k+1)}) \) for all \( t = 1, \ldots T \). This procedure is iterated until convergence is achieved, usually only a small (< 15) number is necessary.\(^3\)

The complete multivariate NAIS algorithm reads as follows,

1. Initialize the procedure with starting values for \( \varphi_t = (b_t, C_t) \) for each time \( t = 1 \ldots T \). The choice of the initial values is usually not necessary for convergence.

   Generally, a vector of zeros for \( b_t \) and an identity matrix for \( C_t \) suffice.

\(^3\)Convergence of the algorithm is not crucial as only the initial iterations typically generate substantial reductions in the variance of the likelihood estimate (DeJong et al., 2013)
2. Compute smoothed mean $\hat{\alpha}_t^{(k)}$ and variance $V_t^{(k)}$ for each time $t = 1 \ldots T$ based on the current values of $\varphi = \{[b_t]^T_{t=1}, [C_t]^T_{t=1}\}$ and the state-space representation (1.20) via the BF algorithm of Koopman et al. (2015) or the KFS.

3. For each time $t = 1 \ldots T$ compute the set of abscissae, $\{Z_{i,t}\}_{i=1}^{M_p+1}$, and evaluate the density $p^*(y_t|\tilde{\alpha}_t, \gamma)$ by the prediction error decomposition as calculated by the Kalman Filter. Minimize equation (1.26) and obtain a new set of $\varphi_t$.

4. Compute a criteria for convergence. Run steps 2 and 3 until convergence is achieved or preset a sufficient number of iterations.

5. Given the final set of $\varphi = \{[b_t]^T_{t=1}, [C_t]^T_{t=1}\}$, sample S independent trajectories for the signal $\tilde{\alpha}$ via the BF algorithm or the simulation smoother of Durbin and Koopman (2002).

6. Compute the likelihood of the approximating linear state-space model $g(y; \varphi)$ by applying the KFS on (1.20) and estimate the likelihood equation (1.18).

The NAIS method outlined in the above steps yields a well-defined procedure for estimating the log-likelihood function and sampling from the proposal density in most cases. However, a pathology we may encounter is that the matrix $C_t$ might become negative definite, implying that the approximating linear model (1.20) is not defined. We believe this is caused by the importance sampler $g(\alpha|y)$ concentrating draws in a locally convex region of $\log p^*(y|\alpha)$ (Richard and Zhang, 2007). This usually does not hinder the application of NAIS due to the fact that the Backward-Forward sampling scheme is a well-defined process even if some of the matrices $C_t$ are negative definite. Moreover, a modified Kalman Filter simulation smoother is available (see Jungbacker and Koopman 2007). We estimate that only 5% of the applications need intervention as a result of problems in the covariance matrix. Efficient, fast solutions for such occasional pathologies consist of changing the initial condition of the optimization thus avoiding the sampler to concentrate on a tail area.

1.5 Monte Carlo Study

The simulated maximum likelihood method developed in the previous section is based on a prediction error decomposition, therefore the usual asymptotical properties apply. However, the likelihood function is still subject to simulation error and its finite sample behavior is unknown. In this section, we perform a Monte Carlo experiment to investigate the small sample properties of our simulated maximum likelihood method on different data generating process (DGP) scenarios. In addition, we compare the NAIS estimator with a Rao-Blackwellized particle filter, which is an

---

4 A negative covariance matrix is far more likely to occur for a parameter set far from the DGP. We have not encountered problems in the vicinity of the real parameters.
unbiased and consistent estimator for the likelihood (see Creal, 2012 for details on the particle filter). This second study aims to show that the NAIS procedure yields a numerically more efficient procedure. Finally, we investigate whether a central limit theorem is applicable to the likelihood estimator. In a seminal work Geweke (1989) defines the conditions for convergence of a Monte Carlo estimator and argued that importance sampling techniques are only applicable if the variance of the importance weights is finite. We follow Koopman (2009) and apply extreme value theory (EVT) to study tail behavior and make inferences on the moments of the distribution through the well-known likelihood ratio test.

1.5.1 Study Design

In this Monte Carlo study we are interested in verifying the finite sample properties of our SML procedure under different specifications of the time-varying coefficients. To achieve this task we investigate two models. The first one we consider is the following basic time-varying autoregressive model for $t = 2, \ldots, T$,

$$y_t = \phi_{0,t} + \phi_{1,t} y_{t-1} + \exp\left(\frac{h_t}{2}\right) \varepsilon_t, \quad \varepsilon_t \sim N(0,1). \quad (1.30)$$

The state variable dynamics are determined by the subsequent Markovian processes,

$$h_{t+1} = \gamma_{h} h_t + \sigma_h \eta_{h,t}, \quad \eta_{h,t} \sim N(0,1), \quad (1.31)$$

$$\alpha_{t+1} = \gamma_{1} \alpha_t + \sigma_1 \eta_{1,t}, \quad \eta_{1,t} \sim N(0,1), \quad (1.32)$$

$$\phi_{0,t+1} = \gamma_{0} \phi_{0,t} + \sigma_0 \eta_{0,t}, \quad \eta_{0,t} \sim N(0,1), \quad (1.33)$$

with $\phi_{1,t+1} = \tanh(\alpha_{t+1})$. To reduce the number of static parameters to be estimated we fix both the unconditional mean of the autoregressive coefficient and the unconditional mean of the log-volatility to 0. We examine the finite sample properties of the estimation method in 3 cases: (1) high persistency in autoregressive coefficient and low in the intercept; (2) low persistency in autoregressive coefficient and high in intercept; (3) high persistency for both parameters. Therefore, the DGP is set as,

<table>
<thead>
<tr>
<th>DGP</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.80</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>2</td>
<td>0.98</td>
<td>0.80</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

For the variances we fix $\sigma_0 = \sigma_1 = \sigma_h = 0.08$ in all cases.

The design of the experiment allows us to verify whether a substantial difference in the persistence of the time-varying coefficients affects the resulting estimation and how the log-likelihood identifies the model parameters under different degrees of time variation.
The second model is intended to verify how the log-likelihood function identifies the model parameters when the time-varying autoregressive coefficients are driven by a factor structure. In this case we specify a AR(2) with 1 factor driving the autoregressive coefficients for \( t = 3, \ldots, T \),

\[
y_t = \phi_{0,t} + \phi_{1,t} y_{t-1} + \phi_{2,t} y_{t-2} + \exp(\frac{h_t}{2}) \varepsilon_t, \quad \varepsilon_t \sim N(0,1) \quad (1.34)
\]

\[
\phi_t = \phi(\Upsilon(\Phi_0 + \Phi_1 f_t)) \quad (1.35)
\]

\[
f_{t+1} = \varphi f_t + \sigma_f \eta_{1,t}, \quad \eta_{1,t} \sim N(0,1), \quad (1.36)
\]

\[
h_{t+1} = \gamma h_t + \sigma_h \eta_{h,t}, \quad \eta_{h,t} \sim N(0,1), \quad (1.37)
\]

\[
\phi_{0,t+1} = \gamma_0 \phi_{0,t} + \sigma_0 \eta_{0,t}, \quad \eta_{0,t} \sim N(0,1), \quad (1.38)
\]

where \( \phi(\Upsilon(.)) \) are the necessary restrictions to ensure that the time-varying autoregressive coefficients are inside the unit circle. We fix \( \gamma_0 = \gamma_h = \varphi = 0.98 \), and \( \sigma_f = \sigma_h = \sigma_0 = 0.08 \). For the loading vectors we set \( \Phi_c = (0.2, 0.2)' \) and \( \Phi_f = (1, 0.5)' \).

For each model two experiments are considered. The first one presents mean and standard deviations for 100 series generated with sample sizes \( n = 100, n = 250, n = 500, \) and \( n = 1000 \). In the SML estimation procedure we use 10 GH nodes with 400 independent trajectories for the signal drawn to compute the Monte Carlo estimate for the log-likelihood function. The second experiment presents mean and sample standard deviations of the estimates in each scenario. In the second experiment, a single series with size \( n = 1000 \) is generated for each model and scenario and we compute the ML estimates, as well as numerical and average statistical standard errors using 30 different random seeds. The objective of this experiment is to verify how reliable the estimate of the standard errors are. For the first model the results are shown in tables 1.2 and 1.3. Tables 1.4 and 1.5 present results for the second model.

1.5.2 Results

The first experiment results are presented in table 1.2. We note that for small sample sizes \( n = 100 \), all the parameters have large standard deviations, but for \( n = 500 \) and \( n = 1000 \) they become much smaller. In addition, the parameters on a time-varying coefficient are better estimated when its degree of persistence is large. For example, the parameter \( \gamma_0 \) and the parameter \( \gamma_1 \) in DGP 1 and DGP 2 respectively are poorly estimated for \( n = 500 \), while the other parameters are substantially closer to their true values. On the other hand, in case 3 all parameters are very similar to the DGP values for \( n = 500 \). Nevertheless, the parameters seem to exhibit convergence to the DGP as the sample size increases, and at \( n = 1000 \) the parameter estimates are closer to the real ones. We conclude that the likelihood function displays more curvature for greater levels of persistence of the time-varying parameters, and low degrees of time-variation may imply a more flat likelihood function and slow convergence.

In the second experiment we aim to obtain numerical and asymptotical standard errors and to verify their precision in identifying the model parameters. The results
Table 1.2: Finite-Sample properties of the SML estimator

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP 1</th>
<th>DGP 2</th>
<th>DGP 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_1)</td>
<td>0.080</td>
<td>0.133</td>
<td>0.080</td>
</tr>
<tr>
<td>(\gamma_1)</td>
<td>0.980</td>
<td>0.477</td>
<td>0.980</td>
</tr>
<tr>
<td>(\sigma_h)</td>
<td>0.080</td>
<td>0.205</td>
<td>0.080</td>
</tr>
<tr>
<td>(\gamma_h)</td>
<td>0.980</td>
<td>0.718</td>
<td>0.980</td>
</tr>
<tr>
<td>(\sigma_0)</td>
<td>0.080</td>
<td>0.118</td>
<td>0.080</td>
</tr>
<tr>
<td>(\gamma_0)</td>
<td>0.980</td>
<td>0.892</td>
<td>0.980</td>
</tr>
</tbody>
</table>

The table presents the DGP parameters, followed by sample average of estimated parameters from 100 series. Values in parentheses are sample standard deviations of the estimates.

Table 1.3 shows the results of the first experiment for the second model. The objective of this study is to show that the NAIS procedure can identify the model parameters under a factor structure for the autoregressive coefficients. As before, the sample sizes are \(n = 100\), \(n = 250\), \(n = 500\), and \(n = 1000\). Results indicate that the average of the estimates tend to converge to the DGP as the sample size increases, although some variation is still observed at \(n = 1000\). However, the average of the estimates for the factor specification is very close to their real values.

Table 1.4 shows the results of the second experiment for the second model. We observe that when the persistence of the time-varying parameter is low, the statistical standard errors are large compared to their numerical ones. This indicates a low curvature of the log-likelihood function, although the parameters are still statistically significant. In the other cases, the asymptotical standard errors are small and the estimation is precise. Furthermore, the numerical standard errors are very small in all DGPs, highlighting that our multivariate NAIS procedure is very accurate and subject to low MC variation.

Table 1.5 shows the results of the second experiment for the second model. We observe that the asymptotical standard errors are small compared to the average of the estimates, thus the estimation is precise. Moreover, numerical standard errors are very small
Table 1.3: Finite-Sample properties of the standard errors

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP 1</th>
<th></th>
<th>DGP 2</th>
<th></th>
<th>DGP 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1$</td>
<td>0.080</td>
<td>0.078 (0.032) [0.001]</td>
<td>0.080</td>
<td>0.080 (0.024) [0.008]</td>
<td>0.080</td>
<td>0.081 (0.044) [0.000]</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.980</td>
<td>0.973 (0.013) [0.001]</td>
<td>0.800</td>
<td>0.742 (0.176) [0.025]</td>
<td>0.980</td>
<td>0.971 (0.009) [0.000]</td>
</tr>
<tr>
<td>$\sigma_h$</td>
<td>0.080</td>
<td>0.087 (0.033) [0.002]</td>
<td>0.080</td>
<td>0.091 (0.025) [0.001]</td>
<td>0.080</td>
<td>0.093 (0.028) [0.001]</td>
</tr>
<tr>
<td>$\gamma_h$</td>
<td>0.980</td>
<td>0.964 (0.009) [0.001]</td>
<td>0.980</td>
<td>0.970 (0.008) [0.001]</td>
<td>0.980</td>
<td>0.963 (0.007) [0.002]</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.080</td>
<td>0.130 (0.145) [0.001]</td>
<td>0.080</td>
<td>0.083 (0.024) [0.001]</td>
<td>0.080</td>
<td>0.081 (0.055) [0.003]</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0.800</td>
<td>0.662 (0.085) [0.018]</td>
<td>0.980</td>
<td>0.974 (0.010) [0.001]</td>
<td>0.980</td>
<td>0.976 (0.011) [0.002]</td>
</tr>
</tbody>
</table>

The table presents the DGP and sample average of the standard errors using 30 different random seeds for $n = 1000$. Values in parentheses are the average of the asymptotical standard errors. Values in brackets are numerical standard errors.

Table 1.4: Finite-Sample properties of the SML estimator

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP $n = 100$</th>
<th></th>
<th>DGP $n = 250$</th>
<th></th>
<th>DGP $n = 500$</th>
<th></th>
<th>DGP $n = 1000$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_h$</td>
<td>0.080</td>
<td>0.161 (0.122)</td>
<td>0.116 (0.087)</td>
<td>0.096 (0.058)</td>
<td>0.086 (0.036)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_h$</td>
<td>0.980</td>
<td>0.818 (0.174)</td>
<td>0.906 (0.136)</td>
<td>0.947 (0.116)</td>
<td>0.973 (0.030)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.080</td>
<td>0.173 (0.135)</td>
<td>0.161 (0.135)</td>
<td>0.157 (0.133)</td>
<td>0.159 (0.112)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0.980</td>
<td>0.754 (0.156)</td>
<td>0.756 (0.151)</td>
<td>0.799 (0.158)</td>
<td>0.919 (0.147)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Phi_{c,1,1}$</td>
<td>0.200</td>
<td>0.094 (0.287)</td>
<td>0.125 (0.228)</td>
<td>0.169 (0.190)</td>
<td>0.192 (0.151)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Phi_{c,2,1}$</td>
<td>0.200</td>
<td>0.153 (0.162)</td>
<td>0.160 (0.128)</td>
<td>0.187 (0.108)</td>
<td>0.203 (0.084)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Phi_{f,2,1}$</td>
<td>0.500</td>
<td>0.341 (0.501)</td>
<td>0.448 (0.438)</td>
<td>0.537 (0.281)</td>
<td>0.525 (0.152)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\varphi$</td>
<td>0.980</td>
<td>0.725 (0.169)</td>
<td>0.840 (0.169)</td>
<td>0.937 (0.111)</td>
<td>0.966 (0.024)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table presents the DGP and sample average of the standard errors of estimates. Values in parentheses are the sample standard errors of the estimates.

compared to the statistical ones. Thus, again we conclude that NAIS is precise and subject to low variation.

Table 1.5: Finite-Sample properties of the standard errors

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_h$</td>
<td>0.080 0.086 (0.032) [0.003]</td>
</tr>
<tr>
<td>$\gamma_h$</td>
<td>0.980 0.973 (0.023) [0.002]</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.080 0.159 (0.063) [0.012]</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>0.980 0.919 (0.119) [0.022]</td>
</tr>
<tr>
<td>$\Phi_{c,1,1}$</td>
<td>0.200</td>
</tr>
<tr>
<td>$\Phi_{c,2,1}$</td>
<td>0.200</td>
</tr>
<tr>
<td>$\Phi_{f,2,1}$</td>
<td>0.500</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>0.980 0.966 (0.015) [0.009]</td>
</tr>
</tbody>
</table>

The table presents the DGP and sample average of the standard errors using 30 different random seeds for $n = 1000$. Values in parentheses are the average of the asymptotical standard errors. Values in brackets are numerical standard errors.
1.5.3 Testing the Importance Sampler

The final set of experiments are related to testing the quality of the NAIS importance sampler in estimating the likelihood function. For this section we only focus on the first model and the third DGP. In the first study we want to compare the NAIS estimator for the log-likelihood with a Rao-Blackwellized particle filter (RB-PF). The experiment was based on 100 series, each with 30 different likelihood evaluations using different sets of random numbers. The number of particles were selected as a grid of values between 400 and 50000. In figure (1.3) (in appendix), we present the standard deviations across the 100 realizations of the process. The multivariate RB-NAIS yields log-likelihood estimates more efficiently than the RB-PF even for a very high number of particles. For example, when using 50000 particles NAIS is still twice more efficient than the PF. Remarkably NAIS with 400 draws is more efficient than the PF with 10000 draws. Figure (1.3) (in appendix) displays the computational time in seconds per likelihood evaluation and the standard deviation of the likelihood estimate. We note that given a sufficient number of particles NAIS is twice more efficient than the particle filter for the same computational effort. These results highlight the computational and numerical efficiency of NAIS.

The second experiment investigates the conditions of convergence of the NAIS likelihood estimator to the true likelihood function. Although, it is still unknown if these conditions need to be met in practice, a failure may theoretically imply convergence issues, as standard asymptotic theory is not applicable. Geweke (1989) argued that a central limit theorem can be claimed to importance sampling methods only if the variance of the importance weights is known to be finite. On the other hand, Richard and Zhang (2007) suggested that a theoretical unbounded variance is not an issue for empirical purposes as the variance will always be finite in practice, especially if a single set of random numbers is used to evaluate the log-likelihood function. Moreover, they provided empirical illustrations where the variance remains theoretically infinite though the actual probability of a variance explosion has been eliminated. However, it might be interesting to study the behavior of the weights in a time-varying parameter model to determine if there is an indication of possible convergence issues. Thus, we look for evidence of a bounded variance of the importance weights.

The experiment is based on the Extreme Value Theory, which makes inferences of the moments of a given distribution through its tail behavior. Particular, EVT implies that we can obtain information about the variance of the weights by studying the behavior of the right tail of the distribution. More information about EVT and diagnostics for the variance of the weights can be found in the supplementary appendix. We follow the approach of Koopman et al. (2009) and we apply the likelihood-ratio test using the ML estimation results. We generate $M = 100000$ importance weights.

---

5The log-likelihood estimator via particle filtering is described in the supplementary appendix.
6This is possibly due to the resampling step in the particle filter which introduces random variation.
1.6. Modeling and Forecasting Inflation via Time-Varying Autoregressive Models

Inflation forecasting is an important task for many central banks. As discussed in Clarida et al. (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). Moreover, low inflation seems to promote growth and support sustainable employment in the long run (Bernanke, 2007). 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. However, numerous studies have shown that persistence and volatility of inflation changes over time (see among others, Cogley and Sargent, 2005; and Stock and Watson, 2007). In an influential work, Stock and Watson (2007) argued 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 models have performed well until the 1980s, but cannot display the same accuracy in recent decades, when simple statistical models, like the first order autoregressive model, have been dominating forecasts.

with sample size \( n = 1000 \) for each of the 3 cases considered before. In order to make the results based on the EVT sensible, a large number of importance weights is necessary when computing the tail threshold \( (u) \) to evaluate the Generalized Pareto Distribution. Results are shown in table 1.6. The likelihood-ratio is a one-tailed test and it rejects the null hypothesis of the existence of the variance for values greater than 1.64. Therefore, we can conclude that there is evidence supporting the existence of the variance of the importance weights, and the multivariate NAIS is a theoretically sound estimator for the real likelihood function.

<table>
<thead>
<tr>
<th>DGP</th>
<th>( u = 0.4M )</th>
<th>( u = 0.3M )</th>
<th>( u = 0.2M )</th>
<th>( u = 0.1M )</th>
<th>( u = 0.01M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-12.706</td>
<td>-10.895</td>
<td>-8.720</td>
<td>-5.937</td>
<td>-1.693</td>
</tr>
</tbody>
</table>

The table presents likelihood-ratio test statistics for several thresholds used to evaluate the Generalized Pareto Distribution.
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 especially on the estimate of the underlying trend inflation. 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, the recent literature has explored time-varying univariate models for trend inflation forecasting, usually outperforming other approaches such as survey-based forecasts (see Clark and Doh, 2014).

Furthermore, 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 (see for example Stock and Watson, 2007). Consequently, several recent studies have developed time-varying parameter models with stochastic volatility for inflation modeling (e.g. Clark and Doh, 2014; Chan, 2013). In this empirical section we propose time-varying autoregressive models with stochastic volatility (TVP-AR hereafter) to estimate trend inflation. Autoregressive models have been shown to forecast inflation well (see Faust and Wright, 2013). The proposed model captures the following observed dynamics in the inflation process: (i) time-varying trend (e.g. Cogley, 2002; Stock and Watson, 2007); (ii) time-varying persistence (e.g. Cogley and Sargent, 2002) and (iii) time-varying volatility (e.g. Chan, 2013; Stock and Watson, 2007). The time-varying AR we consider is given by,

\[ \pi_t = \phi_{0,t} + \sum_{i=1}^{p} \phi_{i,t} \pi_{t-i} + \exp(\frac{h_t}{2}) \epsilon_t, \quad \epsilon_t \sim N(0,1). \]  

(1.39)

For the time-varying autoregressive coefficients \( \phi_{i,t} \) we consider a random walk for autoregressive processes of order 1,

\[ \alpha_t = \alpha_{t-1} + \sigma_f \eta_t, \quad \eta_t \sim N(0,1), \]  

(1.40)

where \( \alpha_t \) yields the first partial autocorrelation of \( \pi_t \) which are mapped into autoregressive coefficients via the Durbin-Levinson algorithm and the restrictions considered in section 3. For autoregressive process of higher order we propose to let the autoregressive coefficients interact through a common factor structure. A factor specification is able to account for time-variation while maintaining model flexibility,

\[ \alpha_t = \Phi_c + \Phi_f f_t \]  

(1.41)

\[ f_t = \varphi f_{t-1} + \sigma_f \eta_t, \quad \eta_t \sim N(0,1), \]  

(1.42)

where \( f_t \) is a scalar latent factor driving the dynamics of the time-varying parameters \( \phi_{i,t}, \) and \( \Phi_c \) and \( \Phi_f \) are \( p \times 1 \) vectors. We select one factor due to the fact that only
1.6. Modeling and Forecasting Inflation via Time-Varying Autoregressive Models

the first order autoregressive coefficient displays strong time-variation. Moreover, the time-varying intercept and the log-volatility are defined by random walk processes,

\[
\begin{align*}
\phi_{0,t+1} &= \phi_{0,t} + \sigma_0 \eta_{0,t}, \quad \eta_{0,t} \sim N(0, 1), \\
h_{t+1} &= h_t + \sigma_h \nu_t, \quad \nu_t \sim N(0, 1).
\end{align*}
\]

A random walk process for the conditional variance is a popular specification in the inflation literature since the influential paper of Stock and Watson (2007). In addition, it yields a parsimonious specification avoiding the increase of static parameters.

Despite incorporating a unit root, our model specification given by (1.39)-(1.44), provides a convenient expression for estimating the inflation trend and allows the use of efficient Kalman filter and smoothing methods to estimate the time-varying linear parameter \(\phi_{0,t}\). This procedure increases numerical efficiency and reduces the computational cost. Note that the likelihood function is based on a simulated prediction error decomposition, therefore usual asymptotic properties apply.

1.6.1 Statistical Properties of the model

The TVP-AR captures 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 of the process (unconditional mean). The unconditional mean of \(\pi_t\) is the value to which the series is expected to converge once the transitory component dies out,

\[
\mu_t = \frac{\phi_{0,t}}{1 - \sum_{i=1}^{p} \phi_{i,t}}.
\]

Note that the trend inflation is a well-defined process due to the restrictions on the time-varying autoregressive coefficients, \(\phi_{i,t}\). Therefore, the stability constrains can be understood as necessary due to the central bank activity in controlling and anchoring trend inflation. Moreover, given the specification of the intercept as a random walk, the trend inflation has a unit root, therefore the detrended inflation, \(\tilde{\pi}_t = \pi_t - \mu_t\), is a zero mean locally stationary process. Although any random walk process hits an upper or lower bound with probability one, we follow the recent literature on modeling inflation and impose that any unit root originates in the trend component (see Chan et al., 2013; for a discussion on non-stationary trend inflation models).

Furthermore, it is interesting to highlight the differences between the model given by equation (1.39) and the UCSV of Stock and Watson (2007). The latter can be described as a local-level with stochastic volatility in the trend and in the transitory
components,

\[
\begin{align*}
\pi_t &= \mu_t + \exp(h_{1,t}/2)e_t, \quad e_t \sim N(0,1), \\
\mu_t &= \mu_{t-1} + \exp(h_{2,t}/2)\eta_t, \quad \eta_t \sim N(0,1), \\
h_{1,t} &= h_{1,t-1} + \gamma v_{1,t}, \quad v_{1,t} \sim N(0,1), \\
h_{2,t} &= h_{2,t-1} + \gamma v_{2,t}, \quad v_{2,t} \sim N(0,1),
\end{align*}
\]

its only parameter \(\gamma\) controls the smoothness of both stochastic volatility processes. The UCSV estimates the long-run inflation by its trend component \(\mu_t\), whereas the model \(1.39\) via time-varying parameters. Nevertheless, the UCSV results are regarded as a benchmark for inflation modeling and comparisons will be made with our model specification. Moreover, despite being also a challenging model to estimate our numerical procedure can easily handle two stochastic volatilities with just minor changes in the baseline algorithm.

### 1.6.2 Bounded Trend Models

Most time-varying parameter models assume that the coefficients are driven by random walk processes. Although an economic-based interpretation is often provided, in the inflation literature the presence of a random walk implies a trend component that can hit any upper or lower bound in the long-run, which is inconsistent with a central bank that pursues inflation stability. Chan et al. (2013) proposed a model where the central bank controls the inflation expectations and commits to allow them to fluctuate within a fixed interval. In this section we explain how to impose a bounded trend in our modeling specification. We aim to restrict the long-run inflation defined in \(1.45\) to lie within a fixed interval \([\mu_l, \mu_u]\), respectively the lower and upper bound. Therefore, a new reparametrization is necessary, where the state-variable is the trend inflation. We assume the following parametrization,

\[
\mu_t = \Gamma(\alpha_{\mu,t}).
\]

The function \(\Gamma(.)\) restricts \(\mu_t\) to be within the required bounds. The state-variable \(\alpha_{\mu,t}\) is the unrestricted trend, and it is specified as a random walk\(^7\)

\[
\alpha_{\mu,t} = \alpha_{\mu,t-1} + \sigma_0 \eta_{0,t}, \quad \eta_{0,t} \sim N(0,1).
\]

For \(\Gamma(.)\) any function restricted between \([\mu_l, \mu_u]\) suffices. In this implementation we follow Monache and Petrella (2017) and apply the following transformation,

\[
\mu_t = \Gamma(\alpha_{\mu,t}) = \frac{\mu_u \exp(\alpha_{0,t}) + \mu_l}{\exp + 1}.
\]

\(^7\)For the sake of simplicity we use \(\sigma_0\) to represent the standard deviation of the trend inflation. Note that in the unbounded model \(\sigma_0\) is the standard deviation of the intercept.
In this paper we decide to fix the upper and lower bound in 5 and 0 respectively. This is in line with the findings in Chan et al. (2013). From equation (1.45) we can write the intercept as,

$$\phi_{0,t} = \mu_t \left(1 - \sum_{i=1}^{P} \phi_{i,t}\right).$$  \hspace{1cm} (1.53)

Finally from an estimation perspective the Kalman filter cannot be used to evaluate the density $p(y_t|\alpha_t;\gamma)$ in the likelihood integral (1.7). This is due to the fact that the intercept is now a highly nonlinear state-variable, therefore it needs to be simulated jointly with the other non-linear time-varying parameters. Fortunately, just minor changes are required in the baseline algorithm previously described. The main step is to remove the marginalized density $p^*(y_t|\alpha,\gamma)$ in (1.7) by a general Gaussian density $p(y_t|\alpha,\gamma)$. In the supplementary appendix we provide more details about the estimation for this non-linear model.

### 1.6.3 Results

The time-varying autoregressive model with stochastic volatility was used to forecast U.S. inflation and its predictive power was measured by an out-of-sample forecasting exercise in comparison to univariate benchmark models. Moreover, smoothed estimates of the time-varying parameters are used to shed light on the evolution of the inflation trend in the last decades.

We estimate the model using annualized deseasonalized quarterly U.S. C.P.I. inflation from 1960Q1 to 2016Q4. More specifically, let $P_t$ 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 Federal Reserve economic database (FRED). For the autoregressive order we select $p = 1, 2, 4$. The maximum likelihood estimation and state smoothing were performed using the Rao-Blackwellized multivariate NAIS previously described with 400 draws for the importance sampler and 15 GH nodes. The estimates are presented in table 1.7. Most parameters are statistically significant, especially for the TVP-AR(1) and TVP-AR(2) models.

After estimating the parameters, the NAIS procedure was used to obtain smoothed estimates of the state variables. Figure 1.1 presents the estimated long-run inflation across specifications. Note that we also show the trend inflation estimated via the UCSV in view of comparing the results. 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 UCSV is much closer to the actual inflation. This is particular relevant to explain the inflation of the 1970s. While in the UCSV model the Great Inflation was mainly caused by an increase in the trend component, in our specification the trend was less affected and the random shocks play a much important role. After the 1980s the long-run inflation seems to be stable and it was not disturbed by the recent Financial Crisis.
### Table 1.7: Estimation Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TVP-AR(1)</th>
<th>TVP-AR(1)-B</th>
<th>TVP-AR(2)</th>
<th>TVP-AR(2)-B</th>
<th>TVP-AR(4)</th>
<th>TVP-AR(4)-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_0$</td>
<td>0.127</td>
<td>0.257</td>
<td>0.110</td>
<td>0.183</td>
<td>0.502</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.057)</td>
<td>(0.040)</td>
<td>(0.368)</td>
<td>(0.075)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>$\sigma_h$</td>
<td>0.184</td>
<td>0.269</td>
<td>0.208</td>
<td>0.244</td>
<td>0.363</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.067)</td>
<td>(0.052)</td>
<td>(0.118)</td>
<td>(0.088)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>$\sigma_f$</td>
<td>0.050</td>
<td>0.080</td>
<td>0.072</td>
<td>0.065</td>
<td>0.366</td>
<td>0.586</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.027)</td>
<td>(0.017)</td>
<td>(0.023)</td>
<td>(0.267)</td>
<td>(0.880)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>NA</td>
<td>NA</td>
<td>0.988</td>
<td>0.984</td>
<td>0.984</td>
<td>0.860</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.014)</td>
<td>(0.025)</td>
<td>(0.012)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>$\Phi_{c,1,1}$</td>
<td>NA</td>
<td>NA</td>
<td>0.257</td>
<td>0.397</td>
<td>−0.008</td>
<td>−0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.254)</td>
<td>(0.215)</td>
<td>(1.176)</td>
<td>(1.000)</td>
</tr>
<tr>
<td>$\Phi_{c,2,1}$</td>
<td>NA</td>
<td>NA</td>
<td>0.110</td>
<td>0.083</td>
<td>−0.152</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.064)</td>
<td>(0.067)</td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>$\Phi_{f,2,1}$</td>
<td>NA</td>
<td>NA</td>
<td>−0.062</td>
<td>0.038</td>
<td>−0.007</td>
<td>−0.130</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.160)</td>
<td>(0.064)</td>
<td>(0.034)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>$\Phi_{c,3,1}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>−0.037</td>
<td>0.399</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.181)</td>
<td>(0.262)</td>
</tr>
<tr>
<td>$\Phi_{f,3,1}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.224</td>
<td>0.699</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.036)</td>
<td>(0.091)</td>
</tr>
<tr>
<td>$\Phi_{c,4,1}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>−0.152</td>
<td>−0.114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.060)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>$\Phi_{f,4,1}$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>−0.052</td>
<td>−0.330</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.037)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>LogL</td>
<td>−432.039</td>
<td>−436.390</td>
<td>−429.213</td>
<td>−431.322</td>
<td>−418.593</td>
<td>−428.870</td>
</tr>
</tbody>
</table>

The table reports parameter estimates of our time-varying autoregressive model with stochastic volatility for U.S. inflation. The statistical standard errors are in parentheses. TVP-AR-B means time-varying autoregressive model with a bounded trend.

The results for the bounded model indicate that the trend inflation remains very stable throughout the sample, including during the Great Inflation of the 1970s, when it does not seem to approach the upper bound. This is consistent with the findings in Chan et al. (2013), despite the fact that they used a different specification and estimation procedure. Finally, it is interesting to highlight that the models with autoregressive terms point out the importance of the random innovations in the inflation process. Therefore, our modeling framework suggest that supply shocks such as energy prices are not necessarily incorporated in inflation through its trend component, but are merely transitory effects. This contrasts with results in the literature based on unobserved components with stochastic volatility (see Stock and Watson, 2007). On the other hand, Chan et al. (2013) and Monache and Petrella (2017) found similar conclusions when using time-varying autoregressive models.

The volatility results in Figure 1.2 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, possibly contributing to the Great Moderation of output volatility (Cecchetti et al., 2007). Recently, volatility peaked again in the aftermath of the Global Financial Crisis, reaching levels as high as observed during the 1970s. Note that the model with
1.6. MODELING AND FORECASTING INFLATION VIA TIME-VARYING AUTOREGRESSIVE MODELS

Figure 1.1: Inflation and its estimated trend. The dark line is actual inflation. The blue line is the unbounded trend and red line is the bounded trend inflation.

Figure 1.2: Inflation volatility. The blue line is the log-volatility in the unbounded model. The red line is the log-volatility in the bounded model.

bounded trend provides the same conclusions, although higher levels of volatility are observed. This is explained by the greater variation of the random innovations caused by the presence of a restricted trend.

1.6.4 Forecasting Evaluation

This section presents results on the out-of-sample predictive power of the time-varying autoregressive model with stochastic volatility in comparison to univariate models. We consider a range of univariate models covering the features of inflation. Simple autoregressive models: the random walk, and the AR(1). Models with time-
varying trend: the local-level. Finally, models with time-varying trend, persistence and volatility: time-varying AR with stochastic volatility and unobserved components stochastic volatility model.

- RW: Random Walk
  \[
  \pi_t = \pi_{t-1} + \sigma \varepsilon_t, \quad \varepsilon_t \sim N(0, 1)
  \]  
  (1.54)

- AR(1): Time-invariant AR(1)
  \[
  \pi_t = \phi_0 + \phi_1 \pi_{t-1} + \sigma \varepsilon_t, \quad \varepsilon_t \sim N(0, 1)
  \]  
  (1.55)

- LL: Local Level
  \[
  \pi_t = \mu_t + \sigma \varepsilon_t, \quad \varepsilon_t \sim N(0, 1),
  \]
  (1.56)

  \[
  \mu_t = \mu_{t-1} + \sigma \mu \eta_t, \quad \eta_t \sim N(0, 1)
  \]  
  (1.57)

- UCSV: Unobserved components stochastic volatility

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 Great Moderation of the 1990s, and the Global Financial Crisis. Forecasts are computed for 1, 2, 4, and 8 steps ahead and are all evaluated on the basis of the root mean square error (RMSE) and the mean absolute error (MAE) defined as,

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} v_{t,h}^2}, \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |v_{t,h}|,
\]  
(1.58)

where \( n \) is the total number of forecasts, and \( v_{t,h} = y_{t+h} - \hat{y}_{t+h|t} \) is the \( t + h \) forecast error. The variable \( \hat{y}_{t+h|t} \) is the importance sampling estimate for \( E(y_{t+h}|y_{1:t}) \) and can be obtained from the last estimate of the time-varying parameters calculated via the multivariate NAIS with sample size up to time \( t \). To construct \( E(y_{t+h}|y_{1:t}) \) the time-varying parameters are updated until time \( t \) and are treated as constants forward in time. We can estimate \( E(y_{t+h}|y_{1:t}) \) for the TVP-AR as,

\[
\hat{y}_{t+h|t} = \hat{\phi}_{0,t-1} + \sum_{i=1}^{p} \hat{\phi}_{i,t-i} \hat{y}_{t+h-i|t},
\]  
(1.59)

The variable \( \hat{\phi}_{i,t} \) is an estimator for \( E(\phi_{i,t}|y_{1:t}) \), and can be computed from the last estimate of the time-varying parameters via our importance sampling methodology. More details on state estimation by importance sampling methods can be found in the supplementary appendix.
1.6. MODELING AND FORECASTING INFLATION VIA TIME-VARYING AUTOREGRESSIVE MODELS

The results of the point forecasts are presented in Table 1.8. We note that all models with time-varying autoregressive coefficients forecast inflation well and outperform most benchmark specifications, especially in medium and longer horizons. Moreover, the TVP-AR(1) seems to be the best predictor for \( h = 1 \) and \( h = 2 \), while the TVP-AR(2) yields the best forecasts in longer horizons. For one-step ahead forecast our models are able to slightly outperform the UCSV. However for longer horizons the UCSV is easily surpassed by our time-varying autoregressive specifications. These results suggest that models based on unobserved components are only appropriate for very short term forecasting, and in longer horizons more time-varying parameters are necessary to reduce forecast errors. Given that the competing specifications cover a range of nested and non-nested models, we use the superior predictive ability test of Hansen (2005) to compare the differences in the point forecasts. A quadratic loss function is used as measure of point forecasting performance. Table 1.10 (in appendix) presents the results. Results show that the differences on the forecasting accuracy are often statistically significant.

Table 1.8: Point Forecast Results

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( h = 1 )</td>
<td>( h = 2 )</td>
</tr>
<tr>
<td>RW</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.931</td>
<td>0.882</td>
</tr>
<tr>
<td>LL</td>
<td>0.898</td>
<td>0.838</td>
</tr>
<tr>
<td>UCSV</td>
<td>0.889</td>
<td>0.816</td>
</tr>
<tr>
<td>TVP-AR(1)</td>
<td>0.876</td>
<td>0.760</td>
</tr>
<tr>
<td>TVP-AR(1)-B</td>
<td>0.859</td>
<td>0.734</td>
</tr>
<tr>
<td>TVP-AR(2)</td>
<td>0.887</td>
<td>0.778</td>
</tr>
<tr>
<td>TVP-AR(2)-B</td>
<td>0.890</td>
<td>0.777</td>
</tr>
<tr>
<td>TVP-AR(4)</td>
<td>0.857</td>
<td>0.775</td>
</tr>
<tr>
<td>TVP-AR(4)-B</td>
<td>0.858</td>
<td>0.760</td>
</tr>
</tbody>
</table>

The table presents root mean square errors and mean absolute errors relative to the RW. Forecasts are computed using an expanding window initially ending in 1985Q1, yielding 128, 127, 125, and 121 out-of-sample forecasts for \( h = 1, 2, 4, 8 \) respectively.

Additionally, we evaluate the density forecasts of our time-varying AR process in comparison to the benchmark models. The \( h \)-step ahead density forecast \( d_t^h(\cdot) \) can be defined for the TVP-AR as,

\[
y_{t+h|t} \sim N\left(\hat{\phi}_{0,t-1} + \sum_{i=1}^{p} \hat{\phi}_{i,t-1} \hat{y}_{t+h-i|t}, \exp(h_t)\right). \tag{1.60}
\]

This density can be approximated by the following discrete sample,

\[
y_{t+h|t} \sim \left\{ \omega_t^{(j)}, N(\hat{\phi}_{0,t-1} + \sum_{i=1}^{p} \hat{\phi}_{i,t-1} \hat{y}_{t+h-i|t}, \exp(h_t^{(j)}) \right\}. \tag{1.61}
\]
Initially we compare the whole calibration of the density forecasts. According to Diebold et al. (1998) this can be verified based on the probability integral transform (PIT). Let $D_h^t(.)$ be the cumulative distribution function of the $h$-step density forecast $d^h_t(.)$, where the forecasts are conditional on the time $t$. For non-linear state-space models the PIT of the $h$-step ahead forecast $\hat{y}_{t+h|t}$ is $D_h^t(y_{t+h|t})$ can be approximated as,

$$D_h^t(y_{t+h|t}) = \int_{-\infty}^{y_{t+h|t}} d^h_t(s) ds \approx \sum_{i=1}^{M} \bar{\omega}_t^{(i)} \Phi(v_t^{(i)}0, F_t^{(i)}),$$

(1.62)

where $\bar{\omega}$ are the normalized importance weights, $\Phi(.)$ is the cumulative Gaussian function, and $v_{t+h}$ and $F_{t+h|t}$ are the forecasting error and its associated variance. If the forecasting density $d_{t+h|t}$ is correctly calibrated the PIT, $D_t(y_{t+h|t})$ is a uniform distributed random variable (see Diebold et al., 1998).

In figure 1.4 (in appendix) we report PIT histograms with 5 bins for our three proposed specifications and the UCSV for $h = 1$. To compare the results of the calibration we also show the PITs for the local level model. Under correct calibration each bin should contain 20% of all PITs. We note that the forecasting ability of the time-varying AR models and the UCSV are very good with uniformly distributed PITs for $h = 1$. The TVP-AR(1) seems to be a particularly well-calibrated forecasting model. On the other hand, the local level does not show satisfactory forecasting calibration, with too many PITs in the third bin. This indicates the importance of considering time-varying persistence and stochastic volatility for inflation forecasting.

Secondly, we use the weighted Continuous Ranked Probability Score (wCRPS) introduced by Gneiting and Raftery (2007) to evaluate density forecast performance. The $wCRPS$ avoids some of the drawbacks of the usually employed log-score (the logarithm of the predictive density), as log-score does not reward values from the predictive density that are close but not equal to the actual realized value. Moreover, average log-score also does not consider regions of interest in the integration, and as discussed before, for monetary policy setting the center of the is more relevant than the tails of the distribution. The $wCRPS$ computes the average absolute distance between the empirical cumulative distribution function of $y_{t+h}$ and the predicted CDF. In addition, the comparison can be weighted by a function that emphasizes certain areas of interest. The $wCRPS$ can be defined as,

$$wCRPS_{t+h|t} = \int w(z) \left( \hat{F}_{t+h|t} - 1_{(y_{t+h} < z)} \right)^2 dz,$$

(1.63)

where $w(z)$ is a weight function, and $\hat{F}_{t+h|t}$ is the $h$-step ahead cumulative density function evaluated at $z$. However, equation (1.63) is not available analytically, therefore, we use the following approximation,

$$wCRPS_{t+h|t} = \frac{y_u - y_l}{K-1} \sum_{k=1}^{K} w(y_k) \left( \hat{F}_{t+h|t} - 1_{(y_{t+h} < z)} \right)^2,$$

(1.64)
where \( y_u \) and \( y_l \) are the range of integration, and \( K \) defines the accuracy of the approximation. Moreover, \( y_k \) is given by,

\[
y_k = y_l + k \frac{y_u - y_l}{K},
\]

(1.65)

To define \( y_u \) and \( y_l \) we apply an interval around the center of the predictive density forecast, i.e. \( y_u = \hat{y}_{t|t+h} + 10 \), and \( y_l = \hat{y}_{t|t+h} - 10 \). We have found that a range of 10 is able to cover the whole space of integration. However, as the mean and the variance of the predictive density vary for each particle we define the \( wCRPS \) at time \( t \) as the following weighted average,

\[
\bar{wCRPS}_{t+h|t} \approx \sum_{i=1}^{M} \bar{\omega}_t^{(i)} wCRPS_{t+h|t}^{(i)}, \quad (1.66)
\]

where \( \bar{\omega}_t \) are the normalized importance weights. The average of all \( \bar{wCRPS}_{t+h|t} \) yields the \( wCRPS \) for the whole sample. The model with lower average \( wCRPS \) is always preferred. Thus, if the ratio of \( wCRPS \) for model \( i \) over \( j \) is greater then one, model \( j \) is preferred to model \( i \) and vice versa. Finally, in the uniform case \( w(.) = 1 \) the same weight is given to all areas of the predictive density. A more interesting weighting procedure for inflation forecasting is to define \( w(.) \) as the probability density function of a standard normal distribution, \( w(y_k) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{z^2}{2} \right) \),

(1.67)

where we define \( z \) as the standardized set of points \( y_k \). Using the standard normal density as weight function more emphasis is given to the area around the center of the predictive density.

The results are presented in table [1.9]. Generally, models based on time-varying autoregressive terms perform well, especially in medium and longer horizons. We note that our specification outperforms the UCSV for forecasts longer than one-step ahead. This indicates that the the TVP-AR models have a higher probability of delivering forecasts which are closer to the true observed inflation than the UCSV. Therefore, merely including stochastic volatility in an unobserved components model is not enough to minimize forecast errors. In addition, the presence of a bounded trend improves the forecasting power in longer horizons. Finally, both weighting methods point out similar results, and the TVP-AR(2) can be regarded as the model that delivers the best density forecasts, although all models with time-varying parameters produce better forecasts than the UCSV in the long-run.

1.7 Conclusions

This paper considers estimation of time-varying autoregressive models with stochastic volatility. We propose an estimation procedure for state-space models with linear
and non-linear time-varying parameters and stationarity restrictions. In addition, we discuss extensions to a high number of non-linear time-varying parameters and a reparametrization to guarantee stability conditions are satisfied at all times. A novel specification for non-linear state-space models is proposed based on common factor models, greatly improving estimation.

Differently from the literature the focus of this paper is to develop a classical maximum likelihood alternative. The methodology is built on a multivariate extension of the numerically accelerated importance sampling of Koopman et al. (2015) together with the Rao-Blackwelization step discussed in Moura and Turatti (2014). This yields a numerically and computationally efficient procedure, as only non-linear time-varying parameters need to be simulated. Monte Carlo results show that the procedure can identify model parameters, and it estimates the log-likelihood at least two times more efficiently than a particle filter for the same computational time.

We apply the model to forecast U.S. quarterly C.P.I. inflation. Smoothed 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 highlights the predictive power of the proposed model, outperforming simpler and more complex univariate benchmark specifications, especially in medium and longer horizons. Finally, well-calibrated density forecasts are only obtained when time-varying persistence and volatility is considered.
1.8 References


1.9 Appendix

1.9.1 Tables

In this section we report the table with results of the superior test ability (SPA) for the differences in the RMSE between the TVP-AR and the competing models. Bold entries denote rejection of the null of equal point predictive ability. P-values were simulated using a stationary bootstrap with 10000 replications.
### Table 1.10: Superior predictive ability test results

<table>
<thead>
<tr>
<th>Model</th>
<th>Model</th>
<th>h</th>
<th>RW</th>
<th>AR(1)</th>
<th>LL</th>
<th>UCSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVP-AR(1)</td>
<td>h</td>
<td>1</td>
<td>-0.203</td>
<td>-0.187</td>
<td>-0.120</td>
<td>-0.094</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.021)</td>
<td>(0.026)</td>
<td>(0.087)</td>
<td>(0.134)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-0.163</td>
<td>-0.186</td>
<td>-0.154</td>
<td>-0.162</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.048)</td>
<td>(0.031)</td>
<td>(0.054)</td>
<td>(0.043)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-0.179</td>
<td>-0.259</td>
<td>-0.223</td>
<td>-0.235</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.037)</td>
<td>(0.004)</td>
<td>(0.012)</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-0.164</td>
<td>-0.225</td>
<td>-0.140</td>
<td>-0.155</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.052)</td>
<td>(0.006)</td>
<td>(0.071)</td>
<td>(0.050)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TVP-AR(1)-B</td>
<td>1</td>
<td>-0.209</td>
<td>-0.208</td>
<td>-0.173</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.016)</td>
<td>(0.018)</td>
<td>(0.034)</td>
<td>(0.045)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-0.171</td>
<td>-0.202</td>
<td>-0.179</td>
<td>-0.195</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.047)</td>
<td>(0.020)</td>
<td>(0.038)</td>
<td>(0.026)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-0.189</td>
<td>-0.295</td>
<td>-0.259</td>
<td>-0.290</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
<td>(0.002)</td>
<td>(0.004)</td>
<td>(0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-0.177</td>
<td>-0.262</td>
<td>-0.178</td>
<td>-0.201</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.038)</td>
<td>(0.000)</td>
<td>(0.033)</td>
<td>(0.015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TVP-AR(2)</td>
<td>1</td>
<td>-0.193</td>
<td>-0.201</td>
<td>-0.159</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.023)</td>
<td>(0.020)</td>
<td>(0.043)</td>
<td>(0.085)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-0.164</td>
<td>-0.190</td>
<td>-0.154</td>
<td>-0.160</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.044)</td>
<td>(0.032)</td>
<td>(0.054)</td>
<td>(0.047)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-0.177</td>
<td>-0.250</td>
<td>-0.195</td>
<td>-0.190</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.038)</td>
<td>(0.006)</td>
<td>(0.022)</td>
<td>(0.023)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-0.191</td>
<td>-0.392</td>
<td>-0.203</td>
<td>-0.232</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.029)</td>
<td>(0.000)</td>
<td>(0.020)</td>
<td>(0.007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TVP-AR(2)-B</td>
<td>1</td>
<td>-0.199</td>
<td>-0.150</td>
<td>-0.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.023)</td>
<td>(0.051)</td>
<td>(0.227)</td>
<td>(1.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-0.178</td>
<td>-0.242</td>
<td>-0.210</td>
<td>-0.239</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.040)</td>
<td>(0.007)</td>
<td>(0.016)</td>
<td>(0.007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-0.193</td>
<td>-0.297</td>
<td>-0.263</td>
<td>-0.286</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.029)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-0.1900</td>
<td>-0.290</td>
<td>-0.205</td>
<td>-0.245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.030)</td>
<td>(0.000)</td>
<td>(0.020)</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TVP-AR(4)</td>
<td>1</td>
<td>-0.184</td>
<td>-0.186</td>
<td>-0.158</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.028)</td>
<td>(0.029)</td>
<td>(0.049)</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-0.166</td>
<td>-0.192</td>
<td>-0.163</td>
<td>-0.172</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.048)</td>
<td>(0.026)</td>
<td>(0.047)</td>
<td>(0.035)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-0.179</td>
<td>-0.250</td>
<td>-0.218</td>
<td>-0.213</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.040)</td>
<td>(0.004)</td>
<td>(0.013)</td>
<td>(0.012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-0.155</td>
<td>-0.192</td>
<td>-0.117</td>
<td>-0.114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.054)</td>
<td>(0.020)</td>
<td>(0.101)</td>
<td>(0.102)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TVP-AR(4)-B</td>
<td>1</td>
<td>-0.161</td>
<td>-0.139</td>
<td>-0.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.045)</td>
<td>(0.064)</td>
<td>(0.126)</td>
<td>(0.179)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-0.198</td>
<td>-0.197</td>
<td>-0.175</td>
<td>-0.183</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.043)</td>
<td>(0.023)</td>
<td>(0.038)</td>
<td>(0.027)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-0.190</td>
<td>-0.292</td>
<td>-0.254</td>
<td>-0.279</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.030)</td>
<td>(0.001)</td>
<td>(0.004)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-0.171</td>
<td>-0.249</td>
<td>-0.156</td>
<td>-0.168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.043)</td>
<td>(0.002)</td>
<td>(0.051)</td>
<td>(0.035)</td>
</tr>
</tbody>
</table>

The table presents superior predictive ability test statistics and their p-value. Bold entries denote rejection of the null of equal point predictive ability. P-values were simulated using a stationary bootstrap with 10000 replications.
1.9.2 Figures
In this section we report the figure with results of the comparison between NAIS and PF. Furthermore, we report the probability density function of the PITs for one-step ahead forecasting.

![Numerical Efficiency](image1)

**Numerical Efficiency**

![Computational Efficiency](image2)

**Computational Efficiency**

**Figure 1.3:** Comparison NAIS and PF. The standard deviations are computed as averages across 100 realizations. For each realization the standard deviation is computed using 30 random seeds.
1.10 Supplementary Appendix

This supplementary appendix provides additional material supporting the main text. Section 2 provides additional details on the RB-NAIS developed in the main text, especially concerning the estimation of the bounded-trend model. Section 3 presents...
methods of filtering and smoothing based on importance sampling. Section 4 reviews some basics of Extreme Value Theory we use in the Monte Carlo experiment in the main text.

### 1.10.1 RB-NAIS for Bounded Trend Models

The RB-NAIS described in the main text is applicable for applications where there are linear and non-linear time-varying parameters. This is the case for most models in time-series econometrics, from TVP vector autoregressions to unobserved component models. However, sometimes we are interested in restricting a linear time-varying parameter to satisfy some bounds, thus all the time-varying parameters can become non-linear. Notable examples are models with bounded trend discussed in the main text. Therefore, an adaptation is required. The main point is to substitute the RB-likelihood equation \((1.14)\) for,

\[
L(\gamma, y) = \int_{T} \prod_{t=p+1}^{T} p(y_t|\alpha_t; \gamma) p(\alpha_t|\alpha_{t-1}; \gamma) d\alpha_{p+1} d\alpha_T, \tag{1.68}
\]

where \(p(y_t|\alpha_t; \gamma)\) can be written as,

\[
p(y_t|\alpha_t; \gamma) \sim N(\phi_{0,t} + \sum_{i=1}^{p} \phi_{i,t}, \exp(h_t)). \tag{1.69}
\]

For models with bounded trend the intercept needs to be simulated jointly with the other time-varying parameters, thus the dimensionaly of the problem is increased. The remaining steps, such as compute the means and sampling are identical as the RB version.

Finally, to compute the least square problem equation \((3.56)\) a large number of combined nodes can be necessary for a high number of non-linear time-varying parameters (>3). Note that the number of combined nodes is \(M^d\), thus when using \(M = 15\) a total of 3375 nodes can be required to compute the LS estimates for an TVP-AR(1) with stochastic volatility and bounded trend. Fortunately, most of these nodes can be discarded using a technique called pruning. The main argument is that many nodes have a low weight on the numerical integration, thus they can be eliminated. Formally, let \(\theta_m\) be a threshold in the two-dimensional case,

\[
\theta_m = \frac{w_1 w_{m+1}}{m}. \tag{1.70}
\]

Then the two-dimensional GH quadrature has the following approximation,

\[
\sum_{i=1}^{S} \sum_{j=1}^{S} \{w(z_i) w(z_j) > \theta_m\} w(z_i) w(z_j) f(z_i, z_j). \tag{1.71}
\]

With this technique a large number of draws with little impact in the integration can be discarded, thus improving computational tractability.
1.10.2 Filtering and smoothing with Importance Sampling methods

This section presents methods of filtering, signal extraction and sampling with importance sampling. These methods were used throughout the main text. We initially discuss filtering methods, especially the Rao-Blackwellized particle filter used to compare the likelihood estimation with the multivariate NAIS.

1.10.2.1 Filtering methods

The multivariate NAIS previously developed is a smoothing method, therefore, it provides a smoothed likelihood estimate. However, there are cases when we are interested in filtered estimates, for example in residuals diagnostics. The main difference is that smoothed estimates use information from the whole sample, thus we obtain $E(\alpha_{1:t|T})$, and filtered estimates uses information up to time $t E(\alpha_{1:t|t})$. In this section we discuss a method to obtain filtered estimates.

The likelihood function (1.14) can be estimated by the particle filter (PF). However, differently from the NAIS method, the particle filter draws from a proposal density that does not carry information about the underlying states. More specifically, the PF sets the proposal sampler $g(\alpha_t|y_t)$ to be the transition density $p(\alpha_t|\alpha_{t-1})$. Therefore, the draws are not conditional on the observables. For a detailed discussion on particle filtering we refer to Creal (2012). The complete PF for a time-varying AR model with stochastic volatility reads as follows,

1. Initialize the procedure with an exact or diffuse initial condition for the linear time-varying parameters.

2. Draw the volatility terms, $\{\sigma^2_t\}_{t=1}^T$ and the dependence coefficients, $\{\alpha_t\}_{t=1}^T$ from the transition density $p(\alpha_t|\alpha_{t-1})$.

3. For each time $t = 1 \ldots T$ evaluate the density $p^*(y_t|\alpha_t, \sigma^2_t)$ by the prediction error decomposition given the draws from step 2 and obtain the importance weights as, $\omega(\theta, y) = p^*(y_t|\alpha_t, \sigma^2_t)$.

4. Normalize the importance weights, $\bar{\omega}^{(i)} = \frac{\omega^{(i)}}{\sum_{j=1}^S \omega^{(j)}}$.

5. Resample with replacement the draws on the basis of the normalized weights $\bar{\omega}^{(i)}$.

6. Apply the Kalman Filter to estimate the the linear time-varying parameters.

7. Estimate the likelihood as $\hat{L}(\theta, y) = \frac{1}{N} \sum_{i=1}^N \left[ \prod_{t=1}^T p^*(y_t|\alpha_t, \sigma^2_t) \right]$.

Using the particle filter an estimate of the filtering distribution is given by,

$$p(\alpha_{1:t}|y_{1:t}) \approx \sum_{i=1}^S \bar{\omega}^{(i)} \delta_{a_{1:t}^{(i)}}(\alpha_{1:t}), \quad (1.72)$$
where $\delta_{x_{1:t}^{(i)}}$ is a dirac measure located at $x_{1:t}^{(i)}$. This equation approximates the predictive density equation (1.61) in the main text, which is required to compute density forecasts.

### 1.10.2.2 Signal extraction via multivariate NAIS

In this section we discuss how to compute state estimates and signal extraction with importance sampling methods. The NAIS procedure is a special case of the general importance sampling technique, thus it is able to deliver estimates of the state variables. As the importance sampler uses information from the whole sample the NAIS method is a smoothing procedure. An importance sampling estimator for $h(x)$, where $x \sim p(x)$ is,

$$E(h(x)) = \frac{\sum_{j=1}^{S} \omega_j(y, y_t) h(x_j)}{\sum_{j=1}^{S} \omega_j(y, y_t)}, \quad \text{(1.73)}$$

$$= \sum_{j=1}^{S} \tilde{\omega}^{(j)} h(x_j), \quad \text{(1.74)}$$

where $\tilde{\omega}^{(j)} i = 1, \ldots, S$ are the normalized importance weights. Convergence holds under the weak law of large numbers when $x^{(i)}$, $i = 1, \ldots, S$ are sampled from the importance density $g(.)$ independently. To estimate $\alpha_{1:T}$ from the observables $y_{1:T}$ we fix the parameter vector in its ML estimate and compute the quantity,

$$h_t(x) = \sum_{i=1}^{S} \tilde{\omega}^{(i)} (y_{1:T}) h(x_{1:T}^{(i)}) \to E(\alpha_{1:T}). \quad \text{(1.75)}$$

Note that the importance weights are computed with information from the whole sample, therefore (1.75) delivers a smoothed estimator. We can forecast $x_t$ on basis of the importance sampling estimator (1.75),

$$h_{T+h}(x) = \sum_{i=1}^{S} \tilde{\omega}^{(i)} (y_{1:T}) h(x_{1:T}^{(i)}) \to E(\alpha_{1:T}). \quad \text{(1.76)}$$

where the importance weights are kept fixed at sample size $T$. The quantity (1.76) is the predictive estimator for $x_{T+h}$.

### 1.10.3 Extreme value theory for Importance Sampling

The seminal work of Geweke (1989) argued that a central limit theorem can only be applicable to importance sampling estimates if the variance of the importance weights is known to be bounded. This is usually accomplished by using an importance sampler with fatter tails than the likelihood integrand. However, in state-space models the likelihood is a mixture of densities, thus overdispersion and heavier tails may happen even in the Gaussian case. This implies that it is difficult to guarantee that
the importance sampler cover the whole support of the target integrand. Although it is still unclear if this condition needs to be satisfied in practice. For example, Richard and Zhang (2007) suggested that a theoretical unbounded variance is not an issue for empirical purposes as the variance will always be finite in practice, especially if a single set of random numbers is used to evaluate the log-likelihood function. Moreover, note that the problem is by no means specific to Importance Sampling and actually applies to all Monte Carlo techniques which rely upon ratios of the form \( \omega(.) = g(.) / p(.) \), including therefore Metropolis-Hastings and Markov Chain Monte Carlo methods (Richard and Zhang, 2007).

Suppose we are interested in the conditional expectation of a measurable function \( \varphi: \mathbb{R}^{(k)} \rightarrow \mathbb{R} \), which can be understood as the log-likelihood function in our context. Under some regularity conditions (see Geweke, 1989) the variance of the importance weights is given by

\[
\int \varphi(\alpha)^2 \omega(\alpha)^2 g(\alpha | y) \, d\alpha < \infty,
\]

where \( \alpha \) is the state variable, \( k \) is its dimension, and \( \omega(.) \) are the importance weights. Given condition (1.77) we have the following central limit theorem for the importance sampling estimator \( \hat{\varphi}(\alpha) \),

\[
\sqrt{S}(\hat{\varphi}(\alpha) - \mathbb{E}(\varphi(\alpha) | y_{1:t})) \rightarrow^d N(0, \sigma_{IS}^2), \quad S \rightarrow \infty
\]

The asymptotical variance \( \sigma_{IS}^2 \) is given by,

\[
\sigma_{IS}^2 = \int (\varphi(\alpha) - \mathbb{E}(\varphi(\alpha) | y_{1:t}))^2 \omega(\alpha) g(\alpha | y_{1:t}) \, d\alpha.
\]

The idea behind extreme value theory is to study the behavior of the tails of the empirical distribution of the importance weights. Assume we have a sequence \( \omega_1, ..., \omega_S \) of i.i.d random variables from a distribution function \( F \). The sequence \( \{\omega\}_{i=1}^S \) does not need to be importance weights, although it is the case in this paper. Denote the endpoint of the distribution function \( F \) as \( \omega_E = \sup\{\omega : F(\omega) < 1\} \leq +\infty \). Pickands (1975) defined a new variable \( z_1, ..., z_N \), created from the original \( \omega_i \) that lie above a threshold \( u \), i.e. \( z_i = \omega_i - u \), with \( N < S \). Pickands (1975) proved that when \( N \) and \( u \) are large the distribution of \( \omega_i \) can be approximated by a generalized Pareto distribution. He showed that,

\[
\lim_{u \rightarrow 0} \sup_{0 < z < \omega_e} |F_u(z) - \text{GDP}(z; \xi, \beta)| = 0,
\]

where \( F_u(z) \) is the distribution of the excesses conditional on threshold \( u \), \( \omega_e = \sup\{\omega : F_\omega < 1\} \) is the right end point for \( F_\omega \), and GDP is the generalized Pareto distribution with density,

\[
f(z; \xi, \beta) = \frac{1}{\beta} \left(1 + \frac{z}{\beta} \right)^{1-\xi} \left(1 + \frac{\beta}{\xi} \right), \quad z \in [0, \infty) \text{ if } \xi > 0, \text{ or } \left[0, \frac{\beta}{\xi} \right] \text{ if } \xi < 0, \text{ and } \beta > 0,
\]
where $\xi$ is the shape parameter which characterises the thickness of the distribution’s tails. In fact, it can be shown that when $\xi > 0$ then $E[\omega^j] = \infty$ for $j \geq 1/\xi$. Consequently, the number of moments that the distribution of the importance weights has can be determined by focusing on $\xi$. In particular, we are interested in establishing that $\xi \leq 1/2$. On this basis Koopman (2009) developed likelihood-based tests to verify if $\xi \leq 1/2$. The one used in the main text is the likelihood-ratio test. Initially, order all importance weights $\omega_{1:S}$ in descending order,

$$\omega_{1:S} \geq \omega_{2:S} \ldots \omega_{S:S}. \quad (1.82)$$

The next step is to fix the threshold value $u$. We follow Koopman (2009) and selected several quantiles of the empirical distribution the importance weights. Then compute the conditional exceedances $z_j = w_j - u$, and eliminate all values below the threshold. Finally, the student-t statistic is given by

$$t = \sqrt{\frac{n}{3\hat{\beta}^2}} \left( \hat{\xi} - \frac{1}{2} \right). \quad (1.83)$$

with $\hat{\xi}$ and $\hat{\beta}$ equal to their ML estimates. The null is reject with large positive $t$-statistic.

1.10.4 References


CLASSICAL ANALYSIS OF TIME-VARYING FACTOR-AUGMENTED VECTOR AUTOREGRESSIONS

Douglas Eduardo Turatti
Aarhus University and CREATE

Abstract

In this paper we discuss classical estimation of time-varying factor-augmented vector autoregressions with stochastic volatility. The model is estimated in two steps. In the first one a consistent estimator of the factors is used to form an identified likelihood function. In the second step the FAVAR system is shown to be estimated equation-by-equation. To estimate the resulting time-varying parameters models with stochastic volatility we propose an estimation procedure that combines the Kalman filtering with efficient importance sampling methods, yielding an efficient and fast numerical procedure. Finally, the time-varying parameters are jointly filtered incorporating information from the whole system of equations. In a simulation study we show that the two-step estimation procedure has good finite-sample and asymptotic properties. An empirical application to a large dataset of macroeconomic variables shows that the proposed model produces better forecasts than benchmark approaches in the literature, especially for inflation measures.
2.1 Introduction

The recent econometric literature has seen an increasing interest in large datasets. In this context factor models offer several advantages over other approaches. They provide means to exploit information from a large number of series summarizing it in a small number of variables, thus partially mitigating the omitting variable problem encountered in vector autoregressions. As argued by Stock and Watson (2002a) also practitioners typically examine a large number of variables when making forecasts. Moreover, factor models are less prone to structural instability often found in low-dimensional system. Recently factor-augmented vector autoregressions (FAVAR) originally proposed by Bernanke et al. (2005) became an important tool for macroeconomic analysis of large datasets.

Another recent line of research focuses on small-scaled models with time-varying parameters and stochastic volatility aiming to cope with structural instability. For example, works from Cogley and Sargent (2005) and Primiceri (2005) 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 (2005) estimating time-varying parameter (TVP) models overlooking possible changes in the volatility is likely to generate fictitious dynamics in the coefficients. Univariate models based on unobserved components have been fairly explored in the literature (see for example Stock and Watson, 2007; Chan, 2013). Moreover, time-varying vector autoregressions with stochastic volatility became a standard tool for monetary policy analysis.

This paper propose two extensions of the basic FAVAR. We let the conditional mean parameters and the volatility be time-varying. This means that the FAVAR parameters will be able to accommodate time-variation between the factors and the target variables. A few references have attempted to combine factor models and vector autoregressions with time-varying parameters (e.g. Korobilis, 2013). However, all contributions to the literature focus on Bayesian methods. Instead, we present a classical maximum likelihood alternative. This paper suggests a simple approach to estimate the time-varying FAVAR models. The TVP-FAVAR is estimated in two steps. In the first one the factors are consistently estimated by principal components. Treating the estimated factors as observables, the relationship between the target and the factors are modeled as time-varying vector autoregressions with stochastic volatility. We consider univariate processes for the covariance matrix, meaning the model can be estimated equationwise. In this paper, we mainly investigate factor-augmented vector autoregressive models with stochastic volatility, but the procedure can be applicable to other cases of interest in the literature, for example, models with multivariate unobserved components (e.g. Stock and Watson, 2007), and general vector autoregressions (e.g. Primiceri, 2005).

However, the presence of stochastic volatility implies a nonlinear state-space model, thus no closed form solution for the likelihood function is available. To over-
come this nuisance we develop an estimation procedure based on a Kalman filter inside an importance sampling methodology, yielding a computationally simple and numerically efficient algorithm. We base our methodology on the Numerically Accelerated Importance Sampling (NAIS) proposed by Koopman et al. (2015) together with a Rao-Blackwellization step for importance sampling. After the estimation procedure the time-varying parameters are jointly extracted using the maximum likelihood estimates for static parameters and the importance sampling estimates for the stochastic volatilities. This process can be carried out by a standard Kalman filter. Finally, the extracted time-varying parameters can be used for forecasting, for example. The purpose of the present paper is twofold. The first one is to provide an estimation framework for classical time-varying (FA)VAR models. The second objective is to assess the estimation procedure in a simulation study and the forecasting ability in an empirical application.

The remainder of the paper is organized as follows. Section 2 lays out the proposed model and introduces the necessary notation. Section 3 presents the estimation and filtering procedure. Section 4 discusses a Monte Carlo study. Section 5 presents the empirical study based on a large dataset and section 6 concludes.

2.2 Time-varying factor-augmented vector autoregressive models

Let \( y_t \) be the \( 1 \times n \) target variables, and \( x_t \) a \( 1 \times N \) dimensional set of macroeconomic variables and \( N \gg n \). \( T \) is the last available time-point in the estimation window. The data are assumed to be zero-mean stationary variables. The dynamic factor model is given by,

\[
x_t = \Lambda f_t + \varepsilon_t, \quad \varepsilon_t \sim N(0_N, H)
\]  

(2.1)

where \( f_t \in \mathbb{R}^k \) is a vector of \( k \) latent factors, \( \Lambda \) is a matrix of factor loadings, \( \varepsilon_t \in \mathbb{R}^N \), and \( H \) is a \( N \times N \) diagonal covariance matrix. We assume that all correlation through time is modeled via time dependence in the factors, and \( \text{cov}(\varepsilon_t, \varepsilon_u) = 0, \forall t \neq u \). Moreover, it is assumed that \( T, N \to \infty \). The target variables \( y_t \) are placed in the state-equation together with the latent factors,

\[
\begin{bmatrix}
  f_{t+1} \\
  y_{t+1}
\end{bmatrix} = T_t^{(1)} \begin{bmatrix}
  f_t \\
  y_t
\end{bmatrix} + \cdots + T_t^{(p)} \begin{bmatrix}
  f_{t-p} \\
  y_{t-p}
\end{bmatrix} + \eta_t, \quad \eta_t \sim N(0_{k+n}, Q_t),
\]  

(2.2)

where \( T_t^{(i)} \) is a \( k+n \times k+n \) matrix of autoregressive coefficients, \( Q_t \) is the \( k+n \times k+n \) covariance matrix, and \( \eta_t \in \mathbb{R}^{k+n} \). Since \( x_t \) and \( y_t \) are assumed to be zero-mean variables no intercept is included in (2.2). Furthermore, the innovations in the measurement \( \varepsilon_t \) and in the transition \( \eta_t \) are uncorrelated random variables. Without loss of generality we assume that the target variable \( y_t \) is a scalar.

The state-equation parameters \( T_t^i \) and \( Q_t \) are allowed to be time-varying in a stochastic way, thus they are a function of idiosyncratic shocks. For the conditional
mean time-varying parameters we specify random walk processes,

\[ T_t^{(i)} = T_{t-1}^{(i)} + v_t^{(i)}, \quad v_t^{(i)} \sim N(0_{k+1}, \Sigma_v^{(i)}), \quad \forall i = 1, \ldots, p. \quad (2.3) \]

To avoid the proliferation of static parameters the matrices \( \Sigma_v^{(i)} \) are assumed to be diagonal. The choice of random walk processes requires fewer static parameters that need to be estimated and it is a common assumption in macroeconometrics (see for example Cogley and Sargent, 2005). For the covariance of the states and the target \( Q_t \), a diagonal matrix is assumed,

\[
\begin{pmatrix}
\text{e}^{(h_{1,t})} & 0 & \cdots & 0 \\
0 & \text{e}^{(h_{2,t})} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \text{e}^{(h_{k+1,t})}
\end{pmatrix}
\]

where \( h_{i,t} \) is the log-volatility and is also modeled as a random walk,

\[ h_t = h_{t-1} + w_t, \quad w_t \sim N(0, \Sigma_w), \quad \Sigma_w = \text{diag}(\sigma_{h,i}^2) \quad \forall i = 1, \ldots, k+1. \quad (2.4) \]

The diagonal covariance matrix \( Q_t \) implies that the relationship between the target and the factors are capture only by the time-varying parameters. Although this assumption may seem restrictive it is fairly common in factor analysis as the factors are often assumed to be orthogonal between each other (see for example Eickmeier et al., 2015).

Moreover, following standard practice in TVP-VAR models we decided not to impose stationarity restrictions in the time-varying autoregressive coefficients \( T_t^{(i)} \). The main issue is related to the random walk assumption, which implies that these parameters may reach any upper and lower bound when \( T \to \infty \). Although theoretically possible, explosive paths for \( T_t^{(i)} \) are not expected as the data is assumed stationary by construction. Therefore, despite the fact that temporary non-stationarity may happen, in the long-run the estimated TVP-FAVAR will follow the data resulting in stationary paths for \( f_t \).

However, it is not feasible to estimate the log-likelihood function of the model discussed above. There are two main complications. First, the factors \( f_t \) and the time-varying parameters in the \( T_t \) matrix cannot be jointly identified as both are allowed to change over time. We therefore replace the unobserved factors \( f_t \) by a consistent estimate \( \hat{f}_t \) to form a feasible likelihood function. The second issue is related to the presence of stochastic volatility (SV) in the FAVAR equation. Given that the model delivers a nonlinear state space representation, the likelihood function cannot be expressed in the form of the prediction error decomposition computed by the Kalman

\[ \text{Note that a stationary law of motion for } T_t^{(i)} \text{ does not eliminate the possibility of temporary non-stationarity and it implies a large number of static parameters to be estimated. Furthermore, as the data is stationary by construction the estimated TVP-VAR will generate stationary paths in the long-run.} \]
Filter. To tackle this nonlinear state-space estimation we propose an estimation procedure based on the NAIS method of Koopman et al. (2015) together with the Rao-Blackwellization step proposed by Moura and Turatti (2014) to efficiently construct an importance sampling approximation to the likelihood function. Although simulation-based methods are still required, the estimation procedure takes advantage of the structure of the model and it simplifies to a sequence of Kalman filters and weighted least squares problems.

### 2.3 Estimation and Filtering

In this section we develop our estimation and filtering procedure for the TVP-FAVAR model. The methodology is based on the two-step approach originally developed by Stock and Watson (2002a). In the first step the factors are consistently estimated by principal components and treated as observable variables forward in the estimation. In the second step the factor-augmented equations are estimated conditional on these estimates as independent time-varying autoregressions. The factor model we consider is,

\[ x_t = \Lambda f_t + \epsilon_t, \quad \epsilon_t \sim N(0, H). \]  

(2.5)

A consistent estimate of space spanned by the factors can be obtained by the principal components estimator (PCA). The PCA estimator solves the following the minimization problem,

\[ \hat{f} = \min_f (N)^{-1} \sum_{i=1}^{T} \sum_{t=1}^{T} (x_{i,t} - \lambda_i f_t)^2, \]  

(2.6)

where \( \lambda_i \) is the \( i^{th} \) row of the matrix \( \Lambda \). To uniquely identify the minimizers, it is necessary to impose identifying restrictions on the estimators. By concentrating out \( \Lambda \) and using the normalization \( f' f / T = I_k \) the problem is equivalent to maximizing \( \text{tr}(f'(xx')f) \). The resulting estimator \( \hat{f} \) is given by \( \sqrt{T} \) times the eigenvectors corresponding to the \( k \) largest eigenvalues of the \( T \times T \) matrix \( xx' \). The solution is not unique, as any other orthogonal rotation of \( \hat{f} \) is also a solution. However, we treat the PCA factors as observable which is justified only when \( N \) grows faster than \( T^{0.5} \).

As shown in Stock and Watson (2002a), when \( N \) is large and the number of principal components used is at least as large as the true number of factors, the principal components consistently recover the space spanned by \( f_t \).

In the second step the PCA factors \( \hat{f} \) are used to estimate the time-varying vector autoregressions. Given that the latent factors \( f_t \) are replace by their consistent estimator \( \hat{f}_t \), and the time-varying matrix \( Q_t \) is diagonal, the measurement densities of the model equations are conditionally independent. Furthermore, we will show that the likelihood function of the whole TVP-FAVAR can be factorized as the product of each equation likelihood function, meaning each equation can be estimated separately. Let \( \theta_{i,t} = (T_{i,t}, h_{i,t})' \) be the state vector of all time-varying parameters at time \( t \) for
where each sampler is distributed as $g(\tilde{y}_t, \theta; \gamma_i)$. State-space models are usually formulated in terms of a measurement and a transition density.

$$p(\tilde{y}_t, \theta; \gamma_i) = \prod_{t=1}^{T} p(\tilde{y}_t, \theta; \gamma_i)p(\theta|\theta_{t-1}; \gamma_i), \quad (2.7)$$

where $p(\tilde{y}_t, \theta; \gamma_i)$ is the joint density of observable and state variables, and $\gamma_i$ is the vector of static parameters for equation $i$. The joint likelihood for the TVP-FAVAR can be written as,

$$L(\gamma, \tilde{y}) = \int \cdots \int p(\tilde{y}_1, \theta_1; \gamma_1) \cdots p(\tilde{y}_{k+1}, \theta_{k+1}; \gamma_{k+1}) d\theta_1 \cdots d\theta_{k+1}, \quad (2.8)$$

Note that the joint density can be decomposed as the product of the marginal densities as the equations are conditionally independent.

The likelihood (2.8) is a very high-dimensional analytically intractable function. To approximate this complicated integral we introduce the multivariate Gaussian importance sampler $g(\theta|\tilde{y}, \varphi)$. Following Richard and Zhang (2007) this multivariate sampler is itself high-dimensional and it can be decomposed into a sequence of low-dimensional importance samplers. Given the representation of a state-space model as the product of a measurement and a transition density, the joint density of latent states over time is given by,

$$p(\theta|\tilde{y}, \gamma) = \prod_{t=1}^{T} p(\theta|\theta_{t-1}, \tilde{y}_t, \gamma). \quad (2.9)$$

The sampler $g(\theta|\tilde{y}, \varphi)$ is partitioned conformably as,

$$g(\theta|\tilde{y}, \varphi) = \prod_{t=1}^{T} g(\theta|\theta_{t-1}, \tilde{y}_t, \varphi_t), \quad (2.10)$$

where each sampler is distributed as $g(\theta|\theta_{t-1}, \tilde{y}_t, \varphi_t) \sim N(\mu_{t,g}, V_{t,g})$, $\mu_{t,g}$ is a $l$-dimensional vector and $V_{t,g}$ is $l \times l$ covariance matrix, $l = p(k+1)^2 + k + 1$, and $\varphi_t$ are the importance parameters. Using this importance sampler the likelihood function for the TVP-FAVAR becomes,

$$L(\gamma, \tilde{y}) = \int \cdots \int \left[ \frac{p(\tilde{y}_1, \theta_1; \gamma_1) \cdots p(\tilde{y}_{k+1}, \theta_{k+1}; \gamma_{k+1})}{g(\theta|\tilde{y}, \varphi)} \right] g(\theta|\tilde{y}, \varphi) d\theta_1 \cdots d\theta_{k+1}. \quad (2.11)$$

Note that although the measurement equations are conditionally independent the importance sampler $g(\theta|\tilde{y}, \varphi)$ introduces interdependence between all states across equations, thus (2.11) cannot be simplified without additional assumptions. However, it is reasonable to expect that the interdependence between states captured by the importance sampler should be very small as the natural sampler $p(\theta|\theta_{t-1}, \tilde{y}_t, \gamma)$ itself has a diagonal covariance matrix. This means that the state variables are uncorrelated and the importance sampler should reflect the DGP over time. Thus, we impose that
the covariance matrix $V_{t,g}$ is diagonal, and the joint importance sampler $g(\theta|\tilde{y}, \varphi)$ can be rewritten as,

$$g(\theta|\tilde{y}, \varphi) = \prod g(\theta_1|\tilde{y}_1, \varphi_1) \cdots g(\theta_{k+1}|\tilde{y}_{k+1}, \varphi_{k+1}),$$  \hspace{1cm} (2.12)

where each sampler is distributed as $g(\theta_1, t|\theta_{1,t-1}, \tilde{y}_{1,t}, \varphi_{1,t}) \sim \mathcal{N}(\mu_{1,t,g}, V_{1,t,g})$, $\mu_{1,t,g}$ is a $l^*$-dimensional vector, $V_{1,t,g}$ is $l^* \times l^*$ covariance matrix, and $l^* = p(k + 1) + k + 1$.

Given this decomposition of the importance sampler the likelihood function [2.11] can be rewritten as,

$$L(\gamma, \tilde{y}) = \int \cdots \int \left[ \frac{p(\tilde{y}_1, \theta_1; \gamma_1) \cdots p(\tilde{y}_{k+1}, \theta_{k+1}; \gamma_{k+1})}{\prod g(\theta_1|\tilde{y}_1, \varphi_1) \cdots g(\theta_{k+1}|\tilde{y}_{k+1}, \varphi_{k+1})} \right] \times$$

$$\times \prod g(\theta_1|\tilde{y}_1, \varphi_1) \cdots g(\theta_{k+1}|\tilde{y}_{k+1}, \varphi_{k+1})\, d\theta_1 \cdots d\theta_{k+1}. \hspace{1cm} (2.13)$$

Rearranging the integrand we find that,

$$L(\gamma, \tilde{y}) = \int \left[ \frac{p(\tilde{y}_1, \theta_1; \gamma_1) g(\theta_1|\tilde{y}_1, \varphi_1)}{g(\theta_1|\tilde{y}_1, \varphi_1)} \right] d\theta_1 \cdots \int \left[ \frac{p(\tilde{y}_{k+1}, \theta_{k+1}; \gamma_{k+1}) g(\theta_{k+1}|\tilde{y}_{k+1}, \varphi_{k+1})}{g(\theta_{k+1}|\tilde{y}_{k+1}, \varphi_{k+1})} \right] d\theta_{k+1}, \hspace{1cm} (2.14)$$

The joint likelihood function [2.14] is now a product of individual likelihood functions. Thus, we can treat each equation separately in the estimation procedure. Note, however, that [2.14] is not the same as [2.11], as the covariance matrix of the joint sampler [2.10] is almost surely not diagonal, therefore some information loss may happen when imposing that its covariance matrix is diagonal. However, given the structure of the model this efficiency loss should be outperformed by the increasing numerical and computational tractability of an equation-by-equation estimation. Note that the results in Geweke (1989) apply and the joint importance sampler [2.12] delivers consistent estimators for the likelihood function. We perform a Monte Carlo study to determine the potential efficiency loss implied by the diagonal form of the importance sampler chosen in this paper. Results suggest that the equation-by-equation procedure yield very similar average parameter estimates as the joint estimation of all parameters.

We proceed by developing an estimation procedure for each equation of the TVP-FAVAR. This requires to estimate $k + 1$ equations with $p(k + 1) + 1$ state variables each. However, the system of equations can still be further simplified reducing the dimension of the state vector. We convert this still multivariate non-linear state-space model by marginalizing all the time-varying parameters that enter linearly in the likelihood function of each equation. This can be accomplished by applying a conditional Kalman filter. Finally, to tackle this now low-dimensional likelihood estimation we develop a fast importance sampling (IS) technique based on the NAIS method of Koopman et al. (2015). Let $\theta_{i,t} = (T_{i,t}, h_{i,t})'$ be the state vector of all time-varying parameters at time $t$ for equation $i$, and $\tilde{y}_t = (\hat{f}_t, y_t)'$ be the vector of observables. The likelihood function for equation $(i)$ is the integral over the joint density of latent
CHAPTER 2. CLASSICAL ANALYSIS OF TIME-VARYING FACTOR-AUGMENTED VECTOR AUTOREGRESSIONS

states $\theta_{i,t}$ and observables $\tilde{y}_{i,t},$

$$L(\gamma, \tilde{y}) = \int \prod_{t=1}^{T} p(\tilde{y}_{i,t}, \theta_{i,t}; \gamma) d\theta_{i,t},$$

(2.15)

where $\gamma$ is the vector of static parameters. The likelihood (2.15) can be decomposed as the product of the measurement and the transition densities,

$$L(\gamma, \tilde{y}_{i}) = \int \prod_{t=1}^{T} p(y_{i,t}^{*}|\theta_{i,t}; \gamma) p(\theta_{i,t}|\theta_{i,t-1}; \gamma) d\theta_{i,1} \ldots d\theta_{i,T},$$

(2.16)

The likelihood function (2.16) is a $p(k+1) + 1 \times T$ analytically intractable integral. Although it is a very high dimensional integration problem and it cannot be evaluated by standard numerical methods, we show it is possible to significantly reduce the dimension of this integral and construct an efficient importance sampling approximation.

Initially, observe that the integration over the density $p(y_{i,t}^{*}|\theta_{i,t}; \gamma) p(\theta_{i,t}|\theta_{i,t-1}; \gamma)$ can be decomposed into two terms: an analytical and an unknown term. In fact it is possible to exploit the structure of the model to reduce the computational cost even further. Note that given the stochastic volatility the time-varying parameters in $T_{t}^{(i)}$ can have their moments analytically estimated via the Kalman Filter and Smoother (KFS). Therefore, the model belongs to the class of conditionally linear and Gaussian state-space models. This procedure is usually called Rao-Blackwellization, and it always reduces the Monte Carlo (MC) variation of the likelihood estimator. Therefore, conditional on the log-volatility $h_{i,t}$, equation (2.16) can be rewritten as,

$$L(\gamma, \tilde{y}_{i}, t) = \int \prod_{t=1}^{T} p^{*}(y_{i,t}^{*}|h_{i,t}; \gamma) p(h_{i,t}|h_{i,t-1}; \gamma) dh_{i,1} \ldots dh_{i,T},$$

(2.17)

where the density $p^{*}(y_{i,t}^{*}|h_{i,t}; \gamma)$ can be obtained by the Kalman filter conditional on a particular value of $h_{i,t}$. Note that the likelihood function (2.17) is now $1 \times T$, thus a large dimension reduction while at the same time increasing efficiency. The density $p^{*}(y_{i,t}^{*}|h_{i,t}; \gamma)$ is evaluated by the prediction error decomposition for a specific draw of $h_{i,t},$

$$p^{*}(y_{i,t}^{*}|h_{i,t}; \gamma) = p(v_{t}|h_{i,t}, \gamma) = N(0, F_{t}),$$

(2.18)

where $v_{t}$ is the prediction error and $F_{t}$ its associated variance. Finally, it is still necessary to obtain an approximation for the low-dimensional likelihood function equation (2.17), and we apply an importance sampling approximation based on the NAIS method of Koopman et al. (2015).

\[\text{This result is an implication of the Rao-Blackwell theorem. For more details see Robert and Casella (2005).}\]

\[\text{The intuition behind this result is that we replace the IS estimator for } T_{t} \text{ by the KFS estimator. As the KFS is an analytical filter its MC variance is zero, thus always smaller than the IS filter.}\]
2.3.1 Rao-Blackwellized Numerically Accelerated Importance Sampling

The key criteria for our importance sampling strategy is minimizing the distance between the unknown likelihood function and its IS estimator. The NAIS procedure together with a Kalman filter delivers a numerical tool for sampling and estimation that can be carried out by standard state-space methods. We aim to show that the sampling from the proposal density can be done by the Kalman filter and the selection of the importance density is based on a least square problem. Initially, it is necessary a Gaussian importance sampling

\[ g(h|\tilde{y};\varphi) = \frac{g(\tilde{y}|h;\varphi)g(h;\varphi)}{g(\tilde{y};\varphi)}, \]

(2.19)

where \(\varphi\) is the parameter vector of the importance density \(g(h|\tilde{y};\varphi)\). \(g(h;\varphi), g(\tilde{y};\varphi)\) are all Gaussian densities, and \(g(\tilde{y};\varphi)\) can be understood as a normalizing constant.

In light of the discussion in the first chapter it is possible to rewrite the likelihood function given some regularity conditions (see Geweke, 1989),

\[ L(\gamma, \tilde{y}) = g(\tilde{y}) \int \omega(h, \tilde{y})g(h|y, \varphi)dh. \]

(2.20)

The importance weights \(\omega(h, \tilde{y})\) are defined as,

\[ \omega(\gamma, y) = \frac{p^*(y|h)}{g(y|h, \varphi)}, \]

(2.21)

evaluated at \(S\) independent trajectories of the non-linear signal vector \(h\). The likelihood estimator is given by,

\[ \hat{L}(\gamma, \tilde{y}) = g(\tilde{y})*\tilde{\omega}(\gamma, y), \tilde{\omega}(\gamma, y) = \frac{1}{S} \sum_{s=1}^{S} \omega(h^s, \tilde{y}). \]

(2.22)

To evaluate the likelihood estimator (2.22) it is necessary a procedure to define the auxiliary parameters \(\varphi = \{\varphi_t\}_{t=1}^T\) and a sampling scheme from the proposal density. We apply the NAIS method of Koopman et al. (2015), where the auxiliary parameters are efficiently selected and the sampling procedure is based on standard methods in the state-space literature. The main result of NAIS is the representation of \(g(\tilde{y}_t|h_{i,t}, \varphi_t)\) as a linear and Gaussian state space density given a set of importance parameters \(\varphi_t = (b_t, c_t)\),

\[ y^*_t = h_{i,t} + \epsilon_t, \quad \epsilon_t \sim N(0, 1/c_t), \]

(2.23)

\[ y^*_t = b_t/c_t. \]

(2.24)

Note that (2.23) is a measurement equation for the importance density model, and it implies a linear and Gaussian state-space model for the log-volatility \(h_{i,t}\). Therefore, the moments of the importance sampler \(g(h|\tilde{y})\) can be computed by the Kalman filter.
and smoothing, and the sampling from $g(h|\tilde{y})$ can be carried out by the simulation smoother of Durbin and Koopman (2002). This is possible as the smoothing density $g(h_{i,t}|\tilde{y}_{i,t},\varphi_t)$ is available analytically and is given by,

$$g(h_{i,t}|\tilde{y}_{i,t},\varphi_t) = \frac{1}{\sqrt{2\pi V_t}} \exp \left( -0.5 \frac{(h_{i,t} - \tilde{h}_t)^2}{V_t} \right), \quad (2.25)$$

where $\tilde{h}_t$ and $V_t$ are the mean and the variance of the log-volatility process $h_{i,t}$, and they are obtained by the KFS applied to the linear state-space (2.23) for a given $\varphi = \{\varphi_t\}_{t=1}^T$.

However, to sample from (2.25) it is necessary to select the values of the auxiliary parameters $\varphi$. In this paper we apply the criteria originally proposed by Richard and Zhang (2007), in which the proposal density is selected to mimic as close as possible the likelihood integral. Richard and Zhang (2007) showed that this is equivalent to solving for every time $t$,

$$\arg\min_{\varphi_t} = \int \lambda^2(\tilde{y}_t, h_{i,t}|\varphi_t)\omega(h_{i,t}, \tilde{y}_t|\varphi_t)g(h_{i,t}|\tilde{y}_t,\varphi_t) \, dh_{i,t}, \quad (2.26)$$

with $\lambda(.)$ being defined as,

$$\lambda(\tilde{y}, h_{i,t}|\varphi) = (\log p^*(\tilde{y}_t|h_{i,t}) - \log g(\tilde{y}_t^*|h_{i,t},\varphi)). \quad (2.27)$$

Equation (2.26) needs to be numerically implemented in a feasible manner as the sampler itself is a function of the auxiliary parameters $\varphi_t$. Richard and Zhang (2007) sequentially applied Monte Carlo methods to approximate (2.26) using draws from their intermediate efficient importance sampler until $\varphi = \{\varphi_t\}_{t=1}^T$ satisfy a convergence criteria. Koopman et al. (2015) refine this approach and consider numerical integration via Gauss-Hermite (GH) methods using intermediate samplers until convergence. Note that the dimension of (2.26) is the dimension of $h_{i,t}$, in this case a one-dimensional integration problem. Numerical methods for a one-dimensional integration are very fast and almost exact procedures. An important aspect of the Gauss-Hermite integration is that only a very small number of draws (nodes) is necessary to efficiently evaluate (2.26). For instance, typical applications only require 10 to 15 nodes.

Let $\{z_i\}_{i=1}^M$ be the set of $M$ predefined GH nodes and $h(z_i)$ be the respective Gauss-Hermite weights. We define the set of abcissae $\tilde{z}_{i,t} = \tilde{h} + \sqrt{V_t}z_i$. Thus, the minimization (2.26) can be rewritten as a least square problem,

$$\log p^*(\tilde{y}_t|\tilde{z}_{i,t}) = \text{constant} + \kappa \tilde{z}_{i,t} - \frac{1}{2} \xi \tilde{z}_{i,t}^2 + \text{error}, \quad (2.28)$$

with weights $w_{i,t} = h(z_i)\exp(\frac{1}{2} z_i^2)$, as the nodes are assumed to be associated with the standard normal distribution. The weighted least squares estimates for $\kappa$ and $\xi$ are the new estimates of $b_t$ and $c_t$ respectively. From these new estimates it is possible to
obtain a new linear representation in equation 2.23 for the stochastic volatility and draw from its Gaussian density. A comprehensive treatment of the RB-NAIS method can be found in the first chapter of this thesis.

The RB-NAIS algorithm is given by the following steps,

1. Initialize the procedure with starting values for \( b_t = 0 \) and \( c_t = 1 \) for each time \( t = 1 \ldots T \). The choice of the initial values is usually not necessary for convergence.

2. Compute smoothed mean \( \tilde{h}_t \) and variance \( V_t \) for each time \( t = 1 \ldots T \) based on the current values of \( \varphi_t = (b_t, c_t) \) and the state-space representation 2.23 via the KFS.

3. Evaluate the density \( p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma_i) \) by the prediction error decomposition. Minimize equation 2.26 and obtain a new set of \( \varphi_t = (b_t, c_t) \) for every time \( t \).

4. Compute a criteria for convergence. Run steps 2 and 3 until convergence is achieved or preset a sufficient number of iterations.

5. Given the final set of \( \varphi = \{(b_t)_{t=1}^T, \{C_t\}_{t=1}^T\} \), sample \( S \) independent trajectories for the log-volatility via the simulation smoother of Durbin and Koopman (2002). Evaluate the density \( p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma_i) \) by the prediction error decomposition conditional on the \( S \) independent trajectories for the log-volatility.

6. Compute the estimator for the likelihood function equation 2.22.

### 2.3.2 Filtering and Smoothing

The set of static parameters are estimated by our two-step simulated maximum likelihood scheme for each equation. The time-varying parameters \( T_i^{(i)} \) and the log-volatility \( h_{i,t} \) need to be estimated as well. An estimator can be constructed based on the importance sampling procedure. For example, let \( h_{i,t} \) be the log-volatility for equation \( i \), an estimator for the variance \( \mathbb{E}(\exp(h_{i,t})) \) can be written as,

\[
\mathbb{E}(\exp(h_{i,t})) \approx \frac{\sum_{j=1}^S \omega_j(y, y_t) \exp(h_{i,t}^{(j)})}{\sum_{j=1}^S \omega_j(y, y_t)} \tag{2.29}
\]

where \( \omega_j(y, y_t) \) are weights as defined in 2.21. Note that the estimator 2.29 yields smoothed estimates using information from the whole sample. The consistency of 2.29 holds under the weak law of large numbers when \( h_{i,t} \) are sampled from the importance density \( g(.) \) independently and as long as the variance of the importance weights are bounded and, therefore, the importance sampling procedure is well-behaved for each equation \((i)\). Finally, for convergence under the weak law of large numbers it is necessary that \( T, S, N \to \infty \).
The filtering process for the time-varying parameters $T_t^{(i)}$ can also be carried out by a similar procedure as for the stochastic volatilities. However, this would imply that the smoothed matrices $\hat{T}_t^{(i)}$ would not necessarily satisfy stability restrictions and could become explosive. The main issue is related to the random walk assumption for the time-varying parameters in the FAVAR equation. Note that we can disregard long-run non-stationarity as the data is stationary by construction, nevertheless, temporary non-stationarity may still be present. Therefore, we devise the following scheme to check for the stationarity condition at each time $t$, and discard the estimates that deliver explosive parameters. The procedure is based on a multivariate Kalman filter conditional on the IS estimates of the stochastic volatilities.

1. Fix the static parameters with their MLE estimate and the stochastic volatility with their IS estimate.
2. Initialize the time-varying autoregressive coefficients with a diffuse initialization or the OLS estimates for a time-invariant FAVAR.
3. Perform the prediction step $\hat{T}_{t|t-1}^{(i)}$ and update step $\hat{T}_{t|t}^{(i)}$ of the Kalman filter.
4. Check for stability of the matrix $\hat{T}_{t|t}^{(i)}$. If the VAR structure satisfies the non-explosiveness condition proceed to time $t+1$.
5. On the other hand, if the matrices $\hat{T}_{t|t}^{(i)}$ do not satisfy stability restrictions set $\hat{T}_{t|t}^{(i)} = \hat{T}_{t-1|t-1}^{(i)}$ and do not perform the update step.

Note that if no update step is performed it does not mean that the $\hat{T}_t^{(i)}$ will be stuck at $\hat{T}_{t-1|t-1}^{(i)}$, but with a new observation $\tilde{y}_{t+1}$ an update can be performed in a further point in time (Eickmeier et al., 2015).

The algorithm outlined in the previous steps yields a well-defined procedure to estimate and filter the TVP-FAVAR model. In addition, given the properties of the model and the numerical integration step for equation (2.26) the estimation is faster than most existing approaches in the literature for model of similar complexity. Finally, the results in Geweke (1989) apply and the simulated maximum likelihood converges to the true one as $S, T, N \to \infty$. In the next section we present a simulation exercise to verify the estimation properties of the equationwise procedure.

### 2.4 Simulation Study

We conduct two MC studies to assess the performance of the estimation procedure. In the first study we investigate the finite sample properties of the estimation procedure. We simulate a state-space system according to equations (2.1)-(2.2) where the state-vector follows a three dimensional stationary VAR(1). The first two components of the state-vector $f_{t,1}$ and $f_{t,2}$ are treated as latent factors and the third factor $f_{t,3}$ is regarded as the observable target. The simulated data are imposed to be stationary.
by checking the stability of the matrix $T_t$ at every time $t$ and dropping those that
do not satisfy this condition. We fix the matrix $\Lambda$ as random draws from a standard
normal distribution, the matrix $H = 0.1 I_{200}$, and $N = 200$. For the static parameters
$\sigma^2_{i,j}$ and $\sigma^2_{h,i}$ we follow Eickmeier et al. (2015) and select two different specifications.
Let $\sigma^2_{i,j}$ be the variance of the time-varying parameters $T_{i,j}$ when $i \neq j$, and $\sigma^2_{i,i}$ is
the variance of the time-varying parameters $T_{i,i}$ if $i = j$. Table (2.1) summarizes the
considered data generating processes (DGP).

<table>
<thead>
<tr>
<th>DGP</th>
<th>$\sigma_{i,i}$</th>
<th>$\sigma_{i,j}$</th>
<th>$\sigma_{h,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>II</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Case I is the one with high parameter variation. Case II is of lower variation intensity.
For both cases the time dimension is 250 and 500, and 750. The choices of sample
size are intended to reflect the usual availability of large macroeconomic datasets,
thus our objective is to study the behavior of the estimator in finite samples. The
larger sample size ($T = 750$) aims to verify asymptotical properties.

The second MC study aims to verify how the diagonal assumption for the covariance
matrix of the importance sampler affects the average parameter estimates. In
this study we simulate a state-space system according to equations (2.1)-(2.2) where
the state-vector follows a two dimensional stationary VAR(1). The first component
of the state-vector $f_{t,1}$ is treated as latent factor and the second $f_{t,2}$ is regarded as
the observable target. As before the simulated data are imposed to be stationary by
checking the stability of the matrix $T_t$ at every time $t$. We fix the matrix $\Lambda$ as random
draws from a standard normal distribution, the matrix $H = 0.1 I_{200}$, and $N = 200$. We
center the same DGPs as in the first study with sample sizes $T = 250$ and $T = 500$.

For each MC study and sample size we standardize $x_t$, $t = 1, \ldots, T$ and apply the PC
analysis to extract the estimate of the factors $\hat{f}$. However, the estimation results using
$\hat{f}$ are not directly comparable to the true DGP, as the true factors $f$ are not identified.
Therefore, in order to compare the DGP and the simulated maximum likelihood
estimates we change the principal components estimator orthogonal rotation to
resample the simulated factors. Specifically, we solve for the orthogonal $k \times k$ matrix
$A^*$ that maximizes $\text{tr}\{\text{corr}(f, \hat{f}(A))\}$ \footnote{The solution to this is $A^* = VU'$, where $V$ and $U$ are the orthogonal matrices of the singular value decomposition $\text{corr}(f, \hat{f}) = USV'$.} The PC factors are then rescaled to have the
same standard deviation of the simulated factors $f_t$,

$$\tilde{f}_i = \frac{\sigma(f_{i1})}{\sigma(\hat{f}_{i1})} \hat{f}_{i1}, \quad i = 1, \ldots, k \quad (2.30)$$
where $\hat{f}_i^*$ is the $i^{th}$ column of the rotated principal components $\hat{f} A^*$. The rotated and rescaled principal components $\tilde{f}_i$ are treated as data, and used to form a likelihood function. This rotation allows us to directly compare the estimated parameter values with the data generating process.

### 2.4.1 Results

We obtain two results from the first model. The first one presents mean and sample standard deviations of the estimates in each scenario for the third factor $f_{t,3}$, which is treated as observable variable. In the second experiment we investigate how suitable is the PC estimator for the latent real factors $f_t$ when the DGP is characterized by stochastic volatility and time-varying parameters. The rotated PC estimator and the simulated factors are compared on the basis of their degree of correlation. The model is estimated by the equationwise NAIS procedure with 200 draws and 10 GH nodes to evaluate the likelihood function. The time-varying parameters $T_t$ and $h_t$ are initialized by a diffuse procedure. Table (2.2) (in appendix) shows the results of the first experiment. We note that the parameter estimates are closer to the DGP values in the larger sample size ($T = 750$). Nevertheless, the estimates are sufficiently even in the case $T = 250$, although the stochastic volatility parameter seems to be overestimated. When $T = 750$ all parameters are closer to the DGP and the sample standard deviations become smaller. It is observed that the parameter in the stochastic volatility process is somewhat slower to converge. Finally, no difference is observed in terms of the estimates for DGP I and DGP II.

In the second result from the first MC study we aim to determine whether the rotated estimated factors $\tilde{f}_i$ are a suitable substitute for the simulated path $f_i$.\footnote{The same strategy for the MC study was used by Eickmeier et al. (2015) and Mikkelsen et al. (2015).} It is important to verify if the PC estimator is able to properly estimate the factor structure even in the presence of stochastic volatility and time-varying parameters. We evaluate the fitness of the PC estimator in terms of the degree of correlation, i.e. $\text{corr}(\tilde{f}_1, f_1)$ and $\text{corr}(\tilde{f}_2, f_2)$. The results are reported in table (2.3) (in appendix). As before, the sample sizes are $T = 250$, $T = 500$, and $T = 750$. Results indicate that the correlation between the simulated factors and their estimated counterparts is very high. The median correlation across the DGPs and sample sizes is at least 0.96, and the 5%-quantile is not lower than 0.88. Furthermore, the median correlation and the lower quantile seem to increase with the sample size, indicating that the PC estimator is consistent. The results suggest that the PC estimator is able to recover the underlying factor structure in a very sensible way even if the factor process is characterized by stochastic volatility and time-varying parameters.

The second MC study is intended to verify if the estimation results are affected by diagonal covariance matrix imposed on the joint importance sampler $g(\theta|\tilde{y}, \varphi)$.\footnote{Note that the PCA factors depend on a particular orthogonal rotation, thus they can only be compared to the real factors after changing the rotation and scale.}
In this study we focus on a TVP-FAVAR(1) with 2 factors, where $f_{t,2}$ is regarded as the observable target. The model is estimated by the equation-wise NAIS procedure for the observable target $f_{t,2}$ with 10 GH nodes and 200 draws to evaluate the likelihood function. Moreover, the joint estimation of the TVP-FAVAR system is carried out by the multivariate RB-NAIS using 10 GH nodes and 400 draws to evaluate the likelihood. As before the time-varying parameters $T_t$ and $h_t$ are initialized by a diffuse procedure. Results for observable factor $f_{t,2}$ are presented in table (2.4) (in appendix). We note that the equationwise procedure and the joint method yield very similar average parameter estimates and standard deviations across all DGPs and sample sizes. This result suggests that the loss of efficiency on imposing a diagonal covariance matrix for the importance sampler is very small, and it can be justified by the computational advantages of estimating several low-dimensional TVP regressions.

Therefore, we have found MC evidence that the estimation procedure is able to deliver good estimate of the model parameters even in small samples sizes. Additionally, the PC estimator can recover the factor structure when its dynamics are driven by time-variation and stochastic volatility. Finally, the second MC study suggests that the parameter estimates from the equationwise procedure are comparable to the joint estimation of the TVP-FAVAR system. The next section presents an empirical application to macroeconomic forecasting with a large dataset.

### 2.5 Empirical Application: Macroeconomic Forecasting with Large Datasets

In this empirical section we aim to assess the forecasting performance of the TVP-FAVAR model when compared to other models commonly employed in the macroeconomics literature. The following models are selected,

- **Model 1**: Time-Varying Factor-Augmented Vector Autoregression with Stochastic Volatility (TVP-FAVAR).

  \[
  \begin{bmatrix}
  f_{t+1} \\
  y_{t+1}
  \end{bmatrix}
  =
  \begin{bmatrix}
  f_t \\
  y_t
  \end{bmatrix}
  + \cdots +
  \begin{bmatrix}
  f_{t-p} \\
  y_{t-p}
  \end{bmatrix}
  + \eta_t,
  \quad \eta_t \sim N(0, Q), \tag{2.31}
  \]

- **Model 2**: Time-Invariant FAVAR model of Bernanke et al. (2005).

- **Model 3**: Local-level model,

  \[
  y_t = \mu_t + \sigma_{\varepsilon} \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \tag{2.32}
  \]

  \[
  \mu_t = \mu_{t-1} + \sigma_{\eta} \eta_t, \quad \eta_t \sim N(0, 1). \tag{2.33}
  \]
Chapter 2. Classical Analysis of Time-Varying Factor-Augmented Vector Autoregressions

- Model 4: First-order autoregressive process,

\[ y_t = \phi_0 + \phi_1 y_{t-1} + \sigma \varepsilon_t, \quad \varepsilon_t \sim N(0, 1). \] (2.34)

We choose models that incorporate information from large datasets and two simple univariate models. The time-invariant FAVAR is intended to compare whether time-varying parameters and SV can improve forecasts. The FAVAR is estimated by an equationwise ordinary least squares conditional on the factor estimates given by the PC estimator. The local-level and the AR(1) are standard benchmark models for univariate forecasting. The hypothesis is that using a large dataset should improve on univariate models. We present two forecasting studies on different datasets. The first one simulates two series according to DGP I and II from the previous section. The idea is to verify whether our proposed estimation procedure is able to outperform simpler models when the DGP is characterized by stochastic volatility and time-varying parameters.

The second forecasting study is an empirical application using the Jurado, Ludvigson and Ng dataset as used in Jurado et al. (2015). The dataset is composed of 132 variables, where 3 incomplete variables were excluded. We applied the same transformation as in Jurado et al. (2015) to achieve stationarity. The data runs from March 1960 to December 2011 after transformations. A total of 624 observations are used. The data refers to the U.S. economy. We select six target variables: three price indexes: the consumer price index (CPI), the producer price index: finish goods (PPI), and the personal consumption expenditure deflator (PCEd); and three real variables: the Industrial production total (IP), the Unemployment Rate: All Workers (U), and Personal Income (PEI). We forecast each variable separately with the factors. All the variables are standardized prior to estimation and forecasting. We conduct forecasting analysis in the whole sample, and in its halves.

The number of factors is selected on the basis of the information criteria discussed in Bai and Ng (2002). We decided according to the Akaike information criteria and considering the most parsimonious specification. Although most information criteria for the selection of factors tend to choose a high number there is evidence that models using a low number of factors tend to forecast better (see Stock and Watson, 2002b; for the case of output and inflation). Motivated by this we rely on five latent factors and one lag in the TVP-FAVAR. As a check of robustness, we also carried out the forecasting analysis with six latent factors. Therefore, this means that we estimate a TVP-VAR with 6 and 7 variables including the forecasting target, which can be a computational burden. However, given the equationwise NAIS procedure the estimation window process can be carried out in standard software and hardware facilities without major difficulties.

---

8 The data are available with descriptions at [https://www.sydneyludvigson.com/data-and-appendixes/](https://www.sydneyludvigson.com/data-and-appendixes/)
Forecasts are computed for $h = 1, 2, 4, 8$ and are evaluated on the basis of the root mean square error (RMSE). We consider a rolling window estimation, and the RMSEs are compared on the basis of the Giacomini and White (2006) test. Let $S$ be number of forecasts and $R$ be the initial window size, the adopted forecasting strategy for $h$-step ahead forecasts for the TVP-FAVAR is the following for each $i = 1, \ldots, S$,

1. Estimate the system parameters starting in $T - R$ for each equation by the simulated maximum likelihood previously developed in section 3. Compute smoothed estimates of the stochastic volatilities via equation (2.29).

2. Apply a multivariate Kalman filter conditional on the MLE estimates for the static parameters and the IS estimates for the stochastic volatilities. Obtain an estimate for the $T_t$ matrix.

3. The forecast is then computed by equation,

$$\begin{bmatrix} \hat{f}_{t+h} \\ \hat{y}_{t+h} \end{bmatrix} = T_t \begin{bmatrix} \hat{f}_t \\ \hat{y}_t \end{bmatrix},$$  \hspace{1cm} (2.35)

where $\hat{y}_{t+h}$ is the forecast of the target variable $y_{t+h}$, and $t$ is the end of the estimation window. The factor estimates $\hat{f}_t$ are the PC estimates.

Note that in our model the time-varying parameters incorporate information about the target rather than the factors. Therefore, each element in $T_t$ is informed about the target forecast in a time-varying manner. This is true even if the covariance matrix of the whole vector $\text{vec}(T_t)$ is diagonal, as the filtered covariance matrix delivered by the Kalman filter is never diagonal even though the DGP specification can be. For $h$-step ahead forecasting we treat the time-varying coefficients $T_t$ as constants in time when $h > 1$. This is consistent with the idea of random walk coefficients. We follow the rule of thumb in Clark and McCracken (2011) and take a sample split ratio $\pi = \frac{S}{R}$ approximately equal to one.

### 2.5.1 Results

Results for the simulated series are presented in table 2.5 (in appendix). In this forecasting study we generated two samples from DGP I and II treating the third factor $f_3,t$ as the observable forecasting target. We select $S = 308$, $R = 300$, and $T = 608$. The correct number of factors is used (2 in this case), in a view of isolating the effects of the estimation procedure. One lag is used in the estimation. The time-varying parameters $T_t$ and $h_t$ are initialized by a diffuse procedure. Results indicate that the estimation procedure is able to properly capture the time-variation in the DGP, and the RMSEs are almost always smaller for the TVP-FAVAR. In fact they are also often statistically significant. In addition, as greater the time-variation in the DGP the better the relative RMSEs results of the TVP-FAVAR. Finally, note that there is little difference between
the TVP-FAVAR and the FAVAR when $h = 8$. This is due to the non-explosiveness restrictions imposed in the joint filtering of the parameters and the stationarity restrictions when generating the data. Therefore, the TVP-FAVAR converges to the unconditional long-term mean, which is captured by the time-invariant FAVAR.

We first analyze the empirical data by estimating a TVP-FAVAR with 5 factors in the full sample for the CPI and IP. The idea is to verify the presence of time variation in the conditional variance and in the autoregressive parameters. Figure (2.1) (in appendix) presents the time evolution of $\mathbb{E}(\exp(h_t))$ for the CPI. Results indicate a remarkable time-variation not only in the volatility of inflation but in the macroeconomic factors as well. Interestingly, all factors display time-varying volatility, some during the Great Inflation of the 1970s, others during the Financial Crisis of 2008 or in both periods. The volatility of the CPI follows closely the reported in the first chapter of this dissertation. An increase is observed during the Great Inflation of the 1970s followed by a tranquil period in the 1990s and by a volatile period during the Financial Crisis of 2008. Figure (2.2) (in appendix) shows the estimated trajectories of $\mathbb{E}(B_t)$. We observe that most parameters are stable or change smoothly throughout the sample, although some of them are very volatile.

Figure (2.3) (in appendix) presents the time evolution of $\mathbb{E}(\exp(h_t))$ for the TVP-FAVAR with IP as target. Again the results indicate strong time-variation in the conditional variance of the factors and the forecasting target. The time-varying volatility of the factors are very similar to the previous specification. However, the IP seems to be less volatile than the CPI, displaying a more cyclical behavior. Moreover, the estimated trajectories for $\mathbb{E}(B_t)$ are somewhat more stable than the TVP-FAVAR for the CPI.

The forecasting results for the models with 5 and 6 factors are presented in tables 2.6 - 2.11 (in appendix). The TVP-FAVAR yields as good as or better forecasts than the FAVAR model in 55.5% of the cases for inflation measures, and in 40.2% of the cases for the real variables. Better results are obtained when comparing to the AR(1) and the LL, often statistically significant. We note that the TVP-FAVAR seems to improve on the FAVAR especially for the CPI, which sees the best results in comparison to the time-invariant FAVAR, especially for short-term forecasting. Interestingly, the time span does not seem to interfere, and the TVP-FAVAR produces better short-term forecasts in the first and in the second half of the sample for the CPI. For the PPI the best results are observed in the first half of the sample, which covers the aftermath of the Great Inflation of the 1970s. The PCE deflator seems to be less volatile than other inflation measures, and the TVP-FAVAR did not improve on the time-invariant FAVAR. This is not a surprise as CPI inflation is known for being more volatile than other inflation indicators, such as the PCE deflator and the PPI.

In addition, the TVP-FAVAR produces very good results for the real variables in relation to the univariate models, but it does not improve over the time-invariant FAVAR. Overall, models based on factors seem to generate large forecast improve-
ments for real variables over univariate approaches. These findings are similar to the ones in Stock and Watson (2002b), in which it was found that the factors have more predictive power for real variables rather than for inflation measures.

2.6 Conclusions

In this paper we study time-varying factor-augmented vector autoregressive models with stochastic volatility. We let the conditional mean parameters and the volatility be time-varying as driftless random walk processes. This means that the FAVAR parameters will be able to accommodate time-variations in the relationship between the macroeconomic factors and the target variables.

This paper suggests a classical estimation procedure in two-steps. In the first one principal components are used to form a feasible likelihood function. In the second we show that the model can be estimated equation-by-equation by assuming a specific factorization of the Gaussian multivariate importance sampler. Therefore, the model is estimated equationwise as time-varying autoregressive processes with stochastic volatility. Finally, we present an estimation procedure for this sequence of univariate regressions on the basis of the NAIS method of Koopman et al. (2015) together with a Kalman Filter step to integrate out linear state-variables. The resulting estimation procedure is computationally fast and numerically precise.

A simulation study is performed to verify the properties of the estimation procedure. It is shown that the equationwise NAIS method is able to properly identify the model parameters even in very small sample sizes. In addition, we investigate whether the PC factors are a suitable substitute for the real factors. The results suggest that the PC estimator is able to recover the underlying factor structure in a very sensible way even if the factor process is characterized by stochastic volatility and time-varying parameters. Finally, another MC study shows that the equation-by-equation procedure is comparable to the joint estimation of the model parameters.

In the empirical application the TVP-FAVAR is used to forecast macroeconomic variables using a large dataset. The macroeconomic variables are taken from the Jurado, Ludvigson and Ng dataset as used in Jurado et al. (2013). We forecast three inflation measures: the consumer price index, the producer price index, and the personal consumption expenditure deflator; and three real variables: the industrial production total, the unemployment rate, and the personal income.

We find evidence that the macroeconomic factors possess time-varying volatility, and some parameters of the TVP-FAVAR are very volatile. Additionally, the TVP-FAVAR can improve the forecasting performance over the time-invariant FAVAR and the univariate benchmark models. In fact the TVP-FAVAR seems to be a good predictor of price indexes, and the results were particularly good for the CPI. In this case we note that the TVP-FAVAR outperforms the FAVAR in short term forecasting across all samples sizes and number of factors. For real variables the TVP-FAVAR produces very
good results in relation to the univariate models, but it does not improve over the time-invariant FAVAR.

2.7 References


2.8 Appendix

2.8.1 Tables

In this section we report the tables with the results of the MC studies, and the RMSEs ratios corresponding to the two forecasting exercises presented in section 2.5. Tables 2.5 - 2.11 present RMSE relative to the TVP-FAVAR. Bold values implies RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the equal conditional predictive ability test of Giacomini and White (2006).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP</th>
<th>( T = 250 )</th>
<th>( T = 500 )</th>
<th>( T = 750 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{3,3} )</td>
<td>DGP I</td>
<td>0.04 (0.014)</td>
<td>0.036 (0.009)</td>
<td>0.035 (0.009)</td>
</tr>
<tr>
<td>( \sigma_{1,3} )</td>
<td></td>
<td>0.03 (0.013)</td>
<td>0.025 (0.009)</td>
<td>0.036 (0.009)</td>
</tr>
<tr>
<td>( \sigma_{2,3} )</td>
<td></td>
<td>0.03 (0.012)</td>
<td>0.025 (0.008)</td>
<td>0.033 (0.010)</td>
</tr>
<tr>
<td>( \sigma_{h,3} )</td>
<td></td>
<td>0.04 (0.043)</td>
<td>0.053 (0.026)</td>
<td>0.051 (0.020)</td>
</tr>
<tr>
<td>( \sigma_{3,3} )</td>
<td>DGP II</td>
<td>0.03 (0.013)</td>
<td>0.027 (0.009)</td>
<td>0.028 (0.009)</td>
</tr>
<tr>
<td>( \sigma_{1,3} )</td>
<td></td>
<td>0.02 (0.013)</td>
<td>0.016 (0.008)</td>
<td>0.016 (0.007)</td>
</tr>
<tr>
<td>( \sigma_{2,3} )</td>
<td></td>
<td>0.02 (0.010)</td>
<td>0.016 (0.014)</td>
<td>0.017 (0.009)</td>
</tr>
<tr>
<td>( \sigma_{h,3} )</td>
<td></td>
<td>0.03 (0.036)</td>
<td>0.043 (0.022)</td>
<td>0.039 (0.018)</td>
</tr>
</tbody>
</table>

The table presents the DGP parameters, followed by sample average of estimated parameters from 250 series. Values in parentheses are sample standard deviations of the estimates.
### Table 2.3: Correlation between simulated and estimated factors

<table>
<thead>
<tr>
<th>Correlation</th>
<th>$T = 250$</th>
<th>$T = 500$</th>
<th>$T = 750$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{corr}(\hat{f}_1, f_1)$</td>
<td></td>
<td>DGP I</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.971</td>
<td>0.978</td>
<td>0.977</td>
</tr>
<tr>
<td>5%-quantile</td>
<td>0.897</td>
<td>0.906</td>
<td>0.907</td>
</tr>
<tr>
<td>95%-quantile</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>$\text{corr}(\hat{f}_2, f_2)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.968</td>
<td>0.973</td>
<td>0.971</td>
</tr>
<tr>
<td>5%-quantile</td>
<td>0.899</td>
<td>0.909</td>
<td>0.899</td>
</tr>
<tr>
<td>95%-quantile</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>$\text{corr}(\hat{f}_1, f_1)$</td>
<td>DGP II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.980</td>
<td>0.980</td>
<td>0.981</td>
</tr>
<tr>
<td>5%-quantile</td>
<td>0.896</td>
<td>0.908</td>
<td>0.907</td>
</tr>
<tr>
<td>95%-quantile</td>
<td>0.998</td>
<td>0.998</td>
<td>0.997</td>
</tr>
</tbody>
</table>

The table presents the median, 5%-quantile, and 95%-quantile of the correlation between the estimated (rescaled) and the simulated factors across 250 replications.

### Table 2.4: Finite-Sample properties of the equationwise procedure

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP</th>
<th>$T = 250$</th>
<th>$T = 500$</th>
<th>$T = 250$</th>
<th>$T = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DGP</td>
<td>Equationwise procedure</td>
<td>Joint procedure</td>
<td>DGP</td>
<td>Equationwise procedure</td>
</tr>
<tr>
<td>$\sigma_{1,1}$</td>
<td>0.04</td>
<td>0.032 (0.017)</td>
<td>0.034 (0.011)</td>
<td>0.032 (0.017)</td>
<td>0.034 (0.010)</td>
</tr>
<tr>
<td>$\sigma_{1,2}$</td>
<td>0.03</td>
<td>0.027 (0.022)</td>
<td>0.027 (0.010)</td>
<td>0.026 (0.019)</td>
<td>0.026 (0.010)</td>
</tr>
<tr>
<td>$\sigma_{h,2}$</td>
<td>0.04</td>
<td>0.075 (0.054)</td>
<td>0.058 (0.032)</td>
<td>0.074 (0.051)</td>
<td>0.058 (0.032)</td>
</tr>
<tr>
<td></td>
<td>DGP</td>
<td>Equationwise procedure</td>
<td>Joint procedure</td>
<td>DGP</td>
<td>Equationwise procedure</td>
</tr>
<tr>
<td>$\sigma_{1,1}$</td>
<td>0.03</td>
<td>0.023 (0.014)</td>
<td>0.025 (0.010)</td>
<td>0.024 (0.014)</td>
<td>0.025 (0.010)</td>
</tr>
<tr>
<td>$\sigma_{1,2}$</td>
<td>0.02</td>
<td>0.018 (0.019)</td>
<td>0.018 (0.010)</td>
<td>0.019 (0.019)</td>
<td>0.018 (0.009)</td>
</tr>
<tr>
<td>$\sigma_{h,2}$</td>
<td>0.03</td>
<td>0.062 (0.040)</td>
<td>0.049 (0.026)</td>
<td>0.062 (0.040)</td>
<td>0.049 (0.026)</td>
</tr>
</tbody>
</table>

The table presents the DGP parameters followed by sample average of estimated parameters from 250 series. Values in parentheses are sample standard deviations of the estimates.
Table 2.5: Forecasting results for the simulated series

<table>
<thead>
<tr>
<th>DGP</th>
<th>h</th>
<th>TVP-FAVAR</th>
<th>FAVAR</th>
<th>LL</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>1</td>
<td>1.128**</td>
<td>1.317**</td>
<td>1.087*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.113**</td>
<td>1.118**</td>
<td>1.099*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.076</td>
<td>1.060*</td>
<td>1.061</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.002</td>
<td>1.018</td>
</tr>
<tr>
<td>II</td>
<td>1</td>
<td>1</td>
<td>1.092*</td>
<td>1.125**</td>
<td>1.185**</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.012</td>
<td>1.069*</td>
<td>1.050*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.003</td>
<td>1.010</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.025</td>
<td>1.002</td>
</tr>
</tbody>
</table>

The table presents RMSE relative to the TVP-FAVAR. Bold values imply RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the Giacomini and White (2006) test.
Table 2.6: Forecasting results for the 5-factors models (whole sample)

<table>
<thead>
<tr>
<th>Variable</th>
<th>h</th>
<th>TVP-FAVAR</th>
<th>FAVAR</th>
<th>LL</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1</td>
<td>1</td>
<td>1.012</td>
<td>1.039**</td>
<td>1.010</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.012*</td>
<td>1.016</td>
<td>1.021*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.999</td>
<td>0.998</td>
<td>1.002</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.004</td>
<td>1.001</td>
</tr>
<tr>
<td>PCEd</td>
<td>1</td>
<td>1</td>
<td>1.002</td>
<td>1.069*</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.993</td>
<td>0.994</td>
<td>1.008</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.996</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.002</td>
<td>1.003</td>
</tr>
<tr>
<td>PPI</td>
<td>1</td>
<td>1</td>
<td>1.009</td>
<td>1.125**</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.002</td>
<td>0.987</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.000</td>
<td>1.002</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.006</td>
<td>1.002</td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>1</td>
<td>0.965</td>
<td>1.003</td>
<td>1.060</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.972</td>
<td>1.020*</td>
<td>1.052*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.013</td>
<td>1.046</td>
<td>1.045</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.084</td>
<td>1.024*</td>
</tr>
<tr>
<td>U</td>
<td>1</td>
<td>1</td>
<td>0.997</td>
<td>1.076*</td>
<td>1.147**</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.998</td>
<td>1.066*</td>
<td>1.119*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.004</td>
<td>1.042</td>
<td>1.066</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.002</td>
<td>1.043</td>
<td>1.021</td>
</tr>
<tr>
<td>PEI</td>
<td>1</td>
<td>1</td>
<td>0.931</td>
<td>0.973</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.998</td>
<td>1.016</td>
<td>1.021</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.995</td>
<td>1.008</td>
<td>1.018</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.008</td>
<td>1.003</td>
</tr>
</tbody>
</table>

The table presents RMSE relative to the TVP-FAVAR. Bold values implies RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the Giacomini and White (2006) test.
### Table 2.7: Forecasting results for the 5-factors models (first half of the sample)

<table>
<thead>
<tr>
<th>Variable</th>
<th>h</th>
<th>TVP-FAVAR</th>
<th>FAVAR</th>
<th>LL</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1</td>
<td>1</td>
<td><strong>1.009</strong></td>
<td>1.074</td>
<td>1.029</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td><strong>1.017</strong></td>
<td>0.993</td>
<td>1.026</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td><strong>1.001</strong></td>
<td><strong>1.004</strong></td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td><strong>1.002</strong></td>
<td>1.002</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>0.993</td>
<td><strong>1.092</strong></td>
<td><strong>1.010</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td><strong>1.001</strong></td>
<td><strong>0.984</strong></td>
<td><strong>1.006</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td><strong>1.002</strong></td>
<td><strong>1.004</strong></td>
<td><strong>1.001</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td><strong>1.002</strong></td>
<td><strong>1.006</strong></td>
</tr>
<tr>
<td>PCEd</td>
<td>1</td>
<td>1</td>
<td><strong>1.004</strong></td>
<td>1.081</td>
<td><strong>1.012</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td><strong>1.039</strong></td>
<td><strong>0.997</strong></td>
<td><strong>1.031</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td><strong>1.006</strong></td>
<td><strong>1.003</strong></td>
<td><strong>1.003</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>0.998</td>
<td><strong>1.006</strong></td>
</tr>
<tr>
<td>PPI</td>
<td>1</td>
<td>1</td>
<td><strong>1.002</strong></td>
<td><strong>1.025</strong></td>
<td><strong>1.091</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.999</td>
<td><strong>1.033</strong></td>
<td><strong>1.033</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.996</td>
<td><strong>1.038</strong></td>
<td><strong>1.014</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td><strong>1.020</strong></td>
<td><strong>1.013</strong></td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>1</td>
<td>0.997</td>
<td><strong>1.017</strong></td>
<td><strong>1.062</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.986</td>
<td><strong>1.002</strong></td>
<td><strong>1.020</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td><strong>1.001</strong></td>
<td><strong>1.002</strong></td>
<td><strong>1.011</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td><strong>1.013</strong></td>
<td><strong>1.012</strong></td>
</tr>
<tr>
<td>U</td>
<td>1</td>
<td>1</td>
<td>0.948</td>
<td>0.968</td>
<td>0.978</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.990</td>
<td>0.994</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.995</td>
<td>0.993</td>
<td><strong>1.006</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td><strong>1.005</strong></td>
<td><strong>1.004</strong></td>
</tr>
</tbody>
</table>

The table presents RMSE relative to the TVP-FAVAR. Bold values implies RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the Giacomini and White (2006) test.
Table 2.8: Forecasting results for the 5-factors models (second half of the sample)

<table>
<thead>
<tr>
<th>Variable</th>
<th>h</th>
<th>TVP-FAVAR</th>
<th>FAVAR</th>
<th>LL</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1</td>
<td>1</td>
<td>1.013</td>
<td></td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.010</td>
<td>1.030*</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.998</td>
<td>0.997</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.005</td>
<td>1.002</td>
</tr>
<tr>
<td>PCEd</td>
<td>1</td>
<td>1</td>
<td>1.006</td>
<td>1.059</td>
<td>1.024</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.989</td>
<td>0.998</td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.995</td>
<td>0.994</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.003</td>
<td>1.001</td>
</tr>
<tr>
<td>PPI</td>
<td>1</td>
<td>1</td>
<td>1.010</td>
<td>1.137**</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.992</td>
<td>0.984</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.992</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.008</td>
<td>1.001</td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>1</td>
<td>0.943</td>
<td>0.991</td>
<td>1.042</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.957</td>
<td>1.013*</td>
<td>1.062*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.022</td>
<td>1.051</td>
<td>1.061</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.115</td>
<td>1.029*</td>
</tr>
<tr>
<td>U</td>
<td>1</td>
<td>1</td>
<td>0.997</td>
<td>1.141*</td>
<td>1.239*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.012</td>
<td>1.121*</td>
<td>1.211*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.002</td>
<td>1.088</td>
<td>1.113</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.066</td>
<td>1.023</td>
</tr>
<tr>
<td>PEI</td>
<td>1</td>
<td>1</td>
<td>0.906</td>
<td>0.980</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.010</td>
<td>1.048*</td>
<td>1.053*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.994</td>
<td>1.029*</td>
<td>1.034</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.012</td>
<td>1.001</td>
</tr>
</tbody>
</table>

The table presents RMSE relative to the TVP-FAVAR. Bold values implies RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the Giacomini and White (2006) test.
Table 2.9: Forecasting results for the 6-factors models (whole sample)

<table>
<thead>
<tr>
<th>Variable</th>
<th>h</th>
<th>TVP-FAVAR</th>
<th>FAVAR</th>
<th>LL</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1</td>
<td>1</td>
<td>1.007</td>
<td>1.028*</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.014*</td>
<td>1.017</td>
<td>1.022*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.998</td>
<td>0.997</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.004</td>
<td>1.001</td>
</tr>
<tr>
<td>PCEd</td>
<td>1</td>
<td>1</td>
<td>0.991</td>
<td>1.051**</td>
<td>1.004</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.989</td>
<td>0.986</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.998</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.004</td>
<td>1.002</td>
</tr>
<tr>
<td>PPI</td>
<td>1</td>
<td>1</td>
<td>1.010</td>
<td>1.124**</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.001</td>
<td>0.988</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.991</td>
<td>1.004</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.005</td>
<td>1.004</td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>1</td>
<td>0.984</td>
<td>1.031*</td>
<td>1.089*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.960</td>
<td>1.025*</td>
<td>1.056*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.992</td>
<td>1.040</td>
<td>1.039</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.015</td>
<td>1.084</td>
<td>1.023</td>
</tr>
<tr>
<td>U</td>
<td>1</td>
<td>1</td>
<td>0.986</td>
<td>1.081*</td>
<td>1.153**</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.983</td>
<td>1.071*</td>
<td>1.124**</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.004</td>
<td>1.056</td>
<td>1.080</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>0.972</td>
<td>1.042</td>
<td>1.018</td>
</tr>
<tr>
<td>PEI</td>
<td>1</td>
<td>1</td>
<td>0.940</td>
<td>0.981</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.999</td>
<td>1.018</td>
<td>1.022</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.000</td>
<td>1.017</td>
<td>1.028</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.002</td>
<td>1.015</td>
<td>1.009</td>
</tr>
</tbody>
</table>

The table presents RMSE relative to the TVP-FAVAR. Bold values implies RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the Giacomini and White (2006) test.
Table 2.10: Forecasting results for the 6-factors models (first half of the sample)

<table>
<thead>
<tr>
<th>Variable</th>
<th>h</th>
<th>TVP-FAVAR</th>
<th>FAVAR</th>
<th>LL</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1</td>
<td>1</td>
<td>1.020</td>
<td>1.064</td>
<td>1.021</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.025</td>
<td>0.996</td>
<td>1.030</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.001</td>
<td>1.002</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>1.001</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>0.995</td>
<td>1.085</td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.004</td>
<td>0.985</td>
<td>1.006</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.999</td>
<td>1.002</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.001</td>
<td>1.002</td>
<td>1.000</td>
</tr>
<tr>
<td>PCEd</td>
<td>1</td>
<td>1</td>
<td>1.001</td>
<td>1.076</td>
<td>1.007</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1.039*</td>
<td>0.996</td>
<td>1.030*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.006</td>
<td>1.003</td>
<td>1.002</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.000</td>
<td>0.998</td>
<td>1.004</td>
</tr>
<tr>
<td>PPI</td>
<td>1</td>
<td>1</td>
<td>1.003</td>
<td>1.045</td>
<td>1.115</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.996</td>
<td>1.040</td>
<td>1.040</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.999</td>
<td>1.045</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>1.006</td>
<td>1.017</td>
<td>1.011</td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>1</td>
<td>0.993</td>
<td>1.035</td>
<td>1.081*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.981</td>
<td>1.005</td>
<td>1.023*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>1.001</td>
<td>1.002</td>
<td>1.013*</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>0.986</td>
<td>1.008</td>
<td>1.007</td>
</tr>
<tr>
<td>U</td>
<td>1</td>
<td>1</td>
<td>0.958</td>
<td>0.977</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.991</td>
<td>0.995</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.998</td>
<td>0.996</td>
<td>1.010</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>0.999</td>
<td>1.004</td>
<td>1.004</td>
</tr>
</tbody>
</table>

The table presents RMSE relative to the TVP-FAVAR. Bold values implies RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the Giacomini and White (2006) test.


### Table 2.11: Forecasting results for the 6-factors models (second half of the sample)

<table>
<thead>
<tr>
<th>Variable</th>
<th>h</th>
<th>TVP-FAVAR</th>
<th>FAVAR</th>
<th>LL</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1</td>
<td>1</td>
<td><strong>1.004</strong></td>
<td><strong>1.018</strong></td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td><strong>1.011</strong></td>
<td><strong>1.022</strong></td>
<td><strong>1.020</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.998</td>
<td>0.996</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td><strong>1.003</strong></td>
<td><strong>1.005</strong></td>
<td><strong>1.005</strong></td>
</tr>
<tr>
<td>PCEd</td>
<td>1</td>
<td>1</td>
<td>0.990</td>
<td>1.038**</td>
<td><strong>1.004</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.982</td>
<td>0.987</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.998</td>
<td>0.997</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td><strong>1.002</strong></td>
<td><strong>1.005</strong></td>
<td><strong>1.002</strong></td>
</tr>
<tr>
<td>PPI</td>
<td>1</td>
<td>1</td>
<td><strong>1.012</strong></td>
<td>1.137**</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.991</td>
<td>0.987</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.997</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td><strong>1.002</strong></td>
<td><strong>1.004</strong></td>
<td><strong>1.000</strong></td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>1</td>
<td>0.973</td>
<td><strong>1.023</strong></td>
<td>1.075</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.941</td>
<td><strong>1.017</strong></td>
<td><strong>1.065</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>0.988</td>
<td><strong>1.038</strong></td>
<td><strong>1.049</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td><strong>1.019</strong></td>
<td>1.116</td>
<td><strong>1.030</strong></td>
</tr>
<tr>
<td>U</td>
<td>1</td>
<td>1</td>
<td>0.979</td>
<td><strong>1.130</strong></td>
<td><strong>1.227</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.985</td>
<td><strong>1.138</strong></td>
<td><strong>1.225</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td><strong>1.008</strong></td>
<td>1.106</td>
<td><strong>1.141</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>0.962</td>
<td>1.067</td>
<td><strong>1.026</strong></td>
</tr>
<tr>
<td>PEI</td>
<td>1</td>
<td>1</td>
<td>0.912</td>
<td>0.987</td>
<td><strong>1.028</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td><strong>1.010</strong></td>
<td><strong>1.050</strong></td>
<td><strong>1.054</strong></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td><strong>1.003</strong></td>
<td><strong>1.048</strong></td>
<td><strong>1.053</strong></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td><strong>1.007</strong></td>
<td><strong>1.028</strong></td>
<td><strong>1.017</strong></td>
</tr>
</tbody>
</table>

The table presents RMSE relative to the TVP-FAVAR. Bold values imply RMSE values larger than the TVP-FAVAR. One, and two stars mean 0.05 and 0.01 statistical significance, respectively, for the Giacomini and White (2006) test.
2.8.2 Figures

In this section we report the figures of the time-varying conditional variance and time-varying parameters of the TVP-FAVAR with CPI and the IP as target variables.

![Factor 1](image1)

![Factor 2](image2)

![Factor 3](image3)

![Factor 4](image4)

![Factor 5](image5)

![CPI](image6)

Figure 2.1: TVP-FAVAR with CPI: Estimated trajectory of the conditional variance of the factors and the forecasting target.
Figure 2.2: TVP-FAVAR with CPI: Estimated trajectory of the time-varying parameters, i.e. $E(B_t)$. 
Figure 2.3: TVP-FAVAR with IP: Estimated trajectory of the conditional variance of the factors and the forecasting target.
Figure 2.4: TVP-FAVAR with IP: Estimated trajectory of the time-varying parameters, i.e. $E(B_t)$. 
CHAPTER 3

UNIVARIATE MAXIMUM LIKELIHOOD ESTIMATION OF A TVP-VAR

Douglas Eduardo Turatti
Aarhus University and CREATE

Abstract

This paper proposes an estimation procedure for time-varying vector autoregressive models which is not constrained by the dimension of the covariance matrix. The estimation procedure tackles the dimensionality problem of the TVP-VAR by applying a triangularization scheme and replacing unknown disturbances by estimated residuals. Through this process the system becomes a sequence of conditionally independent time-varying autoregressions with stochastic volatility. Finally, a classical maximum likelihood estimator for this very low dimensional sequence of time-varying parameter regressions is proposed based on the importance sampling technique together with a Kalman filter to yield a computationally fast and numerically accurate estimator. A Monte Carlo study suggests that the equationwise procedure has good finite-sample and asymptotic properties and results are comparable to the joint estimation of the model parameters. An empirical application to U.S. data generates results that resemble those reported in the macroeconomic literature.
3.1 Introduction

Vector Autoregressive (VAR) models have become a standard tool in macroeconometrics. They are able to describe the relationship between multiple economic variables, and to address how policy changes affect the rest of the economy. Furthermore, they can be used for economic forecasting and time series analysis. However, traditional VAR models rely on the assumption that economic relations are stable over time although most economic variable are subject to structural changes especially in longer time horizons. For example, Stock and Watson (1996) documented parameter instability in a large set of macroeconomic time series.

In order to cope with structural instability it is necessary to account for two sources of time-variation present in economic data: the time-varying size of the shocks in the endogenous variables and the changes in the mechanism of propagation of these shocks. Time-varying parameter models (TVP) have become a popular approach to allow for time-variation in the conditional mean of a VAR model. Nevertheless, estimating TVP models ignoring possible changes in the exogenous shocks is likely to generate fictitious dynamics in the coefficients (Cogley and Sargent, 2005). Thus, more recently, macroeconomists have been interested in TVP-VAR models with stochastic volatility to allow for time-varying shocks and time-variation in their propagation (examples include Cogley and Sargent, 2005; Primiceri, 2005; and Cogley et al. 2010).

However, statistical inference of TVP-VAR models with stochastic volatility is very challenging. The problem is due to the presence of multiple non-linear state variables. Note that a homoscedastic TVP-VAR can be estimated by the Kalman Filter and Smoother (KFS). However, the introduction of multivariate stochastic volatility implies that the model becomes non-linear in the state variables, rulling out the prediction error decomposition computation by the KFS. The estimation procedure, thus, requires highly intensive and costly computational methods, for example, Markov Chain Monte Carlo algorithms (e.g. Gibbs Sampling). Moreover, the computational cost increases nonlinearly with the number of variables, making CPU time greater with the size of the covariance matrix. This holds even if the time-variation in the covariance matrix is restricted to the volatility states.

To tackle the estimation problem several MCMC algorithms were developed based on the Gibbs Sampling. More complex specifications, for example, with sign restrictions, require Metropolis-Hasting (MH) algorithms (e.g Koop and Potter, 2011). As a matter of fact Negro and Primiceri (2015) found that the original Gibbs sampling proposed by Primiceri (2005) did not generate draws from the correct posterior distribution, indicating the complexity of the estimation procedure. On the other hand, classical analysis by maximum likelihood methods still have to be developed for this class of models. Actually, Primiceri (2005) argued that classical methods are not feasible due to high-dimensionality of the likelihood function and the possible existence of multiple unwanted peaks.
In this paper we introduce a new estimation procedure for TVP-VAR models with multivariate stochastic volatility. We base our analysis on a triangularization scheme which allows an equationwise treatment. In contrast to the literature we develop a likelihood based estimator. The reason is due to the triangularization which greatly reduces the dimensionality of the problem and a likelihood-estimator becomes feasible to develop. The aim of this paper is twofold. The first one is to introduce our estimation procedure for TVP-VAR models with stochastic volatility. The second is to present a Monte Carlo study and an empirical application using the new method.

TVP-VAR models are usually subjected to the curse of the dimensionality. This implies several additional problems and computational aspects, for example the CPU time requirements are highly nonlinear in the number of variables. This is caused by a non-diagonal covariance matrix, which creates difficulties for an equation-by-equation estimation. For this reason most applications of TVP-VAR are limited to low-dimensional cases (usually 3 to 5, e.g. Cogley and Sargent, 2005). On the other hand, large VARs are often homoscedastic, see for example Barbura et al. (2010). Our estimation procedure tackles the dimensionality problem and the required simulation effort by applying a triangularization scheme and approximating unobserved disturbances by estimated residuals. Through this process the TVP-VAR system becomes a sequence of conditionally independent time-varying parameter autoregressions with stochastic volatility. Finally, to estimate this one-dimensional regression we propose a simulated maximum likelihood approach based on the importance sampling technique together with a Kalman Filter step. It is worth mentioning that our estimation procedure does not suffer from the curse of dimensionality even if the covariance matrix possess time-varying covariances and variances.

The remainder of the paper is organized as follows. Section 2 lays out the TVP-VAR model we consider. Section 3 presents the estimation and filtering procedure. Section 4 discusses a Monte Carlo study to evaluate the performance of the estimation method. Section 5 presents an empirical application to U.S. data.

### 3.2 Time-Varying VAR Models with Stochastic Volatility

Let $y_t$ be the $N \times 1$ vector of observables variables at time $t$. $X_t = [C_t', y_{t-1}', \ldots y_{t-p}']'$ is the matrix of regressors where $C_t$ is a $c$-dimensional vector of deterministic regressors. $B_t = [B_{c,t}, B_{1,t}, \ldots , B_{p,t}]$ is the matrix of linear time-varying parameters, where $B_{c,t}$ is of $N \times c$ dimension and $B_{i,t}$ for $i = 1, \ldots, p$ is a $N \times N$ matrix. The general TVP-VAR model with stochastic volatility can be written as,

$$y_t = X_t' B_t + Y(H_t) \varepsilon_t, \quad \varepsilon_t \sim N(0_N, I_N), \quad (3.1)$$

$$\text{vec}(B_{t+1}') = \text{vec}(B_t') + Y(Q) \eta_t, \quad \eta_t \sim N(0_{Nl}, I_{Nl}), \quad (3.2)$$

where $0_N$ is a $N \times 1$ vector zero and $l = c + Np$. $Y(.)$ represents the lower Cholesky factor, and $\text{vec}(.)$ stacks the elements of the columns of a matrix into a vector. As usual
in the macroeconometrics literature the time-varying parameters $B_{i,t}$ are assumed to follow random walk processes. The economic rationale for random walk coefficients is that agents gradually learn about the economy as it is hit by exogenous shocks. Thus, they adapt their decision rules to new circumstances, and this causes the parameters in $B_{i,t}$ to change in unpredictable ways (Cogley and Sargent, 2001). The matrices $H_t$ and $Q$ are the covariances matrices of the measurement and transition equations respectively. The matrix $Q$ is assumed to be diagonal. Without loss of generality, consider the LDL decomposition of of the matrix $H_t$,

$$AH_tA' = \Sigma_t,$$  \hspace{1cm} (3.3)

$$H_t = A^{-1}\Sigma_tA^{-1'},$$  \hspace{1cm} (3.4)

where $A$ is a lower triangular unit matrix, and $\Sigma_t$ is a diagonal matrix consisting of time-varying variance states. Let $\Sigma_t$ be the matrix of time-varying volatilities,

$$\Sigma_t = \begin{bmatrix} \sigma_{1,t}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{2,t}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{N,t}^2 \end{bmatrix}. \hspace{1cm} (3.5)$$

It follows that the lower Cholesky factor of $H_t$ is given by,

$$\Upsilon(H_t) = A^{-1}\Sigma_t^{0.5}. \hspace{1cm} (3.6)$$

Finally, the time-varying volatility states are determined by a Gaussian driftless stochastic volatility model,

$$\sigma_{i,t}^2 = \exp(h_{i,t}), \hspace{1cm} (3.7)$$

$$h_{i,t} = h_{i,t-1} + \upsilon_{i,t}, \hspace{0.5cm} \forall i = 1, \ldots, N, \hspace{1cm} (3.8)$$

where the vector of volatility innovations $\upsilon_t$ is $N(0_N, \Omega)$, and $\Omega$ is a diagonal matrix as in Cogley and Sargent (2005). Note that the specification (3.3) implies time-varying variances and correlations, but in a tightly restricted fashion: as fixed proportions of the innovations’ variance (Koop and Korobilis, 2009). In impulse-response analysis, for instance, a constant $A$ implies that an innovation to the $i$-th variable has a time invariant effect on the $j$-th variable (Primiceri, 2005). Nevertheless, Primiceri (2005) reported only small changes in the impulse-response functions through the decades. Finally, the matrix $A$ can be allowed to be time-varying in our procedure, however, we leave this extension for future research.

Moreover, following standard practice in TVP-VAR models we decided not to impose stationarity restrictions in the time-varying autoregressive coefficients $B_t$. The main issue is related to the random walk assumption, which implies that these
parameters may reach any upper and lower bound when $T \to \infty$. However, if the data is stationary prior to analysis the possibility of explosive paths for $B_t$ remains more a theoretical than a practical issue. Therefore, despite the fact that temporary non-stationarity may happen, in the long-run the estimated TVP-FAVAR will follow the data resulting in a stationary paths for $y_t$.

Although popular and flexible the major drawback of the TVP-VAR is its estimation procedure. Given the structure of the covariance matrix Bayesian methods via MCMC suffer from the curse of dimensionality which hinders many potential empirical applications. Due to these difficulties most implementations are limited to a low-dimension $y_t$ (3 to 5). For example, Cogley and Sargent (2005) rely on a Bayesian posterior analysis based on MCMC technique, and use the single-move Gibbs sampling of Jacquier et al. (1994) to simulate the latent time-varying covariance matrices from the posterior distributions. Primiceri (2005) also developed Bayesian analysis based on the multi-move Gibbs algorithm of Kim et al. (1998) which is based on approximations constructed using mixture of normals. However, as Negro and Primiceri (2015) pointed out, Primiceri (2005) did not include the full set of the specific Gaussian mixture components used in each Gibbs sweep. Despite the fact that the algorithm was corrected by Negro and Primiceri (2015) and the results were found to be qualitatively similar the fact that the flaw in the Gibbs sampling remained for almost a decade can be seen as an indication of the complexity of the estimation method.

In order to improve the statistical inference of TVP-VAR models with stochastic volatility we present a new estimation procedure based on the triangularization scheme proposed by Carriero et al. (2015) for a Bayesian framework. We show that this procedure allows the model to be rewritten as a sequence of time-varying parameter autoregressions, meaning the system can be estimated equation-by-equation. Finally, we provide a classical maximum likelihood estimator for this sequence of one-dimensional models by applying the importance sampling technique. To the best our knowledge classical methods for TVP-VARs were not applied in the literature before.

### 3.3 A New Estimation Procedure for TVP-VAR models

This section develops a procedure to estimate the parameters of the TVP-VAR model outlined in equations (3.1)–(3.7) by classical maximum likelihood methods. Initially

---

1. Note that a stationary law of motion for $B_t$ does not eliminate the possibility of temporary non-stationarity and it implies a large number of static parameters to be estimated. Furthermore, if data is stationary by construction the estimated TVP-VAR will generate stationary paths in the long-run.

2. In regards to this issue Bayesian methods have a simple solution. In MCMC related algorithms non-stationarity draws are usually discarded, avoiding the evaluation of the conditional posterior distribution with explosive paths. Such procedure has not been implemented for simulated maximum likelihood methods. For future research it would be interesting to incorporate these ideas in the importance sampling framework by changing the draws dynamically over time.
the TVP-VAR is triangularized, meaning the system is rewritten in a triangular form according to the contemporaneous dependence from the system equations. In a further step the VAR equations are shown to be sequentially estimated as time-varying autoregressions with stochastic volatility. Finally, to estimate this \( N \) one-dimensional regressions we introduce an estimator based on a simulated maximum likelihood via importance sampling. In this paper we employ the Numerically Accelerated Importance Sampling developed by Koopman et al. (2015) together with a Rao-Blackwellization step to incorporate linear time-varying parameters. Although notationally tedious, the Rao-Blackwellized Numerically Accelerated Importance Sampling (RB-NAIS) is computationally simple and employs standard state-space methods. The procedure simplifies to a sequence of low-dimensional weighted least square problems.

### 3.3.1 Triangularization of a TVP-VAR

In this section we show how a TVP-VAR can be rewritten as a sequence of conditionally independent time-varying autoregressions with stochastic volatility. Recall that the measurement equation of TVP-VAR is given by,

\[
y_t = B_t X_t + A^{-1} \Sigma_t^{0.5} \epsilon_t, \quad \epsilon_t \sim N(0_N, I_N),
\]

(3.9)

Let \( a_{i,j} \) be a generic term of the matrix \( A^{-1} \) and \( b_{i,t} \) be the rows of the matrix \( B_t \) corresponding to dependent variable \( i \). Given the triangular form of the matrix \( A \) the VAR can be written as,

\[
y_{1,t} = b_{1,t} X_t + \sigma_{1,t} \epsilon_{1,t}
\]

(3.10)

\[
y_{2,t} = b_{2,t} X_t + a_{2,1} \sigma_{1,t} \epsilon_{1,t} + \sigma_{2,t} \epsilon_{2,t}
\]

(3.11)

\[...
\]

\[
y_{N,t} = b_{N,t} X_t + a_{N,1} \sigma_{1,t} \epsilon_{1,t} + \cdots + a_{N,N-1} \sigma_{N-1,t} \epsilon_{N-1,t} + \sigma_{N,t} \epsilon_{N,t}.
\]

(3.12)

Therefore, we can write a generic equation for \( y_{i,t} \) as,

\[
y_{i,t} - (a_{i,1} \sigma_{1,t} \epsilon_{1,t} + \cdots + a_{i,i-1} \sigma_{i-1,t} \epsilon_{i-1,t}) = b_{i,t} X_t + \sigma_{i,t} \epsilon_{i,t}.
\]

(3.13)

Consider the term in parentheses \( (a_{i,1} \sigma_{1,t} \epsilon_{1,t} + \cdots + a_{i,i-1} \sigma_{i-1,t} \epsilon_{i-1,t}) \) in the generic equation (3.13). If the error terms from all previous equations are given this value can be determined, and the generic equation (3.13) will depend on only one stochastic volatility. Our estimation strategy is based on developing a maximum likelihood approximation for the generic equation (3.13), which implies a sequential estimation for the TVP-VAR system.

However, equation (3.13) cannot be easily estimated because the error terms in \( (a_{i,1} \sigma_{1,t} \epsilon_{1,t} + \cdots + a_{i,i-1} \sigma_{i-1,t} \epsilon_{i-1,t}) \) are unknown and cannot be estimated\(^3\). Thus,

\(^3\)States and disturbances cannot be consistently estimated. It is possible to estimate aspects of their conditional distribution, but not their value per se.
the literature has focused on joint estimation of all equations in the model by Gibbs Sampling or Metropolis-Hastings methods. Nevertheless, if an approximation for $\sigma_{1,t} \epsilon_{1,t}$ is available the system could potentially be estimated equation-by-equation. We propose to develop an approximation for this unknown error term on the basis of estimated residuals from a time-varying autoregressive model with stochastic volatility. Let’s assume a numerical device is available to estimate the parameters of a time-varying autoregressive model with stochastic volatility. Note that the first equation of the TVP-VAR alone is a time-varying autoregressive model with stochastic volatility. Hence, the completed system can be estimated by the following algorithm,

1. Estimate the parameters of the first equation of the TVP-VAR system and save the residual terms,

$$\hat{\epsilon}_{1,t} = y_{1,t} - \hat{b}_{1,t} X_t.$$  \hfill (3.14)

Compute the particle approximation $\tilde{\epsilon}_{1,t}$. The next section presents the particle approximation for states and disturbances.

2. Replace the unknown error term $\sigma_{1,t} \epsilon_{1,t}$ in the second equation by the approximation $\tilde{\epsilon}_{1,t}$. Note that the equation becomes now dependent on only one random innovation and stochastic volatility. Estimate the model and obtain the approximation for $\sigma_{2,t} \epsilon_{2,t}$ as $\tilde{\epsilon}_{2,t}$.

3. Plug $\tilde{\epsilon}_{1,t}$ and $\tilde{\epsilon}_{2,t}$ in the next equation and repeat the process until all the equations in the system are estimated.

The properties of the above procedure are unknown as it relies on approximations of unobserved quantities. Therefore, in section 4 we investigate the finite-sample and asymptotic properties of the equationwise procedure in an MC study. To preview some results we obtain that the parameter estimates are close to the DGP even in small sample sizes, and results are similar to a joint estimation of all model parameters. The previous algorithm describes how to estimate the TVP-VAR equation-by-equation. The next section introduces a maximum likelihood procedure for the generic equation $\tilde{\epsilon}_{i,t}$ of the TVP-VAR system.

3.3.2 Estimation

This section develops a procedure to estimate each equation of the triangularized TVP-VAR separately. Initially, consider the estimation of the general equation (3.13) for $y_{i,t}$ where we assume that the previous sequence of $(\tilde{\epsilon}_{1,t}, \ldots, \tilde{\epsilon}_{i-1,t})$ is available. Let’s define $\tilde{y}_{i,t} = y_{i,t} - (a_{i,1} \tilde{\epsilon}_{1,t} + \cdots + a_{i,i-1} \tilde{\epsilon}_{i-1,t})$. The likelihood function for the general equation is given by the integral over the joint density of observables $\tilde{y}_{i,t}$, latent linear states $b_{i,t}$, and latent log-volatility states $h_{i,t}$,

$$L(\gamma_i, \tilde{y}_i) = \int \prod_{t=1}^T p(\tilde{y}_{i,t}|b_{i,t}, h_{i,t}; \gamma_i) p(b_{i,t}|b_{i,t-1}; \gamma_i) p(h_{i,t}|h_{i,t-1}; \gamma_i) db_{i,1} \ldots db_{i,T} dh_{1,1} \ldots dh_{1,T}. \hfill (3.15)$$
where $\gamma_i$ is the parameter vector for this equation. The likelihood function equation \[ L(y_i, \tilde{y}_i) = \int p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma_i) p(h_{i,t} | \sigma_{i,t-1}, \gamma_i) d h_{i,1} \ldots d h_{i,T} \tag{3.16} \]

$ p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma) $ is the predictive density, $v_t$ and $F_t$ are the prediction error and its associated variance delivered by Kalman filter given the log-volatility $h_{i,t}$. This procedure is usually called Rao-Blackwellization, and it always reduces the Monte Carlo (MC) variation of the likelihood estimator.

However, it is still necessary to approximate the lower dimensional likelihood integral \[ L(y_i, \tilde{y}_i) = \int p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma_i) p(h_{i,t} | \sigma_{i,t-1}, \gamma_i) d h_{i,1} \ldots d h_{i,T} \tag{3.16} \]
where $ p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma) $ is the predictive density, $v_t$ and $F_t$ are the prediction error and its associated variance delivered by Kalman filter given the log-volatility $h_{i,t}$. This procedure is usually called Rao-Blackwellization, and it always reduces the Monte Carlo (MC) variation of the likelihood estimator.

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

\[ L(y_i, \tilde{y}_i) = \int p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma_i) p(h_{i,t} | \sigma_{i,t-1}, \gamma_i) g(h_i, h_i; \varphi) d h_i, \tag{3.19} \]

where $ g(h_i; \varphi) $, $ g(\tilde{y}_i; \varphi) $ are all Gaussian densities, $ g(\gamma_i; \varphi) $ is a normalizing constant and $ \varphi $ is the vector of parameters of the importance density. It is implied that $ g(h_i; \varphi) \equiv p(h_i; \gamma_i) $. Using this importance sampler it is possible to rewrite the likelihood function given some regularity conditions (see Geweke, 1989),

\[ L(y_i, \tilde{y}_i) = g(\tilde{y}_i; \varphi) \int p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma_i) p(h_{i,t} | \sigma_{i,t-1}, \gamma_i) g(h_i, h_i; \varphi) d h_i, \tag{3.20} \]

\[ L(y_i, \tilde{y}_i) = g(\tilde{y}_i; \varphi) \int \omega(h_i, \tilde{y}_i) g(h_i, h_i; \varphi) d h_i, \tag{3.21} \]

where $ g(\tilde{y}_i; \varphi) $ is the likelihood of the importance density model and the importance weights are defined as,

\[ \omega(y_i, \tilde{y}_i) = \frac{p^*(\tilde{y}_{i,t} | h_i)}{g(\tilde{y}_i; h_i, \varphi)}, \tag{3.22} \]
3.3. A New Estimation Procedure for TVP-VAR Models

evaluated at $S$ independent trajectories of the log-volatility. The likelihood estimator is given by,

$$
\hat{L}(\gamma, y) = g(\bar{y}_i; \varphi) \times \hat{\omega}(\gamma, \bar{y}_i), \quad \hat{\omega}(\gamma, \bar{y}_i) = \frac{1}{S} \sum_{i=1}^{S} \omega(h_i^{(i)}, \bar{y}_i). \quad (3.23)
$$

To evaluate (3.23) it is necessary to define the importance sampler $g(\bar{y}_i| h_i, \varphi)$ by selecting the importance parameters $\varphi$. Although most choices of $\varphi$ yield consistent estimators, blind selections can be highly inefficient. We follow Richard and Zhang (2007) strategy and select $\varphi$ minimizing the distance between the proposal and the target density. Moreover, we apply the NAIS method of Koopman et al. (2015), in which an approximate linear model for the importance density greatly simplifies the sampling procedure.

3.3.2.1 Rao-Blackwellized Numerically Accelerated Importance Sampling

In this section we develop a numerical procedure to evaluate the log-likelihood function of the general equation (3.13). The main elements are the linear representation for the importance density and the selection of the importance parameters $\varphi = \{\varphi_t\}_{t=1}^T$. Let $\varphi_t = (b_t, c_t)$, and as discussed in Koopman et al. (2015) the density $g(\bar{y}_{i,t}| h_{i,t}, \varphi_t)$ can be represented by a linear density,

$$
g(\bar{y}_{i,t}| h_{i,t}, \varphi_t) = \exp \left( a_t + b_t h_{i,t} - 0.5 c_t h_{i,t}^2 \right), \quad (3.24)
$$

where $a_t$ is an integrating constant, and the importance parameters $b_t$ and $c_t$ scalars for $t = 1, \ldots, T$, and are a function of $\bar{y}_{i,t}$ and $\gamma_i$. The expression for the Gaussian density (3.24) is possible given the general expressions for the exponential family of distributions. It can be shown that the measurement equation for density (3.24) is given by the following,

$$
y^*_t = h_{i,t} + \epsilon_{i,t}, \quad \epsilon_t \sim N(0, 1/c_t), \quad (3.25)
y^*_t = b_t/c_t. \quad (3.26)
$$

where the artificial observations $y^*_t$ are given by $y^*_t = b_t/c_t$ for a set of $(b_t, c_t)$. The artificial measurement equation (3.25) implies that it is possible to use standard methods for compute the moments and sampling from the proposal density, greatly improving the computational and numerical tractability of importance sampling methods. Thus, the smoothing density $g(h_{i,t}| \bar{y}_{i,t}, \varphi_t)$ is available analytically and is given by,

$$
g(h_{i,t}| \bar{y}_{i,t}, \varphi_t) = \frac{1}{\sqrt{2\pi V_t}} \exp \left( -0.5 \frac{(h_{i,t} - \bar{h}_{i,t})^2}{V_t} \right), \quad (3.27)
$$

5An important condition is that the variance of the importance weights is bounded (see Koopman et al (2009) for a discussion on the finitness of the variance of the importance weights). In addition, Richard and Zhang (2007) provide a discussion on the importance of selecting the importance parameters efficiently.
where $\tilde{h}_{i,t}$ and $V_t$ are the mean and the variance of the signal vector, which can be obtained by the KFS for a fixed $\varphi_t$. The sampling from (3.27) can be carried out by standard state-space methods, such as the simulation smoother of Durbin and Koopman (2002).

The NAIS methods contrasts especially with the efficient importance sampling (EIS) method of Richard and Zhang (2007). In their approach it is necessary to keep track of the integrating constant for every iteration, increasing the computational cost of computing the likelihood function. In addition, the EIS method strongly depends upon a good initial condition for the importance parameters $(b_t, c_t)$ due to numerical overflows caused by a ill-behaved integrating constant, which is likely to occur in the first iterations of the algorithm. On the other hand, NAIS is less dependent on integrating constants as the importance parameters incorporate information about $h_{i,t}$ in $\tilde{y}_{i,t}$ faster.

The importance parameters $\varphi$ are selected minimizing the variance of the importance weights in the full support of the likelihood function. Following the argument of Richard and Zhang (2007) this is near-equivalent to solving at every time $t$,

$$\arg\min_{\varphi_t} \int \lambda^2(\tilde{y}_{i,t}, h_{i,t} | \varphi) \omega(h_{i,t}, \tilde{y}_{i,t} | \varphi) g(h_{i,t} | \tilde{y}_{i,t}, \varphi_t) \, dh_{i,t},$$

(3.28)

where $\lambda(.)$ is defined as,

$$\lambda(\tilde{y}_{i,t}, h_{i,t} | \varphi_t) = (\log p^*(\tilde{y}_{i,t} | h_{i,t}) - \log g(y_{i,t}^* | h_{i,t}, \varphi_t)).$$

(3.29)

We note that whereas the likelihood integrand is very high-dimensional in most cases of interest, the integral (3.28) is usually of lower dimension.

To complete the procedure it is necessary to explain how the minimization (3.28) can be numerically implemented. The issue is due to the fact that the importance sampler $g(h_{i,t} | \tilde{y}_{i,t}, \varphi_t)$ is also a function of the importance parameters $\varphi_t$. Given an initial condition $\varphi^{(0)}$, this can be resolved by a fixed point argument using intermediate importance samplers to evaluate (3.28) until a criteria of convergence is satisfied. In addition, we follow Koopman et al. (2015) and apply numerical integration via Gauss-Hermite (GH) methods to solve the minimization problem in (3.28). Note that the dimension of (3.28) is the dimension of $h_{i,t}$, in this case a one-dimensional integration problem. Numerical methods for a one-dimensional integration problems are very fast and almost exact procedures. An important aspect of the Gauss-Hermite integration is that only a very small number of draws is necessary to efficiently evaluate (3.28). For instance, typical applications only require 10 to 15 points.

Let $\{z_i\}_{i=1}^M$ be the set of nodes $M$ predefined GH nodes and $h(z_i)$ be the respective Gauss-Hermite weights. We define the set of abcissae $\tilde{z}_{i,t} = \tilde{h}_{i,t} + \sqrt{V_t} z_i$. Thus, the minimization (3.28) can be rewritten as a least square problem,

$$\log p^*(\tilde{y}_{i,t}, \tilde{z}_{i,t}) = \text{constant} + \kappa \tilde{z}_{i,t} - \frac{1}{2} \xi \tilde{z}_{i,t}^2 + \text{error},$$

(3.30)
3.3. A NEW ESTIMATION PROCEDURE FOR TVP-VAR MODELS

with weights \( w_{i,t} = h(z_i) \exp\left( \frac{1}{2} z_i^2 \right) \), as the nodes are assumed to be associated with the standard normal distribution. The weighted least squares estimates for \( \kappa \) and \( \xi \) are the new estimates of \( b_t \) and \( c_t \) respectively. From these new estimates it is possible to obtain a new linear representation in equation \ref{3.25} for the stochastic volatility and draw from its Gaussian density. A comprehensive treatment of the RB-NAIS method can be found in the first chapter of this dissertation. The summarized algorithm is given by the following steps,

1. Initialize the procedure with starting values for \( b_t = 0 \) and \( c_t = 1 \) for each time \( t = 1 \ldots T \). The choice of the initial values is usually not necessary for convergence.

2. Compute smoothed mean \( \hat{h}_{i,t} \) and variance \( V_t \) for each time \( t = 1 \ldots T \) based on the current values of \((b_t, C_t)\) and the state-space representation \ref{3.25} via the KFS.

3. Evaluate the density \( p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma) \) by the prediction error decomposition. Minimize equation \ref{3.28} and obtain a new set of \( \varphi_t = (b_t, c_t) \) for every time \( t \).

4. Compute a criteria for convergence. Run steps 2 and 3 until convergence is achieved or preset a sufficient number of iterations. Usually 10 to 15 iterations suffice for convergence of the likelihood estimate.

5. Given the final set of \( \varphi = (\{b_t\}_{t=1}^T, \{C_t\}_{t=1}^T) \), sample S independent trajectories for the log-volatility via the simulation smoother of Durbin and Koopman (2002). Evaluate the density \( p^*(\tilde{y}_{i,t} | h_{i,t}, \gamma_i) \) by the prediction error decomposition conditional on the S independent trajectories for the log-volatility.

6. Compute the estimator for the likelihood function equation \ref{3.23}.

3.3.3 Filtering and Smoothing

The algorithm outlined in the previous steps yields a well-defined procedure to estimate a particular equation of the triangularized TVP-VAR. To complete the process we have to define an estimator for the state-variables and an approximation for the error terms that allow an equation-by-equation estimation. The set of static parameters is sequentially fixed at their MLE estimate. For the log-volatilities an estimator can be constructed based on the importance sampling procedure. For example, let \( h_{i,t} \) be the log-volatility for equation \( i \), an estimator for the moment \( \mathbb{E}(\exp(h_{i,t})) \) can be written as,

\[
\mathbb{E}(\exp(h_{i,t})) \approx \frac{\sum_{j=1}^{S} \omega_j(\gamma, \tilde{y}_{i,t}) \exp(h_{i,t}^{(j)})}{\sum_{j=1}^{S} \omega_j(\gamma, \tilde{y}_{i,t})} \quad (3.31)
\]
where \( h_{i,t}^{(j)} \) is a draw of the importance sampler \( g(h_{i,t} | \tilde{y}_{i,t}) \). An analogous procedure is applicable to the linear time-varying parameters \( b_{i,t} \),

\[
E(b_{i,t}) \approx \frac{\sum_{j=1}^{S} \omega_{j}(\gamma, \tilde{y}_{i,t}) b_{i,t}^{(j)}}{\sum_{j=1}^{S} \omega_{j}(\gamma, \tilde{y}_{i,t})},
\]

where the linear time-varying parameters \( b_{i,t} \) can be based on the Kalman filter or the Kalman smoother. Note that estimates for \( E(\exp(h_{i,t})) \) are always smoothed as the importance sampler \( g(h_{i,t} | \tilde{y}_{i,t}) \) uses information of the whole sample to draw \( h_{i,t} \).

Moreover, the IS estimators (3.31) and (3.32) allow us to obtain an approximation for the unknown error terms \( (a_{i,1} \epsilon_{1,t} + \cdots + a_{i,i-1} \epsilon_{i-1,t}) \). Recall the residual term of first equation of a TVP-VAR,

\[
\hat{\epsilon}_{1,t} = y_{1,t} - \hat{b}_{1,t} X_t.
\]

Given draws of the time-varying parameters \( b_{i,t} \) the IS approximation \( \tilde{\epsilon}_{1,t} \) can be computed as,

\[
\tilde{\epsilon}_{1,t} = \frac{\sum_{j=1}^{S} \omega_{j}(\gamma_{1}, \tilde{y}_{1,t}) (y_{1,t} - \hat{b}_{1,t}^{(j)} X_t)}{\sum_{j=1}^{S} \omega_{j}(\gamma_{1}, \tilde{y}_{1,t})}.
\]

Note that the quantity (3.34) is merely a weighted average of the residuals of the first equation of the TVP-VAR. The general equation for an estimator of the error terms can be written as,

\[
\tilde{\epsilon}_{i,t} = \frac{\sum_{j=1}^{S} \omega_{j}(\gamma_{i}, \tilde{y}_{i,t}) (y_{i,t} - \hat{b}_{i,t}^{(j)} X_t)}{\sum_{j=1}^{S} \omega_{j}(\gamma_{i}, \tilde{y}_{i,t})}.
\]

It is possible to compute the weighted average (3.35) using filtered or smoothed draws of \( b_{i,t} \). In most cases, smoothed draws are preferable as they incorporate information from all the sample and are better estimates of the real \( b_{i,t} \). Throughout this paper we use smoothed estimates of \( b_{i,t} \) to compute \( \tilde{\epsilon}_{i,t} \) based on equation (3.35). The above IS approximations \( (\tilde{\epsilon}_{1,t}, \ldots, \tilde{\epsilon}_{N,t}) \) does not converge to the true error terms, as disturbances cannot be consistently estimated. However, the estimated residuals incorporate information about the volatility states, and they are readily available. In the next section we perform two Monte Carlo studies to evaluate the properties of the proposed estimation procedure. To preview some MC results we have found that the procedure seems to display convergence to the DGP as \( T \to \infty \), and the standard deviations tend to become smaller with \( T \to \infty \).

Finally, with the IS approximation (3.35) the equationwise procedure for the TVP-VAR is completed and the summarized algorithm reads as follows,

---

6Filtering process for non-linear state-space models are usually based on particle filters. See Scharth and Kohn (2016) for a particle filter based on efficient importance sampling. On the other hand DeJong et al. (2013) developed a filtering process that does not depend on particles and resampling.

7In addition, states cannot be estimated consistently, but only aspects of their distribution.
1. Write the TVP-VAR in a triangular form.

2. Estimate the parameters of the first equation by the RB-NAIS method. Obtain the IS approximation $\tilde{\epsilon}_{1,t}$ using formula (3.34).

3. Replace $\sigma_{1,t}\epsilon_{1,t}$ in second equation by $\tilde{\epsilon}_{1,t}$. Estimate the static parameters by the RB-NAIS method and obtain $\tilde{\epsilon}_{2,t}$ using formula (3.35).

4. Proceed to the next equation. Replace $\sigma_{1,t}\epsilon_{1,t}$ and $\sigma_{2,t}\epsilon_{2,t}$ by $\tilde{\epsilon}_{1,t}$ and $\tilde{\epsilon}_{2,t}$ respectively. Estimate the parameters by the RB-NAIS method. Repeat this process until all equations and parameters are estimated.

### 3.3.3.1 Joint Filtering and Smoothing

The equationwise estimation procedure previously described allows the estimation of the model parameter and the filtering and smoothing of state variables. Although joint estimation of the model parameters is very complicated via maximum likelihood or MCMC, the joint filtering with a fixed vector of parameters is feasible. We present three algorithms that may be used for this purpose. The first one is the multivariate RB-NAIS which is discussed in the supplementary appendix. The multivariate NAIS is able to deliver smoothed estimates when the non-linear state vector is of dimension up to 3. For larger state vector dimensions the algorithm suffers heavily from the curse of dimensionality.

The second method is the Rao-Blackwellized Efficient Importance Sampling (RB-EIS) developed in Moura and Turatti (2014). The RB-EIS is able to produce smoothed estimates for non-linear state vectors of dimension up to 4. The RB-EIS does not suffer much from the dimensionality of the problem but it can be highly inefficient if the number of draws do not increase exponentially with the non-linear state-vector size.

The third procedure is the Particle Filter (PF). The PF does not suffer from the curse of dimensionality but it uses a blind proposal as importance sampler, thus it incorporates information about the non-linear states only through the resampling step. Note that the PF yields only filtered estimates. A description of these three algorithms can be found in the supplementary appendix.

### 3.4 Monte Carlo Study

The equationwise simulated maximum likelihood method developed in the previous section allows for a convenient equation-by-equation estimation. However, its finite sample behavior is unknown. In this section, we perform a Monte Carlo experiment

---

8 Although the RB-EIS can in theory handle high-dimensional cases it is usually found that the procedure becomes too numerically complex and the importance parameters tend to diverge very quickly. Experience suggests that up to dimension 4 it is feasible to filter the state variables.
to investigate the small sample properties of our equation-by-equation procedure on different scenarios of the DGP. We perform Monte Carlo studies on the basis of two models. The first one estimate a TVP-VAR(1) with 3 variables. The idea is to obtain sample average and standard deviations of the parameters estimates. The second experiment estimates a TVP-VAR(1) with 2 variables by our equationwise procedure and by a joint estimation via the RB-NAIS in a view of comparing the results.

### 3.4.1 Study Design

In the first Monte Carlo study we are interested in investigating the finite-sample and the asymptotic properties of our equation-by-equation estimation procedure for the TVP-VAR. For this study we set the following TVP-VAR(1) model with 3 variables,

\[
y_{1,t} = \beta_{1,t} + \beta_{11,t} y_{1,t-1} + \beta_{12,t} y_{2,t-1} + \beta_{13,t} y_{3,t-1} + \epsilon_{1,t}, \tag{3.36}
\]

\[
y_{2,t} = \beta_{2,t} + \beta_{21,t} y_{1,t-1} + \beta_{22,t} y_{2,t-1} + \beta_{23,t} y_{3,t-1} + \epsilon_{2,t}, \tag{3.37}
\]

\[
y_{3,t} = \beta_{3,t} + \beta_{31,t} y_{1,t-1} + \beta_{32,t} y_{2,t-1} + \beta_{33,t} y_{3,t-1} + \epsilon_{3,t}, \tag{3.38}
\]

where the vector of random innovations is \( \epsilon_t \sim N(0_3, H_t) \), and \( H_t = A^{-1} \Sigma_t A^{-1}' \). The state-variables \( B_t \) and \( \Sigma_t \) follow driftless random walk processes,

\[
\beta_{i,j,t} = \beta_{i,j,t-1} + q_{i,j} \eta_{i,j,t}, \quad \forall i, j = 1, 2, 3, \tag{3.39}
\]

\[
\beta_{i,t} = \beta_{i,t-1} + q_i \eta_{i,t}, \quad \forall i = 1, 2, 3, \tag{3.40}
\]

\[
h_{i,t} = h_{i,t-1} + \sigma_i \eta_{i,t}, \quad \forall i = 1, 2, 3. \tag{3.41}
\]

We examine the properties of our estimation method in 2 cases: (I) high parameter variation intensity of time-varying parameters and stochastic volatility (II) moderate parameter variation intensity of time-varying parameters and stochastic volatility. Thus, the DGPs are set as,

<table>
<thead>
<tr>
<th>Table 3.1: Data Generating Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II</td>
</tr>
</tbody>
</table>

where \( q_{ii} \) means the standard deviation when \( i = j \), and \( q_{ij} \) in the case \( i \neq j \). \( \beta_{i,j,t} \) and \( \beta_{i,t} \) are defined as slightly more volatile than \( \beta_{i,j,t} \). The samples sizes are \( T = 250, 500, 1000 \). The last sample size is bigger than most applications of TVP-VAR models and it is intended to verify the asymptotic properties of the estimator. The number of replications is set to 250. 15 GH nodes are used to compute the importance parameters and 100 draws to evaluate the likelihood for each equation.

The second study aims to compare the estimates of our equationwise procedure and the joint estimation of the full model from a finite-sample perspective. For this
3.4. MONTE CARLO STUDY

experiment we work with a small TVP-VAR(1) with 2 variables to minimize the curse of dimensionality when estimating the joint model. The TVP-VAR(1) model is,

\[ y_{1,t} = \beta_{1,t} + \beta_{11,t}y_{1,t-1} + \beta_{12,t}y_{2,t-1} + \epsilon_{1,t}, \quad (3.42) \]
\[ y_{2,t} = \beta_{2,t} + \beta_{21,t}y_{1,t-1} + \beta_{22,t}y_{2,t-1} + \epsilon_{2,t}, \quad (3.43) \]

where the vector of random innovations is \( \epsilon_t \sim N(0_2, H_t) \), and \( H_t = A^{-1}\Sigma_tA^{-1}' \). The state variable dynamics are the same as in the first experiment. The DGPs will be the same as DGP II previously defined, with exception of \( a_{2,1} \),

Table 3.2: Data Generating Processes

<table>
<thead>
<tr>
<th>DGP</th>
<th>( q_{ii} )</th>
<th>( q_{ij} )</th>
<th>( \sigma_i )</th>
<th>( \sigma_j )</th>
<th>( a_{2,1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>III</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.6</td>
</tr>
<tr>
<td>IV</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>1</td>
</tr>
</tbody>
</table>

The idea is to increase the covariance between \( y_{1,t} \) and \( y_{2,t} \) to check whether it affects the comparison between the estimation procedures. DGP III can be understood as medium correlated \( y_t \) meanwhile DGP IV is highly correlated. As before the the samples sizes are \( T = 250, 500 \), and the number of replications is also set to 250. 15 GH nodes are used to compute the importance parameters and 100 draws to evaluate the likelihood for the equation-wise procedure. In the joint estimation case we increase the number of draws to evaluate the likelihood function to 200. This is important for a fair comparison as the non-linear state vector has increased its dimensionality thus a larger number of draws may be necessary to cover the whole space of integration. The joint estimation will be carried out by the Rao-Blackwellized Multivariate Numerically Accelerated Importance Sampling which is discussed in the supplementary appendix. In this section the time-varying parameters and the stochastic volatility were initialized with a diffuse procedure.

Finally, note that both models have time-varying linear states \( B_{i,t} \) which follow a random walk. This means that they may hit any upper or lower bound in sufficiently large sample sizes. To avoid the divergence of \( \{y_t\}_{t=1}^T \) we perform stationarity checks at each point in time and explosive paths were discarded. This is particular relevant for computing estimates when \( T = 1000 \). In addition, to properly obtain a simulated path from the model the burn-in period is set as 500.

3.4.2 Results

Table [3.3] (in appendix) presents mean and standard deviations of the parameter estimates for the first model. We note that for small sample sizes \( (n = 250) \), all the parameters have large standard deviations, but for \( n = 500 \) and \( n = 1000 \) they become much smaller. However, the sample average is sufficiently close to the DGP even for a small size as 250. Most importantly the parameters seem to approach the DGP as
the sample size increases, and at \( n = 1000 \) the parameter estimates are very close to the real ones and the standard deviations are small, even for the terms in \( A^{-1} \). Although the convergence for the terms in the stochastic volatility equation seems to be somewhat slower, possibly because they affect all equations, and errors may accumulate through the steps. Finally, no relevant difference is observed in terms of DGP I and II.

Table 3.4 (in appendix) shows the results of the estimation for the second model. The objective of this study is to compare the estimates of our equationwise procedure and the joint estimation of all model parameters in one step. We aim to compare from a finite-sample perspective. Results indicate that the averages of the estimates are very similar. Exceptions are the intercepts in the case \( T = 250 \), meaning they incorporate the approximation error of the equation-by-equation procedure, especially in the high correlation case. However, for \( T = 500 \) all estimates are in fact very close across methods. This provides strong support for the equationwise procedure as a sound estimator and comparable to the joint procedure in large sample sizes.

We conclude from these experiments that the equation-by-equation estimation procedure is able to properly identify the model parameters and seems to display convergence as the sample size increases. Therefore, we have found MC evidence that the procedure has good finite sample and asymptotic properties and can be applied to improve estimation of TVP-VAR models. Moreover, the average of parameter estimates are very close to their true values even in small sample sizes, and the sample standard deviations are small. Finally, estimates are comparable to the one-step procedure even in finite-samples.

### 3.5 Empirical Application

In order to examine the abilities of our estimation procedure to describe real data we apply it to a dataset similar that of Cogley and Sargent (2005) and Primiceri (2005). The data are composed by the following time series,

- **Short term interest rate:** secondary rate on 3-month Treasury Bills. The data are sampled monthly and then converted to a quarterly series by selecting the first month of each quarter. We use the first difference of these values to ensure stationarity.

- **Unemployment rate:** civilian unemployment seasonally adjusted and sampled monthly. The data are converted from monthly to quarterly series by computing the average. As before the first difference was taken to ensure stationarity.

- **Inflation:** Consumer Price Index for all urban consumers, all items. The data are seasonally adjusted and sampled monthly. It is converted to a quarterly basis by taking the last month of the quarter. For stationarity purposes we work
with changes in inflation computed as the second order log-difference of the consumer price index of the quarter.

The data were obtained from the Federal Reserve Economic Database (FRED) and have codes TB3MS, UNRATE, and CPIAUCSL respectively. The sample runs from 1948:Q1 to 2000:Q4 and consists of 212 observations. The sample end choice is to match those of Cogley and Sargent (2005) and Primiceri (2005). We use the sample from 1948:Q3 to 1958:Q4 as “training” to calibrate the initial conditions for the non-stationary time-varying parameters $B_{i,t}$ and $h_{i,t}$. Therefore, the model is estimated from 1960:Q1 to 2000:Q4, summing up a total of 165 observations. The estimation is designed to resemble as close as possible the one from Cogley and Sargent (2005).

The model we consider is the TVP-VAR described in equations (3.1)- (3.3). Two lags are used in the estimation. Recall the TVP-VAR,

$$
\begin{align*}
y_t &= X_t' B_t + Y(H_t) \epsilon_t, \quad \epsilon_t \sim N(0, I_3), \\
\text{vec}(B_{t+1}') &= \text{vec}(B_t') + Y(Q) \eta_t, \quad \eta_t \sim N(0, I_{21}),
\end{align*}
$$

where $X_t = [y_t', y_{t-1}', y_{t-2}']'$. The time-varying covariance is defined as $H_t = A^{-1} \Sigma_t A^{-1}'$, where $A$ is a $3 \times 3$ lower unit triangular matrix, and $\Sigma_t$ is a $3 \times 3$ diagonal matrix of time-varying volatilities. The sample from 1948:Q3 to 1958:Q4 is used to calibrate the initial conditions for the non-stationary time-varying parameters $B_{i,t}$ and $h_{i,t}$ by estimating a time-invariant vector autoregression. This is analogous to the Bayesian idea of informative priors. The mean for $B_{i,0|0}$ is selected as OLS point estimates for $B_{i,t}$, and the variance is selected as 2 times the OLS estimates. For the log-volatilities $h_{i,t}$ the initial condition is chosen to be the logarithm of the OLS point estimates of the standard errors of the same time invariant VAR, and its variance is set as 10, making this initial condition weakly informative. The adopted initialization strategy is comparable to the one of Cogley and Sargent (2005) and Primiceri (2005).

Next, as discussed in section 2 the order of variables matter in a TVP-VAR under a triangular decomposition of the covariance matrix. For example, Sims (1980) reported that the ordering of variables in an identified VAR mattered for a comparison of interwar and postwar business cycles. We follow Cogley and Sargent (2005) in which the ordering is selected to minimize the time-variation in the coefficients $B_{i,t}$, namely,

$$
y_t = \begin{bmatrix}
\text{interest}_t \\
\text{unemployment}_t \\
\text{inflation}_t
\end{bmatrix}.
$$

This ordering means that the interest rates are the most exogenous at time $t$ and it has its own own source of stochastic volatility, meanwhile inflation suffers influence from all other stochastic volatilities.

9However in their specification the matrix $Q$ is not diagonal, therefore, the results may somewhat differ.
The model is estimated by the equationwise-NAIS procedure and the state-variables are jointly filtered by the multivariate NAIS method discussed in the supplementary appendix. Figures 3.1 and 3.2 (in appendix) show the smoothed volatilities and covariances. Qualitatively, the results are similar to those in Cogley and Sargent (2005). The general patterns in the standard deviations and the correlations interest-inflation, interest-unemployment are the same, changing only the magnitude. In general, our model and method report lower levels of volatilities and correlations. This can be caused by the different model specification or by the different data treatment. Overall, volatilities of inflation and interest rate increased during late 1970s as a result of the Great Inflation and the Volcker disinflation shocks. Unemployment follows a different pattern and its level of volatility only slightly increases during the Great Inflation. Note that our model and method yield a correlation between inflation and unemployment very close to zero, however, in Cogley and Sargent (2005) it is also close to zero throughout the sample apart from the Volcker shock from late 1970s to mid 1980s. Finally, the correlations closely mimic the standard deviations, and this is expected given the triangular decomposition and the time-invariant matrix $A$.

We follow Cogley and Sargent (2005) and use the log-determinant of the covariance matrix, $\log|\mathbb{E}(H_t)|$, as the measure of short-term uncertainty. According to Whittle (1953) this as a measure of the total uncertainty entering the system at each time $t$. Figure 3.3 (in appendix) shows the results. Uncertainty increases in two steps, one between the late 1960s and the other between the late 1970s. Most of the subsequent decrease occurred in the mid-1980s, during the latter part of Volcker’s term. It is worth mentioning that our results are very similar to Cogley and Sargent (2005) in this regard.

In conclusion, the equationwise NAIS for the TVP-VAR is able to estimate the time-variation in real data. The results were mostly similar to those reported by Cogley and Sargent (2005). For future research it would be interesting to allow the matrix $A^{-1}$ be time-varying as in Primiceri (2005).

### 3.6 Conclusions

In this paper we discuss classical maximum simulated likelihood inference for time-varying vector autoregressive models with time-varying variances and covariances. Following the literature in which a special form of the covariance matrix is applied we propose a triangularization scheme where the variables are written from the most exogenous to the most endogenous.

We show that an equationwise estimation procedure is possible using estimated residuals as an approximation for unknown error terms, resulting in a sequence of time-varying autoregressive regressions with stochastic volatility. Finally, we propose the Rao-Blackwellized Numerically Accelerated Importance Sampling which combines Kalman Filtering with importance sampling to deliver a numerically and
A Monte Carlo study shows that the equationwise procedure has good finite-sample and asymptotic properties, and most parameter estimates are very close to the DGP. In large sample sizes, the sample standard deviations become smaller and the estimates are very close to their real values. Another MC study compares the equation-by-equation procedure with the joint estimation of the model parameters, and it shows that the difference between them is very small in finite samples, and in large samples they yield almost equal results.

An empirical application that mimics the one from Cogley and Sargent (2005) shows that the procedure captures most of the patterns discussed in the macroeconomics literature, and correlations and volatilities are close to those reported by Cogley and Sargent (2005) and Primiceri (2005). For future research, it would be interesting to allow the matrix $A$ to be time-varying. Another interesting possibility for future research would be to estimate a large TVP-VAR with stochastic volatility. Most methods available in the literature are hardly applicable for a large dimensional covariance matrix.

### 3.7 References


3.8 Appendix

3.8.1 Tables

In this section we report the tables discussed in the MC study section 3.4. Table 3.3 presents average estimates and sample standard errors from 250 series generated by DGP I and II and estimated by the equationwise NAIS procedure. Table 3.4 shows the average of estimates and sample standard errors generated by DGP III and IV and estimated by the equationwise NAIS method and the Multivariate RB-NAIS.
### Table 3.3: Finite-Sample properties of the equationwise procedure

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP</th>
<th>$T = 250$</th>
<th>$T = 500$</th>
<th>$T = 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_1$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.042 (0.044)</td>
<td>0.039 (0.031)</td>
</tr>
<tr>
<td>$q_{11}$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.034 (0.015)</td>
<td>0.035 (0.010)</td>
</tr>
<tr>
<td>$q_{12}$</td>
<td>DGP I</td>
<td>0.03</td>
<td>0.028 (0.015)</td>
<td>0.028 (0.009)</td>
</tr>
<tr>
<td>$q_{13}$</td>
<td>DGP I</td>
<td>0.03</td>
<td>0.026 (0.013)</td>
<td>0.028 (0.008)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.076 (0.050)</td>
<td>0.061 (0.033)</td>
</tr>
<tr>
<td>$q_2$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.043 (0.043)</td>
<td>0.041 (0.029)</td>
</tr>
<tr>
<td>$q_{21}$</td>
<td>DGP I</td>
<td>0.03</td>
<td>0.027 (0.017)</td>
<td>0.028 (0.009)</td>
</tr>
<tr>
<td>$q_{22}$</td>
<td>DGP I</td>
<td>0.03</td>
<td>0.034 (0.015)</td>
<td>0.036 (0.010)</td>
</tr>
<tr>
<td>$q_{23}$</td>
<td>DGP I</td>
<td>0.03</td>
<td>0.028 (0.016)</td>
<td>0.028 (0.009)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.074 (0.050)</td>
<td>0.061 (0.034)</td>
</tr>
<tr>
<td>$a_{2,1}$</td>
<td>DGP I</td>
<td>0.30</td>
<td>0.306 (0.102)</td>
<td>0.312 (0.070)</td>
</tr>
<tr>
<td>$q_3$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.041 (0.042)</td>
<td>0.041 (0.027)</td>
</tr>
<tr>
<td>$q_{31}$</td>
<td>DGP I</td>
<td>0.03</td>
<td>0.027 (0.015)</td>
<td>0.028 (0.009)</td>
</tr>
<tr>
<td>$q_{32}$</td>
<td>DGP I</td>
<td>0.03</td>
<td>0.028 (0.016)</td>
<td>0.028 (0.009)</td>
</tr>
<tr>
<td>$q_{33}$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.034 (0.015)</td>
<td>0.036 (0.009)</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>DGP I</td>
<td>0.04</td>
<td>0.073 (0.051)</td>
<td>0.059 (0.032)</td>
</tr>
<tr>
<td>$a_{3,1}$</td>
<td>DGP I</td>
<td>0.30</td>
<td>0.300 (0.104)</td>
<td>0.296 (0.067)</td>
</tr>
<tr>
<td>$a_{3,2}$</td>
<td>DGP I</td>
<td>0.30</td>
<td>0.307 (0.102)</td>
<td>0.302 (0.071)</td>
</tr>
<tr>
<td>$q_1$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.033 (0.034)</td>
<td>0.028 (0.022)</td>
</tr>
<tr>
<td>$q_{11}$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.025 (0.013)</td>
<td>0.025 (0.009)</td>
</tr>
<tr>
<td>$q_{12}$</td>
<td>DGP II</td>
<td>0.02</td>
<td>0.018 (0.012)</td>
<td>0.018 (0.009)</td>
</tr>
<tr>
<td>$q_{13}$</td>
<td>DGP II</td>
<td>0.02</td>
<td>0.014 (0.010)</td>
<td>0.018 (0.008)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.059 (0.043)</td>
<td>0.047 (0.029)</td>
</tr>
<tr>
<td>$q_2$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.034 (0.042)</td>
<td>0.029 (0.021)</td>
</tr>
<tr>
<td>$q_{21}$</td>
<td>DGP II</td>
<td>0.02</td>
<td>0.017 (0.013)</td>
<td>0.019 (0.008)</td>
</tr>
<tr>
<td>$q_{22}$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.024 (0.013)</td>
<td>0.026 (0.008)</td>
</tr>
<tr>
<td>$q_{23}$</td>
<td>DGP II</td>
<td>0.02</td>
<td>0.017 (0.012)</td>
<td>0.018 (0.007)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.058 (0.041)</td>
<td>0.047 (0.026)</td>
</tr>
<tr>
<td>$a_{2,1}$</td>
<td>DGP II</td>
<td>0.30</td>
<td>0.299 (0.090)</td>
<td>0.302 (0.060)</td>
</tr>
<tr>
<td>$q_3$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.033 (0.034)</td>
<td>0.030 (0.020)</td>
</tr>
<tr>
<td>$q_{31}$</td>
<td>DGP II</td>
<td>0.02</td>
<td>0.016 (0.014)</td>
<td>0.019 (0.009)</td>
</tr>
<tr>
<td>$q_{32}$</td>
<td>DGP II</td>
<td>0.02</td>
<td>0.018 (0.014)</td>
<td>0.018 (0.008)</td>
</tr>
<tr>
<td>$q_{33}$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.025 (0.014)</td>
<td>0.027 (0.009)</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>DGP II</td>
<td>0.03</td>
<td>0.057 (0.039)</td>
<td>0.044 (0.024)</td>
</tr>
<tr>
<td>$a_{3,1}$</td>
<td>DGP II</td>
<td>0.30</td>
<td>0.300 (0.089)</td>
<td>0.295 (0.062)</td>
</tr>
<tr>
<td>$a_{3,2}$</td>
<td>DGP II</td>
<td>0.30</td>
<td>0.307 (0.085)</td>
<td>0.304 (0.064)</td>
</tr>
</tbody>
</table>

The table presents the DGP parameters followed by sample average of estimated parameters from 250 series. Values in parentheses are sample standard deviations of the estimates.
### Table 3.4: Finite-Sample properties of the equationwise procedure

<table>
<thead>
<tr>
<th>Parameters</th>
<th>DGP III</th>
<th>DGP IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Equationwise procedure</td>
<td>One-Step procedure</td>
</tr>
<tr>
<td>$q_1$</td>
<td>0.03 0.025 (0.015) 0.025 (0.009)</td>
<td>0.03 0.025 (0.015) 0.025 (0.009)</td>
</tr>
<tr>
<td>$q_{11}$</td>
<td>0.03 0.025 (0.015) 0.025 (0.009)</td>
<td>0.03 0.025 (0.015) 0.025 (0.009)</td>
</tr>
<tr>
<td>$q_{12}$</td>
<td>0.03 0.054 (0.038) 0.046 (0.026)</td>
<td>0.03 0.054 (0.038) 0.046 (0.026)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.03 0.054 (0.038) 0.046 (0.026)</td>
<td>0.03 0.054 (0.038) 0.046 (0.026)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.02 0.017 (0.014) 0.018 (0.008)</td>
<td>0.02 0.017 (0.014) 0.018 (0.008)</td>
</tr>
<tr>
<td>$\alpha_{2,1}$</td>
<td>0.03 0.059 (0.042) 0.045 (0.028)</td>
<td>0.03 0.059 (0.042) 0.045 (0.028)</td>
</tr>
</tbody>
</table>

The table presents the DGP parameters followed by sample average of estimated parameters from 250 series. Values in parentheses are sample standard deviations of the estimates.

#### 3.8.2 Figures

This section reports the figures with results of the volatilities and correlations discussed in the text.
Figure 3.1: Standard deviations of interest rates, unemployment and inflation
Figure 3.2: Correlations between interest rates, unemployment and inflation

Figure 3.3: Total variance
3.9 Supplementary Appendix

This supplementary appendix provides additional material supporting the main text. Section 2 provides additional details on the multivariate RB-NAIS used in the main text to jointly filter the state-variables. Section 3 reviews an alternative method for filtering and smoothing also based on importance sampling, the RB-EIS. Section 4 presents a particle filter for the TVP-VAR with fixed static parameters.

3.9.1 Multivariate RB-NAIS

The extension of NAIS to the multivariate case is straightforward, in fact, only minor adjustments are necessary. Let \( h_t \) be the \( q \)-vector of stochastic volatilities at time \( t \). The key element in the multivariate NAIS is the representation of \( g(y|\mathbf{h}, \varphi) \) as a multivariate linear and Gaussian state space density,

\[
g(y_t|h_t) = \exp(a_t + b'_t h_t - \frac{1}{2} h'_t C_t h_t), \quad (3.47)
\]

where \( a_t \) is an integrating constant, and the importance parameters \( b_t \in \mathbb{R}^q \) and \( C_t \in \mathbb{R}^{q \times d} \) for \( t = 1, \ldots, T \) are a function of \( y_t \) and \( \gamma \). Therefore, we have to select \( \{b_t\}_t^{T=1} \) and \( \{C_t\}_t^{T=1} \) to construct the importance density. Note that \( C_t \) is a matrix and it will yield the covariance matrix of the importance sampler.

For a given set of importance parameters, \( \varphi_t = (b_t, C_t) \), the proposal density \( g(y_t|h_t, \varphi) \) can be computed via the artificial observations \( y^*_t = C_t^{-1} b_t \) with measurement equation,

\[
y^*_t = h_t + \epsilon_t, \quad \epsilon_t \sim N(0, C_t^{-1}), \quad (3.48)
\]

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

\[
\log g(y^*|\mathbf{h}, \varphi) = -\frac{1}{2} \log(2\pi)^k + \frac{1}{2} \log|C_t| - \frac{1}{2} [(C_t^{-1} b_t - h_t)' C_t (C_t^{-1} b_t - h_t)] \quad (3.49)
\]

\[
= a_t + b'_t h_t - \frac{1}{2} h'_t C_t h_t, \quad (3.50)
\]

The linear state-space representation \(3.48\) implies that we can use the Kalman filter and smoothing techniques to evaluate and sample from \( g(h_t|y_t, \varphi_t) \), thus resulting in a computationally efficient and simple algorithm. Additionally, the smoothing density \( g(h_t|y_t, \varphi_t) \) is available analytically and is given by,

\[
g(h_t|y_t, \varphi_t) = \frac{1}{2\pi|V_t|} \exp\left(-\frac{1}{2} (h_t - \hat{h}_t)' (V_t)^{-1} (h_t - \hat{h}_t)\right), \quad (3.51)
\]

where \( \hat{h}_t \) and \( V_t \) are the mean and the variance of the signal vector, and are obtained by the KFS applied to the linear state-space \(3.48\) for a given \( \varphi \).
Moreover, the importance parameters $\varphi$ are selected by approximating the importance sampler and the likelihood integrand. Richard and Zhang (2007) proposed that the appropriate Gaussian density can be found by solving for every time $t$,

$$
\varphi_t = \arg\min_{\varphi_t} \lambda^2 (y_t, h_t|\varphi_t) \omega(h_t, y_t|\varphi_t) g(h_t|y_t, \varphi_t) h_t,
$$

(3.52)

where $\lambda^2(.)$ is defined as,

$$
\lambda^2 (y_t, h_t|\varphi_t) = \left( \log p^*(y_t|h_t) - \log g(y_t^*|h_t, \varphi_t) \right)^2.
$$

(3.53)

To solve (3.52) and obtain the importance parameters $(b_t, C_t)$ quadrature methods via Gauss-Hermite are used. Note that a main complication arises as the importance sampler $g(h_t|y_t, \varphi_t)$ itself depends upon the importance parameters $\varphi_t$. As in Richard and Zhang (2007) this can be resolved by a standard fixed point argument using intermediate importance samplers until convergence of $\varphi = \{(\varphi_t)_{t=1}^T\}$. To initiate the NAIS procedure an initial condition is required for $\varphi$, which will be sequentially updated to $\varphi^{(k)}$ through iterations on (3.52) until a convergence criteria is satisfied.

We can define,

$$
\Phi(y_t, h_t, \varphi^{(k)}) = \lambda^2 (y, h|\varphi^{(k)}) \omega(h, y|\varphi^{(k)}),
$$

(3.54)

where $\varphi^{(k)}$ is the set of importance parameters in the $k^{th}$ iteration. Let $\{z_i\}_{i=1}^{M^q}$ be the set of abscissae after all combinations of $M$ predefined GH nodes and $h(z_i)$ be the respective Gauss-Hermite weights. The minimization (3.52) is then implemented as,

$$
\arg\min_{\varphi^{k+1}} \sum_{i=1}^{M^q} w_{i,t} \Phi(y_t, Z_{i,t}^{(k)}, \varphi^{(k)}),
$$

(3.55)

with weights given by, $w_{i,t} = h(z_i) \exp(\frac{1}{2} z_i^2)$ and $Z_{i,t}^{(k)} = \tilde{h}_t^{(k)} + F_t^{(k)} z_i$, for $i = 1, \ldots, M^q$, where $F_t^{(k)}$ is the root matrix computed via the Cholesky decomposition of $V_t^{(k)}$. For this implementation, the distribution of the combined nodes is given by $g(Z_{i,t}^{(k)}|y, \varphi^{(k)}) \propto \exp(-\frac{1}{2} z_i^T z_i)$, as they are associated with the multivariate standard normal distribution. Finally, the minimization (3.52) can be written as a weighted least squares problem applied to $M^q$ observations for the auxiliary regression,

$$
\log p^*(y_t|Z_{i,t}^{(k)}) = \text{constant} + \kappa^T Z_{i,t}^{(k)} - \frac{1}{2} \xi^T \text{vech}(Z_{i,t}^{(k)} - Z_{i,t}^{(k)}) + \text{error},
$$

(3.56)

where $\text{vech}(.)$ stacks the elements of a matrix into a vector, and $\kappa_T$ and $\xi_T$ are regression coefficients determining the new estimates of $b_t^{(k+1)}$ and $C_t^{(k+1)}$ for each time $t = 1 \ldots T$. The weights for the auxiliary regression are given by $w_{i,t} w(Z_{i,t}^{(k)}|y_t, \varphi^{(k)}), g(Z_{i,t}^{(k)}|y, \varphi^{(k)})$.

Using the new estimates $b_t^{(k+1)}$ and $C_t^{(k+1)}$, we can obtain the smoothed means $\tilde{h}_t^{(k+1)}$ and variance $V_t^{(k+1)}$ for $g(h_t|y, \varphi)$. This procedure is iterated until convergence is achieved.
3.9.2 RB-EIS

The RB-EIS extends the Efficient Importance Sampling of Richard and Zhang (2007) to incorporate linear time-varying states. EIS differs from NAIS in the computation of the importance parameters and the sampling from the proposal density. Let \( g(h_t|y, \varphi) \) be the importance sampler density, and \( p^*(y_t|h_t; \gamma)p(h_t|h_{t-1}) \) be the target integrand, where \( p^*(y_t|h_t; \gamma) \) is evaluated by the Kalman filter conditional on a draw of the log-volatility \( h_t \). The importance sampler can be decomposed as the product of a kernel and an integrating constant,

\[
g(h_t|y, \varphi_t) = \frac{k(h_t; \varphi_t)}{\chi(h_{t-1}; \varphi_t)}, \quad \text{where} \quad \chi(h_{t-1}; \varphi_t) = \int k(h_t; \varphi) \, dh_t \quad (3.57)
\]

EIS minimizes the MC variance of likelihood integrand matching the Gaussian kernel \( k(h_t; \gamma) \) period by period not only to \( p^*(y_t|h_t; \gamma)p(h_t|h_{t-1}) \) but also to next period’s integrating constant \( \chi(h_t; \varphi_{t+1}) \). Note that period \( t+1 \) integrating constant, \( \chi(h_t; \varphi_{t+1}) \), depends solely on \( h_t \) and not on \( h_{t+1} \), and transferring it back to period \( t \) allows the efficient importance sampler of period \( t \) to incorporate important information on the dynamics of \( h_t \) contained in \( p^*(y_{t+1}, h_{t+1}; \gamma) \). The process of transferring back the integrating constant works like a smoother bringing information on \( h_t \) contained in \( y_{t+1} \), and requires a backward sequence of optimization, starting from the last period of the sample, \( T \), and going back to its first observation. The EIS minimization problem is then solved as a recursive sequence of least-squares problems going from \( T \to 1 \):

\[
\hat{\varphi}_t = \arg \min \sum_{i=1}^S \left\{ \ln \left[ p^*(y_i|h_t^t, \gamma)\chi(h_t^t; \varphi_{t+1}) \right] - c_t - k(h_t^t; \varphi_t) \right\}^2 \quad (3.58)
\]

where \( \chi(h_T; \varphi_{T+1}) \equiv 1 \), and trajectories \( \{\tilde{h}_t^T\}_{t=0}^T \) are drawn from an initial sampler.

Once estimates of \( \hat{\varphi}_t \) for \( t = 1 \) are obtained, they can be used to estimate the likelihood function for a given value of \( \gamma \). Note that the natural logarithm of density kernels from the exponential family of distributions are linear in their natural parametrization (see Lehmann (1986), Section 2.7), which implies that the least-squares problem in (12) simplifies to a sequence of ordinary least-squares regressions.

There are some drawbacks in the EIS in comparison to NAIS. First, the moments of the importance sampler and the integrating constant are not easily available and the researcher has to compute them analytically prior to the estimation. Moreover, the computation of the importance parameters by Monte Carlo integration implies that a large number of draws is necessary, and it may significantly slow down the algorithm. Finally, the use of Monte Carlo integration usually implies that a large number of iterations may be necessary if the initial sampler is blind.
### 3.9.3 Particle Filter

The multivariate NAIS and the EIS are smoothed method, therefore, it provides a smoothed estimate for the state variables. However, there are cases when we are interested in filtered estimates, for example in residuals diagnostics. The main difference is that smoothed estimates use information from the whole sample, thus we obtain $E(h_{1:T}|T)$, and filtered estimates uses information up to time $t E(h_{1:t}|t)$. In this section we discuss a method to obtain filtered estimates.

The Rao-Blackwellized likelihood function (3.16) can be estimated by the particle filter. However, differently from the NAIS method, the auxiliary particle filter draws from a proposal density that does not carry information about the underlying states. More specifically, the PF sets the proposal sampler $g(h_t|y_t)$ to be the transition density $p(h_t|h_{t-1}; \gamma)$. Therefore, the draws are not conditional on the observables. For a detailed discussion on particle filtering we refer to Creal (2012). The complete PF for the generic equation of the TVP-VAR reads as follows,

1. Initialize the procedure with a diffuse intial condition for the linear time-varying parameters.
2. Draw the volatility terms, $\{h_{t}^2\}_{t=1}^{T}$ and the from the transition density $p(h_t|h_{t-1})$.
3. For each time $t = 1 \ldots T$ evaluate the density $p^*(y_t|h_t; \gamma)$ by the prediction error decomposition given the draws from step 2 and obtain the importance weights as, $\omega(\theta, y) = p^*(y_t|h_t; \gamma)$.
4. Normalize the importance weights, $\tilde{\omega}^{(i)} = \frac{\omega^{(i)}}{\sum_{j=1}^{S} \omega^{(j)}}$.
5. Resample with replacement the draws on the basis of the normalized weights $\tilde{\omega}^{(i)}$.
6. Apply the Kalman Filter to estimate the the linear time-varying parameters.
7. Estimate the likelihood as $\hat{L}(\theta, y) = \frac{1}{N} \sum_{i=1}^{N} \left[ \prod_{t=1}^{T} p^*(y_t|h_t; \gamma) \right]$.

### 3.9.4 References


PhD dissertations since 1 July 2011

2011-4  Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics
2011-5  Christian Bach: The Game of Risk
2011-6  Stefan Holst Bache: Quantile Regression: Three Econometric Studies
2011:12 Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management
2011:13 Christian Gormsen Schmidt: Exploring the Barriers to Globalization
2011:16 Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders
2011:22 Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade
2012-1 Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting
2012-2 Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping
2012-3 Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing
2012-4 Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing
2012-5 Lasse Frisgaard Gunnersen: Income Redistribution Policies
2012-6 Miriam Wüst: Essays on early investments in child health
2012-7 Yukai Yang: Modelling Nonlinear Vector Economic Time Series
2012-9 Henrik Nørholm: Structured Retail Products and Return Predictability
2012-10 Signe Frederiksen: Empirical Essays on Placements in Outside Home Care
2012-11 Mateusz P. Dziubinski: Essays on Financial Econometrics and Derivatives Pricing
<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-12</td>
<td>Jens Riis Andersen: Option Games under Incomplete Information</td>
</tr>
<tr>
<td>2012-14</td>
<td>Laurent Callot: Large Panels and High-dimensional VAR</td>
</tr>
<tr>
<td>2012-15</td>
<td>Christian Rix-Nielsen: Strategic Investment</td>
</tr>
<tr>
<td>2013-1</td>
<td>Kenneth Lykke Sørensen: Essays on Wage Determination</td>
</tr>
<tr>
<td>2013-2</td>
<td>Tue Rauff Lind Christensen: Network Design Problems with Piecewise Linear Cost Functions</td>
</tr>
<tr>
<td>2013-4</td>
<td>Rune Bysted: Essays on Innovative Work Behavior</td>
</tr>
<tr>
<td>2013-5</td>
<td>Mikkel Nørlem Hermansen: Longer Human Lifespan and the Retirement Decision</td>
</tr>
<tr>
<td>2013-7</td>
<td>Mark Strøm Kristoffersen: Essays on Economic Policies over the Business Cycle</td>
</tr>
<tr>
<td>2013-8</td>
<td>Philipp Meinen: Essays on Firms in International Trade</td>
</tr>
<tr>
<td>2013-9</td>
<td>Cédric Gorinas: Essays on Marginalization and Integration of Immigrants and Young Criminals – A Labour Economics Perspective</td>
</tr>
<tr>
<td>2013-12</td>
<td>Paola Andrea Barrientos Quiroga: Essays on Development Economics</td>
</tr>
<tr>
<td>2013-13</td>
<td>Peter Bodnar: Essays on Warehouse Operations</td>
</tr>
<tr>
<td>2013-14</td>
<td>Rune Vammen Lesner: Essays on Determinants of Inequality</td>
</tr>
<tr>
<td>2013-15</td>
<td>Peter Arendorf Bache: Firms and International Trade</td>
</tr>
<tr>
<td>2013-16</td>
<td>Anders Laugesen: On Complementarities, Heterogeneous Firms, and International Trade</td>
</tr>
<tr>
<td>Year</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2013-17</td>
<td>Anders Bruun Jonassen: Regression Discontinuity Analyses of the Disincentive Effects of Increasing Social Assistance</td>
</tr>
<tr>
<td>2014-1</td>
<td>David Sloth Pedersen: A Journey into the Dark Arts of Quantitative Finance</td>
</tr>
<tr>
<td>2014-3</td>
<td>Lukas Bach: Routing and Scheduling Problems - Optimization using Exact and Heuristic Methods</td>
</tr>
<tr>
<td>2014-4</td>
<td>Tanja Groth: Regulatory impacts in relation to a renewable fuel CHP technology: A financial and socioeconomic analysis</td>
</tr>
<tr>
<td>2014-5</td>
<td>Niels Strange Hansen: Forecasting Based on Unobserved Variables</td>
</tr>
<tr>
<td>2014-6</td>
<td>Ritwik Banerjee: Economics of Misbehavior</td>
</tr>
<tr>
<td>2014-7</td>
<td>Christina Annette Gravert: Giving and Taking – Essays in Experimental Economics</td>
</tr>
<tr>
<td>2014-8</td>
<td>Astrid Hanghøj: Papers in purchasing and supply management: A capability-based perspective</td>
</tr>
<tr>
<td>2014-9</td>
<td>Nima Nonejad: Essays in Applied Bayesian Particle and Markov Chain Monte Carlo Techniques in Time Series Econometrics</td>
</tr>
<tr>
<td>2014-10</td>
<td>Tine L. Mundbjerg Eriksen: Essays on Bullying: an Economist’s Perspective</td>
</tr>
<tr>
<td>2014-11</td>
<td>Sashka Dimova: Essays on Job Search Assistance</td>
</tr>
<tr>
<td>2014-12</td>
<td>Rasmus Tangsgaard Varneskov: Econometric Analysis of Volatility in Financial Additive Noise Models</td>
</tr>
<tr>
<td>2015-1</td>
<td>Anne Floor Brix: Estimation of Continuous Time Models Driven by Lévy Processes</td>
</tr>
<tr>
<td>2015-2</td>
<td>Kasper Vinther Olesen: Realizing Conditional Distributions and Coherence Across Financial Asset Classes</td>
</tr>
<tr>
<td>2015-3</td>
<td>Manuel Sebastian Lukas: Estimation and Model Specification for Econometric Forecasting</td>
</tr>
<tr>
<td>2015-4</td>
<td>Sofie Theilade Nyland Brodersen: Essays on Job Search Assistance and Labor Market Outcomes</td>
</tr>
<tr>
<td>Year</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2015-6</td>
<td>The Sibling Relationship Dynamics and Spillovers</td>
</tr>
<tr>
<td>2015-7</td>
<td>Empirical Studies of Earnings Quality</td>
</tr>
<tr>
<td>2015-8</td>
<td>The Relationship between Caseworkers and Unemployed Workers</td>
</tr>
<tr>
<td>2015-9</td>
<td>Empirical Essays on Heterogeneous Firms and International Trade</td>
</tr>
<tr>
<td>2015-10</td>
<td>Modelling and Optimisation of Renewable Energy Systems</td>
</tr>
<tr>
<td>2015-11</td>
<td>Essays in the Economics of Crime</td>
</tr>
<tr>
<td>2015-12</td>
<td>Solution Methods and Inference in Continuous-Time Dynamic Equilibrium Economies (with Applications in Asset Pricing and Income Fluctuation Models)</td>
</tr>
<tr>
<td>2015-13</td>
<td>The Internationalization of Big Accounting Firms and the Implications on their Practices and Structures: An Institutional Analysis</td>
</tr>
<tr>
<td>2015-14</td>
<td>Corporate Innovation, R&amp;D Personnel and External Knowledge Utilization</td>
</tr>
<tr>
<td>2015-15</td>
<td>Empirical Banking</td>
</tr>
<tr>
<td>2015-16</td>
<td>Structural and Empirical Analysis of the Labour Market</td>
</tr>
<tr>
<td>2015-17</td>
<td>Essays on International Economics and Development</td>
</tr>
<tr>
<td>2015-18</td>
<td>Lévy Semistationary Models with Applications in Energy Markets</td>
</tr>
<tr>
<td>2015-20</td>
<td>Business Cycles and Expected Returns</td>
</tr>
<tr>
<td>2015-21</td>
<td>Dynamic Models of the Housing Market</td>
</tr>
<tr>
<td>2016-1</td>
<td>Essays on International Migration: Institutions, Skill Recognition, and the Welfare State</td>
</tr>
<tr>
<td>2016-2</td>
<td>Essays on Forecasting with Linear State-Space Systems</td>
</tr>
<tr>
<td>2016-3</td>
<td>Financial Frictions, Price Rigidities, and the Business Cycle</td>
</tr>
<tr>
<td>2016-4</td>
<td>Volatility and Correlation in Financial Markets: Theoretical Developments and Numerical Analysis</td>
</tr>
</tbody>
</table>
2016-5 Anders Kronborg: Methods and Applications to DSGE Models
2016-6 Morten Visby Krægpøth: Empirical Studies in Economics of Education
2016-7 Anne Odile Peschel: Essays on Implicit Information Processing at the Point of Sale: Evidence from Experiments and Scanner Data Analysis
2016-8 Girum Dagnachew Abate: Essays in Spatial Econometrics
2016-9 Kai Rehwald: Essays in Public Policy Evaluation
2016-10 Reza Pourmoayed: Optimization Methods in a Stochastic Production Environment
2016-11 Sune Lauth Gadegaard: Discrete Location Problems – Theory, Algorithms, and Extensions to Multiple Objectives
2016-12 Lisbeth Palmhøj Nielsen: Empirical Essays on Child Achievement, Maternal Employment, Parental Leave, and Geographic Mobility
2016-13 Louise Voldby Beuchert-Pedersen: School Resources and Student Achievement: Evidence From Social and Natural Experiments
2016-14 Mette Trier Damgaard: Essays in Applied Behavioral Economics
2016-15 Andrea Barletta: Consistent Modeling and Efficient Pricing of Volatility Derivatives
2016-16 Thorvardur Tjörví Ölafsson: Macrofinancial Linkages and Crises in Small Open Economies
2016-17 Carlos Vladimir Rodriguez Caballero: On Factor Analysis with Long-Range Dependence
2016-18 J. Eduardo Vera-Valdés: Essays in Long Memory
2016-19 Magnus Sander: Returns, Dividends, and Optimal Portfolios
2016-20 Ioana Daniela Neamtu: Wind Power Effects and Price Elasticity of Demand for the Nordic Electricity Markets
2016-21 Anne Brink Nandrup: Determinants of Student Achievement and Education Choice
2016-22 Jakob Guldbæk Mikkelsen: Time-Varying Loadings in Factor Models: Theory and Applications
2016-23 Dan Nguyen: Formidability and Human Behavior: An Interdisciplinary Approach

2016-25 Samira Mirzaei: Optimization Algorithms for Multi-Commodity Routing and Inventory Routing Problems

2017-1 Viktoryia Buhayenko: Determining Dynamic Discounts for Supply Chain Coordination

2017-2 Mikkel Bennedsen: Rough Continuous-Time Processes: Theory and Applications

2017-3 Max Weiss Dohrn: Business Cycle Costs for Finitely Living Individuals

2017-4 Oskar Knapik: Essays on Econometric Modelling and Forecasting of Electricity Prices


2017-6 Bo Laursen: Econometric Analysis of Time-Varying Volatility in Financial Markets

2017-7 Federico Carlini: Essays on Fractional Filters and Co-Integration

2017-8 Jonas Juul Henriksen: International Trade and the Labour Market

2017-9 Christian Ellermann-Aarslev: History Dependent Unemployment Insurance

2017-10 Martin Alfaro: Essays on International Trade and Strategic Behavior


2017-12 Morten Holm Jacobsen Fenger: Essays on the Dynamics of Consumer Behavior

2017-13 Carsten P.T. Rosenskjold: Econometric Modelling of Mortality and its Socio-Economic Differences


2018-2 Maria Elbek: Modeling and Optimization of Collection Problems

2018-3 Søren Kærgaard Slipsager: Long-horizon Investing: Pensions and Private Equity

2018-4 Parisa Bagheri Tookanlou: Models for Product Line Design with Customization – Considering Factors in Marketing and Operations

2018-5 Douglas Eduardo Turatti: Monte Carlo Analysis of Time-Varying Parameter Models with Stochastic Volatility