In-Sample Confidence Bands and Out-of-Sample Forecast Bands for Time-Varying Parameters in Observation Driven Models

Francisco Blasques
Siem Jan Koopman
Katarzyna Łasak
André Lucas

Faculty of Economics and Business Administration, VU University Amsterdam, Tinbergen Institute, the Netherlands

In-Sample Confidence Bands and Out-of-Sample Forecast Bands for Time-Varying Parameters in Observation Driven Models[☆]

Francisco Blasques[a], Siem Jan Koopman[a,b], Katarzyna Łasak[a], André Lucas[a]

[a]VU University Amsterdam and Tinbergen Institute
[b]CREATES, Aarhus University

Abstract

We study the performance of alternative methods for calculating in-sample confidence and out-of-sample forecast bands for time-varying parameters. The in-sample bands reflect parameter uncertainty only. The out-of-sample bands reflect both parameter uncertainty and innovation uncertainty. The bands are applicable to a large class of observation driven models and a wide range of estimation procedures. A Monte Carlo study is conducted for time-varying parameter models such as generalized autoregressive conditional heteroskedasticity and autoregressive conditional duration models. Our results show clear differences between the actual coverage provided by the different methods. We illustrate our findings in a volatility analysis for monthly Standard & Poor's 500 index returns.

Keywords: autoregressive conditional duration, delta-method, generalized autoregressive conditional heteroskedasticity, score driven models, time-varying mean.

JEL: C53, C52

1. Introduction

Over the last decades, time-varying parameter models have become increasingly popular in empirical economics and finance. The fast development of new methods for filtering time-varying parameters in dynamic models with nonlinear and non-Gaussian features has made these models more accessible, flexible and attractive. Starting from the 1960s, time-varying parameters for the mean equation in linear Gaussian models were initially typically handled by the Kalman filter and related methods. For a given linear Gaussian state space model, the Kalman filter can be used to calculate conditional means and variances of unobserved time-varying parameters (or linear functions thereof) in a computationally efficient manner; see, for example, Durbin and Koopman [2012] for a detailed treatment. In this modeling framework, the construction of in-sample confidence bands and out-of-sample forecast bands is straightforward and performed on a routine basis as expressions for the conditional variances of the time-varying parameters are explicitly available. In case of nonlinear and/or non-Gaussian extensions of state space models, the computation of confidence bands can be somewhat more involved. An example is the stochastic volatility model for which analysis simulation-based methods are typically used; see the discussions in Shephard [2005].

Since the 1980s, other model classes for time-varying parameters have been developed. Specifically, models for time-varying conditional variances have received much attention in the empirical economics and finance literature. For example, the generalized autoregressive conditional heteroskedasticity (GARCH) model by Engle [1982] and Bollerslev [1986] has led to a range of model formulations for time-varying parameters. In the standard ARCH and GARCH models, the conditional variance is obtained from filtering past observations through a volatility updating equation. The relative simplicity of GARCH models has spurred their widespread use in both the academic and professional world.

In most empirical studies, the estimated volatility from the GARCH model is presented without in-sample bands that reflect parameter uncertainty in the volatility updating equation. Similarly, volatility forecasts may feature bands that reflect innovation uncertainty, but will typically ignore parameter uncertainty. Exact analytical results are not available because the filters are highly nonlinear functions of past observations. As a result, statistical software programs rarely provide in-sample confidence bands or out-of-sample forecast bands for estimates of such time-varying parameters. This argument also applies to other models related to GARCH including the autoregressive conditional duration (ACD) model of Engle and Russell [1998], the multiplicative error model of Engle [2002], the observation driven Poisson count model of Davis, Dunsmuir, and Streett [2003], and the score driven models of Creal et al. [2013]. All these models belong to the class of observation driven models as opposed to the class of parameter driven models. The latter also includes the state space model as briefly discussed above. See Cox [1981] for a detailed description of parameter and observation driven models.

We review and analyze different methods for constructing in-sample and out-of-sample bands. For our in-sample bands, we compare two analytical methods and one simulation-based method. All these bands reflect parameter uncertainty. The approximate analytical bands require simple computations and are not subject to random fluctuations due to simulation error. The analytical bands can be used as long as the updating equation is differentiable and the estimator of the static parameters is asymptotically normally distributed. For the computation of forecast bands, we compare three simulation-based procedures. The first method only takes innovation uncertainty into account. The second and third methods incorporate both parameter and innovation uncertainty. Although these methods require simulations, the forecast bands are relatively fast to compute. In particular, we argue that the necessary computations are more efficient than the bootstrapped forecast bands proposed by for example Pascual et al. [2006] for GARCH mod-

els.

All methods considered can be readily implemented in software packages. We investigate in detail the coverage probabilities of each of these different approaches over a range of different time-varying parameter models. We find that simulation based methods are the most reliable, but that the approximate analytical methods also perform well in many settings.

To provide evidence of how effective the different methods are, we present the results of a Monte Carlo study in which we compute in-sample confidence bands and out-of-sample forecast bands for generated time series from GARCH, score driven, ACD and time-varying mean (local level) models. The results reveal that the actual coverage of our (preferred) analytical bands is close to the nominal coverage level obtained by simulation. The simulation in-sample bands and forecast bands all attain accurate coverage. An empirical application for the GARCH model applied to a time series of monthly log-returns from the Standard & Poor's 500 index reveals the practical importance of these bands. We also show that the choice of method for computing in-sample bands is empirically relevant, and that our analytical bands provide a good approximation to the more computationally intensive simulation based bands.

This paper is organized as follows. Section 2 introduces the class of observation driven models. Section 3 introduces different methods for computing in-sample bands for the time-varying parameter. Section 4 presents different simulation-based methods for the computation of the out-of-sample forecast bands. Section 5 analyzes the relative performance of the bands in a Monte Carlo study. Section 6 presents our empirical findings for the Standard & Poor's 500 monthly returns. Section 7 concludes.

2. Observation driven models

In observation driven models, the time-varying parameter is filtered using an updating equation that depends only on past observations. The main focus in these models is on how to let past realizations of the variable of interest affect the current value of the time-varying parameter. This can be achieved by means of different specifications.

Consider a time series model for an observed time series $y_1, \ldots, y_T$ given by

$$y_t \sim p_y(y_t | f_t; \boldsymbol{\theta}), \qquad t = 1, \ldots, T, \tag{1}$$

where density $p_y(\,\cdot\,)$ is implied by a model equation for y_t and is a function of the time-varying parameter f_t and the static parameter $\boldsymbol{\theta}$, for example, $y_t = f_t + \varepsilon_t$ for a time-varying mean, or $y_t = \mu + f_t^{1/2} \varepsilon_t$ for a fixed mean and time-varying variance, with for instance $\varepsilon_t \sim \mathrm{N}(0,1)$. The time-varying parameter is formally defined as a function $f_t := f_t(y^{1:t-1}, f_1; \boldsymbol{\theta})$ that depends on the past observations $y^{1:t-1} := \{y_1, y_2, \ldots, y_{t-1}\}$, on some initial value f_1, and on a static parameter vector $\boldsymbol{\theta}$. The updating function for the time-varying parameter can be expressed in many different ways. For example, we can consider a linear updating equation consisting of lagged values of y_t and f_t. When only considering single lags, we obtain

$$f_{t+1} = \omega + \beta f_t + \alpha s(y_t, f_t; \boldsymbol{\theta}), \tag{2}$$

with initialization f_1 and where $s(y_t, f_t; \boldsymbol{\theta})$ is some (possibly nonlinear) function of y_t, f_t, and $\boldsymbol{\theta}$. The function $s(\,\cdot\,)$ can be chosen in a flexible way and is often just a transformation of y_t as we will show in the examples below. The coefficients ω, α and β are part of the parameter vector $\boldsymbol{\theta}$. The recursive nature of the formulation implies that f_{t+1} is a (nonlinear) function of $y_t, \ldots, y_1, f_1$ and $\boldsymbol{\theta}$. Hence, the updating equation (2) is consistent

with the definition of f_t, that is $f_t := f_t(y^{1:t-1}, f_1; \boldsymbol{\theta})$. We can also write the updating equation (2) in the general form of a stochastic recurrence equation,

$$f_{t+1} = \phi(y_t, f_t; \boldsymbol{\theta}), \tag{3}$$

with initialization f_1 and recurrence function $\phi(\cdot)$. The choice of the function $s(\cdot)$ in (2) thus defines the type of updating used for f_t.

Example 1. The generalized autoregressive conditional heteroskedasticity model of Engle [1982] and Bollerslev [1986], known as the GARCH model, for a mean-adjusted financial return series $y_1, \ldots, y_T$ is a special case of equations (1) and (2) with

$$y_t = f_t^{1/2} \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0,1), \qquad s(y_t, f_t; \boldsymbol{\theta}) = y_t^2,$$

for $t = 1, \ldots, T$, where $\mathrm{NID}(0,1)$ refers to a standard normally independently distributed sequence. The GARCH-t model of Bollerslev [1986] is obtained when we replace the normal by the Student's t distribution.

Example 2. The autoregressive conditional duration model of Engle and Russell [1998], known as the ACD model, for irregularly spaced data is also a special case of equations (1) and (2) with

$$y_t = f_t \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{Exp}(1), \qquad s(y_t, f_t; \boldsymbol{\theta}) = y_t,$$

for $t = 1, \ldots, T$, where $\mathrm{Exp}(1)$ is the standard Exponential distribution. Different specifications of this modeling framework for durations and intensities are discussed in Grammig and Maurer [2000].

The score driven models of Creal et al. [2013] and Harvey [2013] also belong to the model class represented by equations (1) and (2) where $p_y(y_t | f_t; \boldsymbol{\theta})$ can be any density

6

with f_t as the time-varying parameter. The specific choice of function $s(\cdot)$ distinguishes the score driven model from other models, namely

$$s(y_t, f_t; \boldsymbol{\theta}) = S_t \frac{\partial \log p_y(y_t | f_t, y^{1:t-1}; \boldsymbol{\theta})}{\partial f_t},$$

for some scaling function $S_t := S_t(f_t; \boldsymbol{\theta})$. In many models of practical interest, the predictive density $p_y(y_t | f_t, y^{1:t-1}; \boldsymbol{\theta})$ reduces to the conditional observation density $p_y(y_t | f_t; \boldsymbol{\theta})$ since f_t is also a function of $y^{1:t-1}$. The scaling S_t can be simply set to unity or alternatively be set to reflect the local curvature in the log conditional density function at time t.

When we consider the observation densities of Examples 1 and 2 and apply the score driven framework, we obtain equivalent updating functions if we scale the score by the inverse conditional Fisher information matrix; see Creal et al. [2013]. Indeed the score driven models include many well-known and popular dynamic models. However, in the case of a Student's t distribution for $p_y(y_t | f_t; \boldsymbol{\theta})$ in equation (1), we obtain the model as discussed in Creal et al. [2011, 2013] and Harvey [2013].[1]

Example 3. The univariate Student's t score driven volatility model is given by

$$s(y_t, f_t; \boldsymbol{\theta}) = (1 + 3\lambda^{-1}) \left(\frac{(1 - \lambda^{-1}) y_t^2}{1 + \lambda^{-1} y_t^2 / f_t} - f_t \right),$$

where λ is the degrees of freedom of the Student's t distribution. This expression uses a scaling function based on the inverse Fisher information matrix; see Creal et al. [2011, 2013] for more details.

The parameter vector $\boldsymbol{\theta}$ is unknown and needs to be estimated to obtain estimates of

---

[1]More literature on score driven models with theoretical and empirical developments is provided via the website http://www.gasmodel.com.

the time-varying parameter f_t. The estimation of $\boldsymbol{\theta}$ can be based on the principle of maximum likelihood as we can evaluate the loglikelihood function $\ell(\theta; f_1)$ via the prediction error representation. We have

$$\ell(\theta; f_1) = \sum_{t=1}^{T} \log p_y(y_t | f_t, y^{1:t-1}; \boldsymbol{\theta}),$$

where values for f_t are evaluated via equation (2) with a specific initial value f_1. When it is difficult to find a "natural" value for f_1, it can be included in the parameter vector $\boldsymbol{\theta}$ and estimated simultaneously with the other parameters in $\boldsymbol{\theta}$. We notice that in many, if not most, cases of empirical interest we have $p_y(y_t | f_t, y^{1:t-1}; \boldsymbol{\theta}) = p_y(y_t | f_t; \boldsymbol{\theta})$ such that we can rely simply on the model density when computing the loglikelihood function. Hence maximum likelihood estimation reduces to the basic task of numerically maximizing $\ell(\theta; f_1)$ with respect to $\boldsymbol{\theta}$. This is the standard practice for GARCH and related models.

We write the maximum likelihood estimator of $\boldsymbol{\theta}$ as $\widehat{\boldsymbol{\theta}}_T$. The values of f_t obtained from equation (2) under $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_T$ are denoted by $\widehat{f}_t$ for $t = 1, \ldots, T$. The $\widehat{f}_t$'s can be regarded as a weighted function of f_1 and $y^{1:t-1}$ with the weights determined by $\widehat{\boldsymbol{\theta}}_T$, for $t = 2, \ldots, T$. We can summarize the sources of uncertainty for $\widehat{f}_t$ as follows:

- Parameter uncertainty: we do not know the true parameter vector $\boldsymbol{\theta}$ but we replace it by its maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_T$ to compute $\widehat{f}_t$;

- Filtering uncertainty: we consider the updating equation (2) for f_t but we do not know the underlying time-varying parameter process;

- Model uncertainty: we consider $p_y(\cdot)$ in equation (1) but we do not know the true data generation process for y_t.

In this paper we concentrate on developing bands that reflect the parameter uncer-

tainty for $\widehat{f}_t$, $t = 1, \ldots, T$ while taking the other two sources of uncertainty as given. This is an important step forward in the current literature, where in-sample uncertainty bands around the time-varying parameter are almost never shown. The next section presents several ways to construct the approximate bands. Section 5 investigates the coverage properties of these bands by conducting a Monte Carlo study.

3. In-Sample Bands

Conditional on past information and the initial value f_1, and based on the maximum likelihood estimate $\widehat{\theta}_T$ of θ, we develop confidence bands for $\widehat{f}_t$ that reflect parameter uncertainty about the true value of θ. The bands are based as usual on the variance of $\widehat{f}_t$, which we denote as $V_t := \mathbb{V}\mathrm{ar}(\widehat{f}_t)$. The bands thus reflect the randomness of the estimator $\widehat{\theta}_T$. We propose three different methods to obtain approximations for these bands:

1. simple non-cumulative bands;
2. cumulative delta-method bands;
3. simulation based bands.

The first two methods are analytic and very fast. They approximate the variance of the filter V_t by linearizing the updating function $\phi(\,\cdot\,)$ in (3) or (2) and using the asymptotic variance of $\widehat{\theta}_T$. These analytic methods have the advantage of easy implementation in software packages as they are based on simple calculations. The third method provides probably the most accurate reflection of parameter uncertainty in $\widehat{f}_t$ but comes with the disadvantage that the computations are subject to randomness by construction. This method also uses the asymptotic variance of $\widehat{\theta}_T$, but does not require the linearization of the updating function $\phi(\,\cdot\,)$. Hence this method may be preferable when nonlinearities play an important role in the filter.

The simple non-cumulative bands do not take into account the *accumulation* of parameter uncertainty in the updating or filtering process for $\widehat{f}_t$. They rely on simple expressions, but are also somewhat naive and therefore only deliver accurate results for a restrictive class of observation driven models. Specifically, these bands only become interesting when the autoregressive component is either not present, that is when $\beta = 0$ in equation (2), or when it is sufficiently small. For example, the method is appropriate for the ARCH model of Engle [1982], which is the model of Example 1 with $\beta = 0$.

The *cumulative delta-method bands* take the accumulation of parameter uncertainty in the updating process into account. These bands are obtained by a simple application of the delta-method and are relevant for all observation driven models, including GARCH, ACD, and score driven models where the updating equation for $\widehat{f}_t$ features an autoregressive term, that is $\beta \neq 0$.

The simulation bands deal not only with the accumulation of the uncertainty due to the parameter vector $\boldsymbol{\theta}$, but also treat the effects of possible nonlinear functional relations between $\boldsymbol{\theta}$ and $\widehat{f}_t$. These bands are most appropriate for cases where we have strong nonlinear expressions in the updating process and the linearization error produced by the delta-method cannot be ignored. In contrast to the first two methods for computing bands, this method is not fully analytic and requires repeated simulations of the estimated time-varying parameter paths $\widehat{f}_t$.

All methods for computing bands rely on the asymptotic normality of the estimator $\widehat{\boldsymbol{\theta}}_T$. The bands can therefore be computed for any estimator $\widehat{\boldsymbol{\theta}}_T$ as long as it has a normal distribution asymptotically. The Monte Carlo study in Section 5 investigates the extent to which this reliance on the asymptotic distribution of the estimator may be problematic in small samples.

3.1. Simple In-Sample Non-Cumulative Bands

Our simple non-cumulative bands can generally be obtained by linearizing the function $\phi(y_t, \widehat{f}_t; \widehat{\boldsymbol{\theta}}_T)$, as defined in equation (3) with $f_t = \widehat{f}_t$ and $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_T$, around the true parameter vector used for generating $y_1, \ldots, y_T$, that is $\boldsymbol{\theta}_0 \in \Theta$ where Θ is the parameter space. We use $\widehat{f}_t$ to denote the filter that is a function of the point estimate $\widehat{\boldsymbol{\theta}}_T$. In the case of the non-cumulative bands, we take $\widehat{f}_t$ as given in calculating the approximate variance of $\widehat{f}_{t+1}$. Our proposed simple bands then rely on the approximation

$$\widehat{f}_{t+1}(y^{1:t}, \widehat{f}_1; \widehat{\boldsymbol{\theta}}_T) \approx \phi(y_t, \widehat{f}_t; \boldsymbol{\theta}_0) + \sum_{i=1}^{q} \frac{\partial \phi(y_t, \widehat{f}_t; \boldsymbol{\theta}_0)}{\partial \theta_i}(\widehat{\theta}_{T,i} - \theta_{0,i}), \tag{4}$$

for any $\widehat{f}_t$ and around some point $\boldsymbol{\theta}$, where $\theta_{0,i}$ is the ith element of the $q \times 1$ true parameter vector $\boldsymbol{\theta}_0$ and, similarly, $\widehat{\theta}_{T,i}$ is the ith element of the $q \times 1$ maximum likelihood estimate vector $\widehat{\boldsymbol{\theta}}_T$, for $i = 1, \ldots, q$.

We define the partial derivative of $\phi(y_t, \widehat{f}_t; \boldsymbol{\theta})$ with respect to θ_i, the ith element of $\boldsymbol{\theta}$, as the function

$$\nabla_{i,t} = \nabla_i(y_t, \widehat{f}_t; \boldsymbol{\theta}) := \frac{\partial \phi(y_t, \widehat{f}_t; \boldsymbol{\theta})}{\partial \theta_i}, \qquad i = 1, \ldots, q. \tag{5}$$

Hence we can write the approximation in (4) as

$$\widehat{f}_{t+1} \approx \phi(y_t, \widehat{f}_t; \boldsymbol{\theta}_0) + \sum_{i=1}^{q} \nabla_{i,t} \times (\widehat{\theta}_{T,i} - \theta_{0,i}),$$

for $t = 1, \ldots, T$. The approximate variance of $\widehat{f}_{t+1}$ for given $\widehat{f}_t$ is then obtained by

$$V_{t+1} = \mathbb{V}\mathrm{ar}(\widehat{f}_{t+1}) \approx \sum_{i=1}^{q} \nabla_{i,t}^2 \mathbb{V}\mathrm{ar}(\widehat{\theta}_{T,i}) + 2 \sum_{1 \leq i < j \leq q} \nabla_{i,t} \nabla_{j,t} \mathbb{C}\mathrm{ov}(\widehat{\theta}_{T,i}, \widehat{\theta}_{T,j}), \tag{6}$$

where $\mathbb{V}\mathrm{ar}$ and $\mathbb{C}\mathrm{ov}$ are the variance and covariance operators, respectively. In case the

derivative $\nabla_i(y_t, f; \boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ and/or f, we evaluate the derivative at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_T$ and $f = \widehat{f}_t$.

Let the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_T$ be asymptotically normally distributed as

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathbf{W}), \tag{7}$$

where $\boldsymbol{\theta}_0$ is the true parameter vector and $N(0, \mathbf{W})$ is the multivariate normal distribution with zero mean vector and covariance matrix $\mathbf{W}$. The standard notation $\xrightarrow{d}$ is for *convergence in distribution*. We typically take $\mathbf{W}$ as the robust sandwich covariance matrix estimator and evaluate it at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_T$. We then write explicitly $\mathbf{W} = \widehat{\mathbf{W}}$. It further follows that

$$\mathbb{V}\mathrm{ar}\left[\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right)\right] \approx \widehat{\mathbf{W}} \quad \Leftrightarrow \quad \mathbb{V}\mathrm{ar}(\widehat{\boldsymbol{\theta}}_T) \approx T^{-1}\widehat{\mathbf{W}}.$$

From the asymptotic properties of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_T$, we can deduct that the asymptotic variance of $\widehat{f}_{t+1}$ converges asymptotically to V_{t+1} where

$$V_{t+1} \approx \sum_{i=1}^{q} T^{-1}w_{i,i}\nabla_{i,t}^2 + 2\sum_{1 \leq i < j \leq q} T^{-1}w_{i,j}\nabla_{i,t}\nabla_{j,t}, \tag{8}$$

and with $w_{i,j}$ as the (i, j) element of matrix $\mathbf{W}$. In a standard fashion and based on the asymptotic normal distribution for $\widehat{f}_{t+1}$, we obtain asymptotic 95% in-sample confidence bands for $\widehat{f}_{t+1}$ as

$$\left[\widehat{f}_{t+1} - 1.96\sqrt{V_{t+1}} \quad , \quad \widehat{f}_{t+1} + 1.96\sqrt{V_{t+1}}\right]. \tag{9}$$

Example 4. (Cont'd from Example 1) Consider the ARCH model of Engle [1982] which is defined as the GARCH model of Example 1 with $\beta = 0$ in equation (2) and with $\boldsymbol{\theta} = (\omega, \alpha)'$. The resulting updating function for $\widehat{f}_{t+1}$ is given by $\widehat{f}_{t+1} = \widehat{\omega}_T + \widehat{\alpha}_T y_t^2$ where $\widehat{\omega}_T$ and $\widehat{\alpha}_T$ are respectively the first and second elements of the parameter vector $\widehat{\boldsymbol{\theta}}_T$. It

follows immediately that

$$\nabla_{1,t} = 1, \qquad \nabla_{2,t} = y_t^2, \qquad t = 1, \dots, T.$$

The asymptotically correct bands obtained from our simple non-cumulative approach are then given by (9) with

$$
\begin{aligned}
V_{t+1} &= \mathbb{V}\mathrm{ar}(\widehat{\omega}_T) + y_t^4 \mathbb{V}\mathrm{ar}(\widehat{\alpha}_T) + 2y_t^2 \mathbb{C}\mathrm{ov}(\widehat{\omega}_T, \widehat{\alpha}_T) \\
&\approx T^{-1}(w_{1,1} + y_t^4 w_{2,2} + 2y_t^2 w_{1,2}).
\end{aligned}
$$

These expressions are very simple to compute numerically once $\mathbf{W}$ has been replaced by its estimate $\widehat{\mathbf{W}}$.

In Section 5 we provide evidence that our simple non-cumulative bands are not appropriate for observation driven models where the filter has strong autoregressive dynamics. This applies to the GARCH model with $\beta \neq 0$ and in particular to the case where β is far away from zero.

3.2. In-Sample Cumulative Delta-Method Bands

The previous approximations did not account for the fact that also $\widehat{f}_t$ itself depends on the estimator $\widehat{\boldsymbol{\theta}}_T$. This is an important defect. In filters with an autoregressive component, the parameter uncertainty accumulates as the filter evolves over time. This occurs because $\widehat{f}_{t+1}$ depends on $\widehat{f}_t$, which already contains parameter uncertainty. This accumulation of parameter uncertainty in the filtering process can be analytically tracked by application of the delta-method. It is based on the same linearization (4) of $\phi(\cdot)$, but where we also explicitly account for the dependence of $\widehat{f}_t$ on $\widehat{\boldsymbol{\theta}}_T$. We redefine $\nabla_{i,t}$ as

$$\nabla_{i,t} = \nabla_i(y_t, f_t; \boldsymbol{\theta}) := \frac{\partial \phi(y_t, f_t; \boldsymbol{\theta})}{\partial f_t} \frac{\partial f_t}{\partial \theta_i} + \frac{\partial \phi(y_t, f_t; \boldsymbol{\theta})}{\partial \theta_i}. \tag{10}$$

13

We assume that the point estimate $\widehat{\theta}_T$ is consistent and sufficiently close the true parameter θ_0. Hence when calculating $\nabla_{i,t}$ we take f_t as the value of the filter at time t for the point estimate of $\widehat{\theta}_T$, that is $\widehat{f}_t$. The derivatives $\partial f_t / \partial \theta_i$ can then be computed recursively. For example, for the GARCH model we have

$$\frac{\partial f_{t+1}}{\partial \theta_i} = \left(\frac{\partial \omega}{\partial \theta_i} + \frac{\partial \alpha}{\partial \theta_i} y_t^2 + \frac{\partial \beta}{\partial \theta_i} f_t \right) + \beta \frac{\partial f_t}{\partial \theta_i}, \tag{11}$$

which can be computed in parallel to the recursion for f_t itself.

The variance V_{t+1} of the filtered estimate $\widehat{f}_t$ is again approximated by (8), but with $\nabla_{i,t}$ defined as in equation (10). Also the asymptotic 95% in-sample confidence bands for $\widehat{f}_{t+1}$ can be based on equation (9), but with V_{t+1} computed as indicated above.

The delta-method bands are accurate when: (i) the the updating equation is sufficiently well approximated by a linear function; and (ii) when the sample is sufficiently large for the distribution of the estimator to be approximately normal. These two conditions are met in many situations of empirical interest.

The following example illustrates the application of the current theory to the case of the GARCH model.

Example 5. (Cont'd from Example 1) Consider the GARCH model of Bollerslev [1986] as given by equation (2) and with $\theta = (\omega, \alpha, \beta)'$. We have $y_t = \sqrt{f_t} u_t$ with $u_t \sim NID(0,1)$ and $f_{t+1} = \omega + \alpha y_t^2 + \beta f_t$ for $t = 1, \ldots, T$. Applying equation (10) to the GARCH updating function, it follows that

$$\nabla_{1,t} = 1 + \beta \frac{\partial f_t}{\partial \omega}, \quad \nabla_{2,t} = y_t^2 + \beta \frac{\partial f_t}{\partial \alpha}, \quad \nabla_{3,t} = f_t + \beta \frac{\partial f_t}{\partial \beta}, \qquad t = 1, \ldots, T,$$

where the partial derivatives on the right-hand side are defined via the recursion (11). We evaluate f_t and f_{t-1} by $\widehat{f}_t$ and $\widehat{f}_{t-1}$, respectively, and replace the elements of θ by the

14

corresponding elements of $\widehat{\boldsymbol{\theta}}_T$. The resulting approximation of the variance of $\widehat{f}_{t+1}$ is then given by

$$
\begin{aligned}
V_{t+1} \approx{} & \left(1 + \widehat{\beta}_T \frac{\partial f_t}{\partial \omega}\right)^2 w_{1,1} + \left(y_t^2 + \widehat{\beta}_T \frac{\partial f_t}{\partial \alpha}\right)^2 w_{2,2} + \left(\widehat{f}_t + \widehat{\beta}_T \frac{\partial f_t}{\partial \beta}\right)^2 w_{3,3} \\
& + 2\left(1 + \widehat{\beta}_T \frac{\partial f_t}{\partial \omega}\right)\left(y_t^2 + \widehat{\beta}_T \frac{\partial f_t}{\partial \alpha}\right) w_{1,2} \\
& + 2\left(1 + \widehat{\beta}_T \frac{\partial f_t}{\partial \omega}\right)\left(\widehat{f}_t + \widehat{\beta}_T \frac{\partial f_t}{\partial \beta}\right) w_{1,3} \\
& + 2\left(y_t^2 + \widehat{\beta}_T \frac{\partial f_t}{\partial \alpha}\right)\left(\widehat{f}_t + \widehat{\beta}_T \frac{\partial f_t}{\partial \beta}\right) w_{2,3},
\end{aligned}
$$

where $w_{i,j}$ denotes element (i,j) of the asymptotic variance matrix $\mathbf{W}$ of $\widehat{\boldsymbol{\theta}}_T$. The necessary computations are more involved when compared to those required for the simple non-cumulative method used in Example 4. The bands, however, are still straightforward to implement and fast to compute.

If the updating equation is highly nonlinear, the error in the above linearization may become substantial, resulting in inaccurate approximations for V_{t+1}. The in-sample bands obtained from using the simulation method below avoid the linearization step altogether. The resulting bands are only subject to the error in approximating the distribution of the estimator by its asymptotic distribution.

3.3. In-Sample Simulation-Based Bands

An alternative approach to obtain in-sample confidence bands for $\widehat{f}_{t+1}$ is using simulation methods. Since the bands only reflect the uncertainty in $\widehat{\boldsymbol{\theta}}_T$, one can obtain exact confidence bands by drawing parameter values $\boldsymbol{\theta}^i$ from the distribution of $\widehat{\boldsymbol{\theta}}_T$ and running the filter from $t = 1$ to $t = T$ for every simulated $\boldsymbol{\theta}^i$. However, since the finite-sample distribution of $\widehat{\boldsymbol{\theta}}_T$ is unknown, we draw parameter values from the approximat-

ing asymptotic distribution instead,

$$\widehat{\boldsymbol{\theta}}_T^i \sim N\left(\widehat{\boldsymbol{\theta}}_T, T^{-1}\widehat{\mathbf{W}}\right), \qquad i = 1, \ldots, M, \tag{12}$$

for some predefined number M. For each $\boldsymbol{\theta}^i$, $i = 1, \ldots, M$, the sequence $f_1^i, \widehat{f}_2^i, \ldots, \widehat{f}_T^i$ can then be determined using equations (2) or (3) with $f_1^i = f_1$. In particular, we consider the filtering recursion for each i, that is

$$\widehat{f}_{t+1}^i = \phi(y_t, \widehat{f}_t^i; \widehat{\boldsymbol{\theta}}_T^i), \qquad t = 1, \ldots, T.$$

The simulation method is rather different from the two earlier methods. Instead of linearizing the filtering recursion and working with an approximate Gaussian distribution for $\widehat{f}_t$, we can make use of simulations to obtain more accurate bands. The uncertainty of $\widehat{\boldsymbol{\theta}}_T$ is characterized by the asymptotic distribution (7). Also note that in the simulation method the bands need not be based on V_{t+1}, but can rather be obtained directly by calculating the appropriate percentiles for each t over the M draws of the filtered paths $\widehat{f}_t^i$ for $i = 1, \ldots, M$.

When nonlinearities play a prominent role in the updating equation $\phi(\cdot)$, the simulation bands may become more accurate than those based on the linearization methods. However, simulations can be time consuming when the sample size T is large and when a high level of accuracy (large M) is required. Also simulation methods are inevitably subject to simulation error. Hence the simulation bands may be less attractive for its implementation in software packages.

Example 6. (Cont'd from Example 1) Consider the GARCH model with $y_t \sim NID(0, f_t)$ and $f_{t+1} = \omega + \alpha y_t^2 + \beta f_t$, for initial value f_1. The simulation bands can be simply obtained by drawing M parameter vectors from the approximate density of the estimator

as in (12), we obtain $\widehat{\boldsymbol{\theta}}_T^i = (\widehat{\omega}_T^i, \tilde{\alpha}_T^i, \tilde{\beta}_T^i)$ and

$$\widehat{f}_{t+1}^i = \tilde{\omega}_T^i + \tilde{\alpha}_T^i y_t^2 + \tilde{\beta}_T^i \widehat{f}_t^i, \qquad \text{for } i = 1, \ldots, M.$$

From these simulated paths we compute in-sample bands for any confidence level.

It is important to note the use of an approximate asymptotic distribution may result in drawing values $\widehat{\boldsymbol{\theta}}_T^i$ that lie outside the admissable parameter space Θ that is of interest. For example, for GARCH models it is common practice to impose restrictions on the vector (ω, α, β) that ensure positivity of the filtered conditional variance. Avoiding draws from outside the parameter space of interest can be achieved by re-parameterizing the model in such a way as to ensure that Θ is unbounded for the transformed parameters. For example, by estimating $\omega^* = \log(\omega)$ we can easily ensure that $\omega = \exp(\omega^*)$ is strictly positive.

4. Out-of-Sample Forecast Bands

In observation driven models, the time-varying parameter f_{T+1} depends by construction on the observed sample of data $y_1, \ldots, y_T$. Therefore, innovation uncertainty is not a concern for the construction of a forecast band for f_{T+1}. In effect, the in-sample bands discussed in the previous section are applicable to the parameter sequence $f_1, \ldots, f_{T+1}$. However, when considering bands for f_{T+n}, with $n \geq 2$, it is crucial to take into account the uncertainty of future innovations. In the context of observation driven models, future innovations play a crucial role as they determine the future values of the unknown time-varying parameter via the realizations of y_{T+n-1}.

We describe three different methods for the construction of out-of-sample forecast bands. The first method only considers innovation uncertainty, while the second and third method incorporate both parameter and innovation uncertainty.

4.1. Simple Forecast Bands

A basic approach to the construction of a forecast band for f_{T+n}, for some $n \geq 2$, is to first take the point estimate $\widehat{\boldsymbol{\theta}}_T$ and the resulting filtered value $\widehat{f}_{T+1}$ as given, and then to extrapolate the filter $f_{T+1}, \ldots, f_{T+n}$ multiple times by drawing multiple innovation paths from the estimated innovation density. Bands are constructed from the multiple filtered paths by computing the desired percentiles. This is equivalent to drawing multiple paths $y_{T+1}, \ldots, y_{T+n-1}$ from the conditional density $p(y_t|f_t)$ for $t = T+1, \ldots, T+n-1$, and using these values to obtain multiple paths $f_{T+1}, \ldots, f_{T+n}$. The crucial simplifying aspect of these bands is that they take $\widehat{\boldsymbol{\theta}}_T$ and the resulting $\widehat{f}_{T+1}$ as fixed. Hence the method ignores parameter uncertainty. While this method only captures innovation uncertainty, the resulting forecast bands are interesting in their own right. The basic algorithm used to obtain the forecast bands can be summarized as follows.

1. Given $\widehat{\boldsymbol{\theta}}_T$ and the filtered value $\widehat{f}_{T+1}$, draw S values $y_{T+1}^1, \ldots, y_{T+1}^S$ from the estimated conditional density for time $T+1$

$$y_{T+1}^s \sim p_y(y_{T+1}|\widehat{f}_{T+1}; g\boldsymbol{\theta}_T), \qquad s = 1, \ldots, S.$$

2. Use $y_{T+1}^1, \ldots, y_{T+1}^S$ and the updating equation (3) to obtain $\widehat{f}_{T+2}^1, \ldots, \widehat{f}_{T+2}^S$ conditional on $\widehat{\boldsymbol{\theta}}_T$ and $\widehat{f}_{T+1}$,

$$\widehat{f}_{T+2}^s = \phi(y_{T+1}^s, \widehat{f}_{T+1}; \widehat{\boldsymbol{\theta}}_T), \quad s = 1, \ldots, S.$$

3. For each $\widehat{f}_{T+2}^s$, $s = 1, \ldots, S$, redo steps 1 and 2, for periods $T+2, \ldots, T+n$.

4. Use the S values of $\widehat{f}_{T+n}^s$ to calculate the forecast bands at the desired percentiles.

The forecast band method described above is simple to implement. However, its does not take the randomness of either the estimator $\widehat{\boldsymbol{\theta}}_T$ or the filtered value $\widehat{f}_{T+1}$ into account.

18

4.2. Delta-Method Forecast Bands

To incorporate parameter uncertainty into the forecast bands, we can make use of the asymptotic distribution of the estimator of the parameter. In particular, we follow the same approach as taken in Section 3 where we draw M parameter values $\widehat{\boldsymbol{\theta}}_T^i$, for $i = 1, \ldots, M$, from the estimated approximate distribution (12) and use each $\widehat{\boldsymbol{\theta}}_T^i$ to extrapolate the filter into the future. By drawing parameter values from an approximate distribution, we are able to take into account parameter uncertainty. This method is considerably more efficient than the bootstrap procedure proposed by Pascual et al. [2006] for GARCH models, which requires the re-estimation of the parameter $\widehat{\boldsymbol{\theta}}_T^i$ from M sequences of bootstrapped innovations. To achieve further computational efficiency, we build on the delta-method proposed in Section 3.2 to draw values for the filtered value at time $T + 1$. In this way we obtain approximate values $\widehat{f}_{T+1}^i$ directly for $i = 1, \ldots, M$, instead of applying the filter M times for the full sample to obtain the $\widehat{f}_{T+1}^i$ implied by a given $\widehat{\boldsymbol{\theta}}_T^i$, $i = 1, \ldots, M$.

The key feature of this method of computing forecast bands is its ability to incorporate parameter uncertainty in an efficient way by drawing from approximate distributions for $\widehat{\boldsymbol{\theta}}_T$ and $\widehat{f}_{T+1}$ instead of having to re-estimate and re-filter these. The algorithm used to obtain these bands is summarized as follows:

1. Draw M vectors $(\widehat{\boldsymbol{\theta}}_T^i, \widehat{f}_{T+1}^i)$, $i = 1, \ldots, M$ from the approximate distribution

$$(\widehat{\boldsymbol{\theta}}_T^i, \widehat{f}_{T+1}^i) \sim N\left((\widehat{\boldsymbol{\theta}}_T, \widehat{f}_{T+1}), T^{-1}\widehat{\Sigma}\right), \qquad i = 1, \ldots, M, \qquad (13)$$

where the matrix $\widehat{\Sigma}$ consists of the estimated asymptotic variance of $\widehat{\boldsymbol{\theta}}_T$ and the approximate variance of $\widehat{f}_{T+1}$ derived in Section 3.2 using the delta-method, i.e., $\widehat{\Sigma} = (\mathrm{I}_q, \nabla_t)'\widehat{W}(\mathrm{I}_q, \nabla_t)$.

2. For each pair $(\widehat{\boldsymbol{\theta}}_T^i, \widehat{f}_{T+1}^i)$, $i = 1, \ldots, M$, follow steps 1–3 of the simple forecast bands

algorithm given in Section 4.1 to obtain $M \times S$ values of $\widehat{f}_{T+n}^{i,s}$.

3. Use the $M \times S$ values of $\widehat{f}_{T+n}^{i,s}$ to calculate the bands at the desired percentiles.[2]

As we have done in Section 3.2, it is important to emphasize that the use of an approximate asymptotic distribution may result in drawing values $\widehat{\boldsymbol{\theta}}_T^i$ that lie outside a parameter space Θ that is of interest. In our Monte Carlo exercise of Section 5, we found that a re-parameterization of the GARCH model was important to ensure the positivity of the conditional variance in our forecast bands.

4.3. Multiple Filtering Forecast Bands

In cases where the delta-method does not provide a good approximation to the distribution of f_{T+1}, the delta-method forecast bands may perform poorly as the draws $\widehat{f}_{T+1}^i$, for $i = 1, \ldots, M$, may not reflect accurately the underlying parameter uncertainty. This problem is more likely to occur in observation-driven models with strongly nonlinear parameter updates as the linearization error incurred by the delta-method becomes more relevant. In these cases, it may be preferable to re-filter the entire path $\{\widehat{f}_t^i\}_{t=1}^{T+1}$ for each $\widehat{\boldsymbol{\theta}}_T^i$, for $i = 1, \ldots, M$. This should provide the pairs $(\widehat{\boldsymbol{\theta}}_T^i, \widehat{f}_{T+1}^i)$, for $i = 1, \ldots, M$, that are required for iterating the filter forwards and obtaining the bands that reflect both parameter uncertainty and innovation uncertainty.

When compared to the previous method, the key feature of these forecast bands is their ability to obtain values for $\widehat{f}_{T+1}^i$ that reflect more accurately the parameter uncertainty about the static parameters. This accuracy comes at an additional computation cost because the filtered path has to be calculated multiple times. However, this method is still more efficient than the bootstrap procedure proposed by Pascual et al. [2006] for

---

[2]Note that one can also choose to integrate the in-sample and out-of-sample simulations in the above algorithm by setting $S = 1$ and simulating one out-of-sample path for every in-sample simulation of the static parameters.

GARCH models, as it does not require the re-estimation of $\widehat{\boldsymbol{\theta}}_T$. This alternative method for obtaining forecast bands can be summarized as follows:

1. Draw M parameter values $\widehat{\boldsymbol{\theta}}_T^i$ from its approximate distribution

$$\widehat{\boldsymbol{\theta}}_T^i \sim N\left(\widehat{\boldsymbol{\theta}}_T, T^{-1}\widehat{\mathbf{W}}\right), \qquad i = 1, \ldots, M, \tag{14}$$

2. For each $\widehat{\boldsymbol{\theta}}_T^i$, use the observed data $y_1, \ldots, y_T$ and the updating equation (3) and run the observation-driven filter for $t = 1, \ldots, T$ to obtain $\widehat{f}_{T+1}^i$, $i = 1, \ldots, M$.

3. For each pair $(\widehat{\boldsymbol{\theta}}_T^i, \widehat{f}_{T+1}^i)$, $i = 1, \ldots, M$, follow steps 1–3 of the simple forecast bands algorithm given in Section 4.1 to obtain $M \times S$ values of $\widehat{f}_{T+n}^{i,s}$.

4. Use the $M \times S$ values of $\widehat{f}_{T+n}^{i,s}$ to calculate the forecast bands at the desired percentiles.

5. Monte Carlo Study

5.1. In-Sample Confidence Bands

To verify the performances of the three methods for computing the in-sample bands of time-varying parameters in observation driven models, we carry out a Monte Carlo study. We consider five different models: (i) the GARCH model based on normally distributed innovations; (ii) GARCH based on Student's $t(5)$ innovations; (iii) the Student's $t(5)$ based score driven volatility model of Creal et al. [2011, 2013] and Harvey [2013]; (iv) the autoregressive conditional duration (ACD) model; and (v) a time-varying mean (or local level) model with Gaussian innovations. The GARCH model is introduced in Example 1. In case the density $p_y(y_t|f_t;\boldsymbol{\theta})$ is $NID(0, f_t)$, we obtain Model (i); in case it is Student's $t(5)$ with variance f_t and degrees of freedom λ, we obtain Model (ii). Model (iii) has $s(y_t, f_t;\boldsymbol{\theta}) = (1 + 3\lambda^{-1})((1 + \lambda)y_t^2/(\lambda - 2 + y_t^2/f_t)$ with $\lambda > 2$ degrees of freedom.

Table 1: Models in Monte Carlo study

We present the details of the five observation driven models in our Monte Carlo study. In all cases the updating equation for the time-varying parameter f_t is given by (2), that is $f_{t+1} = \omega + \alpha s_t + \beta f_t$, with $\omega = 0.05$, $\alpha = 0.1$ and for three different values of $\beta = 0.2$, 0.5, 0.8. The remaining details for $p_y(y_t|f_t; \boldsymbol{\theta})$ and $s_t = s(y_t, f_t; \boldsymbol{\theta})$ are given below with a reference to the examples discussed in Section 2. Student's $t(0, f_t, \lambda)$ refers to the Student's t density function with zero mean, variance (rather than scale parameter) f_t and degrees of freedom λ with $\lambda = 5$, $Exp(f_t)$ refers to the Exponential density with intensity parameter f_t, and for model (v) we have set $\sigma^2 = 1$. In all cases, we take $f_1 = 1$.

	Model	$p_y(y_t|f_t; \boldsymbol{\theta})$	$s(y_t, f_t; \boldsymbol{\theta})$	Example
(i)	GARCH	$NID(0, f_t)$	y_t^2	1
(ii)	GARCH-t	iid Student's $t(0, f_t, \lambda)$	y_t^2	1
(iii)	GAS-t	iid Student's $t(0, f_t, \lambda)$	$\frac{(1+3\lambda^{-1})(1+\lambda)y_t^2}{\lambda-2+y_t^2/f_t}$	3
(iv)	ACD	iid $Exp(f_t)$	y_t	2
(v)	Local Level (LL)	$NID(f_t, \sigma^2)$	y_t	–

Model (iv) is introduced in Example 2. The local level (LL) model (v) is a basic application of the score driven model of Example 3. We set $p_y(y_t|f_t; \boldsymbol{\theta})$ to the Gaussian density $NID(f_t, \sigma^2)$ and choose $S_t(f_t; \boldsymbol{\theta}) = 1$ and $s(y_t, f_t; \boldsymbol{\theta}) = y_t$. For models (i) and (iv), the parameter vector is given by $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$, for models (ii) and (iii) it is $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda)'$ and for model (v) it is $\boldsymbol{\theta} = (\omega, \alpha, \beta, \sigma^2)'$. Further details are given in Table 1. We choose ω and f_1 such that the time-varying parameter process is started at its unconditional mean. We consider two different sample sizes, $T = 500$, 1000, and we generate the time series for three values of $\beta = 0.2, 0.5, 0.8$, with $\omega = 0.05$, $\alpha = 0.1$, $\lambda = 5$, $\sigma^2 = 1$ and $f_1 = 1$.

The aim of this Monte Carlo study is to verify how well the three methods studied in Section 3 approximate the correct in-sample bands for $\widehat{f}_{t+1}$. The in-sample bands can be determined in a Monte Carlo setting by repeating the following steps:

1. Generate a time series $y_1, \ldots, y_T$ from one of the models listed in Table 1 for some T.

2. For this generated time series, estimate $\boldsymbol{\theta}$ using the method of maximum likelihood

as discussed in Section 2.

3. Calculate the time-varying parameter sequence $\widehat{f}_{t+1}$ as well as its derivative with respect to $\boldsymbol{\theta}$ using (2) and (11) for a given f_1, based on the generated time series and the estimate $\widehat{\boldsymbol{\theta}}_T$.

These three steps can be repeated N times so that we obtain a set of N paths for $\widehat{f}_t$. From this set we can empirically determine the coverage of the in-sample bands for $\widehat{f}_t$ for nominal confidence levels such as 90%, 95% and 99%.

Within this Monte Carlo framework, for each generated time series, we determine the in-sample bands by the three methods as described in Section 3. We compute the time series average of the number of times the true f_t lies outside the in-sample band. We subsequently take the average of this number over the N replications, and present the results in Table 2.

As expected, the simple non-cumulative bands are not very accurate for the GARCH model with $\beta = 0.8$, while they are somewhat more accurate for the model with $\beta = 0.2$. However, the simple bands provide poor overall results for the three nominal coverage levels of 90%, 95% and 99%. As we have indicated in Example 4, the simple method is appropriate for the ARCH model (that is the GARCH model with $\beta = 0.0$). Our Monte Carlo study reveals that the simple method produces rather inaccurate results even when $\beta = 0.2$. These findings are confirmed by the other models and for both sample sizes considered.

The cumulative delta-method bands are considerably more accurate. The bands are slightly too small, resulting in lower estimated coverage compared to the nominal levels. The differences, however, appear to be tolerable for most practical purposes. The performance for different values of β remains stable, while the performance of the simple non-cumulative bands deteriorates rapidly if the value for β is increased. The estimated

23

Table 2: In-sample bands coverages

For our Monte Carlo study, we present the 90%, 95% and 99% coverages of the in-sample bands, as computed by the three methods "non-cumulative", "cumulative delta" and "simulation", for the observation driven models as described in Table 1. In case of the simulation method, we use $M = 1000$ simulations. The Monte Carlo study is based on $N = 1000$ replications.

T	β	GARCH 90	GARCH 95	GARCH 99	t-GARCH 90	t-GARCH 95	t-GARCH 99	t-GAS 90	t-GAS 95	t-GAS 99	ACD 90	ACD 95	ACD 99	LL 90	LL 95	LL 99
Non-cumulative bands																
500	0.2	76.3	81.9	88.8	74.7	80.1	86.6	74.7	80.2	87.2	76.7	82.1	88.8	79.4	85.0	91.4
500	0.5	67.1	73.4	82.1	62.6	69.0	77.8	63.7	69.8	78.4	68.7	74.7	82.8	69.2	75.6	84.4
500	0.8	46.0	51.8	61.6	40.1	45.4	54.6	43.8	49.4	58.5	50.9	56.8	66.2	46.1	52.0	61.5
1000	0.2	78.9	84.3	91.0	77.7	83.6	89.9	75.8	82.0	89.3	78.1	83.7	90.3	81.1	86.6	92.8
1000	0.5	66.7	73.4	82.7	63.3	69.7	79.2	63.6	69.9	79.1	70.1	76.2	84.5	69.3	76.0	85.1
1000	0.8	46.0	51.9	61.8	38.7	44.0	53.2	43.1	48.6	57.6	51.3	57.4	67.0	46.6	52.6	62.2
Cumulative bands																
500	0.2	86.3	91.8	97.2	87.1	91.7	96.1	87.8	92.1	96.7	86.7	92.3	97.4	86.2	91.9	97.3
500	0.5	85.4	90.7	96.2	86.9	91.6	96.0	89.3	93.1	96.8	86.1	91.6	96.8	86.5	92.0	97.4
500	0.8	86.4	91.5	96.6	86.2	91.1	95.8	91.6	94.5	97.4	85.7	91.3	96.8	88.6	94.3	98.8
1000	0.2	86.1	91.9	97.2	88.2	93.1	97.5	89.6	94.2	98.0	86.1	91.6	97.2	85.9	91.4	96.8
1000	0.5	85.8	91.6	96.9	86.7	91.8	96.6	90.5	94.3	97.7	86.5	91.9	96.9	85.1	91.0	96.8
1000	0.8	87.7	92.9	97.5	87.4	92.8	97.2	93.3	96.0	98.4	86.5	92.0	97.3	88.5	93.7	98.3
Simulation bands																
500	0.2	92.1	96.0	99.0	91.8	95.3	98.3	92.7	96.6	99.5	92.6	96.5	99.3	90.8	95.4	98.8
500	0.5	92.2	96.0	99.0	93.1	96.3	98.6	94.5	97.8	99.6	92.7	96.3	99.0	91.0	95.6	99.0
500	0.8	93.0	96.8	99.2	93.1	96.8	99.2	93.3	98.1	99.8	93.3	96.9	99.1	89.2	94.3	98.5
1000	0.2	91.9	96.2	99.1	92.2	96.3	99.2	93.1	97.0	99.3	91.3	95.6	99.0	91.1	95.6	98.9
1000	0.5	92.3	96.5	99.2	92.1	96.1	99.1	92.4	97.1	99.4	92.4	96.4	99.1	90.5	95.3	98.9
1000	0.8	92.0	96.3	99.2	92.1	96.6	99.2	92.0	96.4	99.4	92.0	96.1	99.0	89.1	94.0	98.0

24

coverage levels for the cumulative delta-method bands remain close to the corresponding nominal levels for $\beta = 0.2, 0.5, 0.8$. The performance slightly improves if the sample size is increased. The overall performance is very stable across the different models.

The reported coverages for the simulation method are presented in the bottom panel of Table 2. The in-sample bands from the simulation method are based on $M = 1000$ simulations. The presented coverages show that the simulation method is most accurate overall. Again, the improvements in accuracy compared to the cumulative bands is clearly noticeable, but appears to be tolerable for empirical applications. This is a relevant finding, as the cumulative bands are easier and faster to compute than the simulation based ones. This finding applies to all levels of persistence for the time-varying parameter process and to all different models considered. In our next section, we investigate whether this conclusion also holds up for empirical data.

5.2. Out-of-Sample Forecast Bands

Next we verify the coverages of the out-of-sample forecast bands computed by the methods proposed in Section 4. In particular, we consider the GARCH, t-GARCH, t-GAS, ACD and LL models to assess the accuracy of the different methods to compute forecast bands. For each of the models, we consider two sample sizes of $T = 500$ and $T = 1000$, and report the actual coverage of the three different forecast bands with nominal coverage of 0.95. Coverage values are reported for k-step ahead forecasts with $k \in \{1, \ldots, 5, 10, 20\}$. As mentioned in Sections 3 and 4, we have avoided negative forecasts of conditional variances by re-parameterizing GARCH, t-GARCH and t-GAS models to ensure that appropriate parameter restrictions are satisfied, for example, by estimating the logarithm of ω rather than ω itself. This slightly alters the derivatives in ∇_t in a straightforward way by application of the chain rule. The results of a comprehensive Monte Carlo study are reported in Table 3. Our main findings are as follows.

First, we find that taking parameter uncertainty into account is of crucial importance for obtaining accurate forecast bands at any horizon. In particular, it is clear that when we do not take parameter uncertainty into account, the forecast bands are not wide enough. This results in a low coverage compared to the nominal value of 0.95 for all considered models and all forecast horizons. Table 3 shows that this problem is especially significant at short forecast horizons as the parameter uncertainty at such horizons is relatively more important than innovation uncertainty. The coverage improves for larger forecast horizons as the innovation uncertainty becomes progressively more important. Specifically, at $k = 1$, the first forecast bands are degenerate and hence fail to contain the true parameter with probability one. This is reflected by a coverage of zero.

Second, Table 3 further shows that even the two forecast bands that take parameter uncertainty into account typically perform better at larger forecast horizons. With the exception of the t-GAS model, the coverage of the bands improves substantially as k increases. It is a reminder of the fact that parameter uncertainty is more difficult to capture than innovation uncertainty.

Third, the out-of-sample forecast bands obtained by repetitive filtering of $\widehat{f}_{T+1}$ perform better than those obtained using the delta-method for $k = 1$. However, both methods perform similarly well for $k \geq 2$. As such, the extra computational effort required by the method based on repetitive filtering $\widehat{f}_{T+1}$ does not seem necessary when considering forecast bands for $k \geq 2$. We can simply conclude that the delta-method forecast bands are the best choice for forecasting horizons where innovation uncertainty matters, that is for $k \geq 2$.

6. An empirical illustration

To illustrate the use and the appearance of the different in-sample and out-of-sample bands in a an empirical study, we consider a monthly time series of Standard & Poor's

Table 3: Out-of-sample forecast band coverages

We present the forecast band coverages of $\widehat{f}_{T+k}$ for $k \in \{1, \ldots, 5, 10, 20\}$. The design of this study is the same as the one carried out for Table 2, with $\beta = 0.8$ and coverage level 95%. We consider three methods. The method *"fixed"* is based on the sample estimate of $\widehat{\theta}$ and the subsequent filtered value $\widehat{f}_{T+1}$ for computing the out-of-sample bands. The method *"delta approx"* is based on simulated values of $\widehat{\theta}$ and $\widehat{f}_{T+1}$ from their asymptotic normal distributions. The method *"filtered"* only simulates $\widehat{\theta}$ from the asymptotic normal approximation and applies the filter over the entire sample to obtain $\widehat{f}_{T+1}$ for computing the out-of-sample bands. All computations are based on $M = 1000$ simulations. The Monte Carlo study is based on $N = 1000$ replications.

| Model	T	f_{T+1}	k						
			1	2	3	4	5	10	20
GARCH	500	fixed	0.0	65.6	73.8	78.5	80.3	84.4	83.5
		delta approx.	81.2	91.1	92.1	92.6	92.4	93.5	91.9
		filtered	86.4	91.5	92.4	92.5	92.2	92.7	91.1
	1000	fixed	0.0	68.8	77.6	83.6	84.2	88.9	89.8
		delta approx.	77.1	87.5	90.0	91.7	91.8	93.7	93.9
		filtered	81.0	88.2	89.6	92.0	92.0	93.7	93.8
t-GARCH	500	fixed	0.0	58.9	68.7	74.1	75.5	77.8	81.0
		delta approx.	87.4	93.4	93.9	95.6	95.2	93.4	94.4
		filtered	92.4	95.3	95.7	95.2	95.1	93.3	94.5
	1000	fixed	0.0	64.7	75.8	79.3	82.4	86.0	86.1
		delta approx.	83.3	92.6	94.2	93.6	93.4	94.3	92.5
		filtered	87.2	93.7	94.1	94.0	93.5	93.9	91.4
t-GAS	500	fixed	0.0	71.0	77.8	80.7	80.8	80.4	80.1
		delta approx.	90.0	95.8	95.8	96.1	95.5	94.5	94.8
		filtered	93.0	94.9	95.5	95.1	95.0	94.3	94.7
	1000	fixed	0.0	74.6	83.2	86.8	87.7	88.0	88.2
		delta approx.	94.0	95.8	95.5	95.8	95.8	95.0	94.8
		filtered	94.7	96.1	95.6	95.4	95.4	94.5	94.8
ACD	500	fixed	0.0	72.5	83.0	86.6	88.7	91.7	91.5
		delta approx.	70.5	91.0	93.7	94.5	95.0	94.8	95.3
		filtered	82.2	91.7	93.3	94.2	94.8	94.6	94.7
	1000	fixed	0.0	75.7	84.3	87.3	89.4	92.0	91.5
		delta approx.	67.5	88.3	92.0	92.7	94.0	93.9	93.5
		filtered	80.0	88.4	91.7	91.7	93.0	93.9	93.4
LL	500	fixed	0.0	78.9	84.7	85.5	87.5	87.7	88.0
		delta approx.	74.4	91.1	93.0	92.7	94.1	94.7	94.5
		filtered	87.3	93.8	95.2	94.4	95.1	95.9	95.4
	1000	fixed	0.0	84.7	86.4	86.9	89.3	89.7	90.7
		delta approx.	64.7	91.3	91.6	91.8	92.5	92.7	93.4
		filtered	75.9	92.7	92.7	92.9	93.1	93.5	94.2
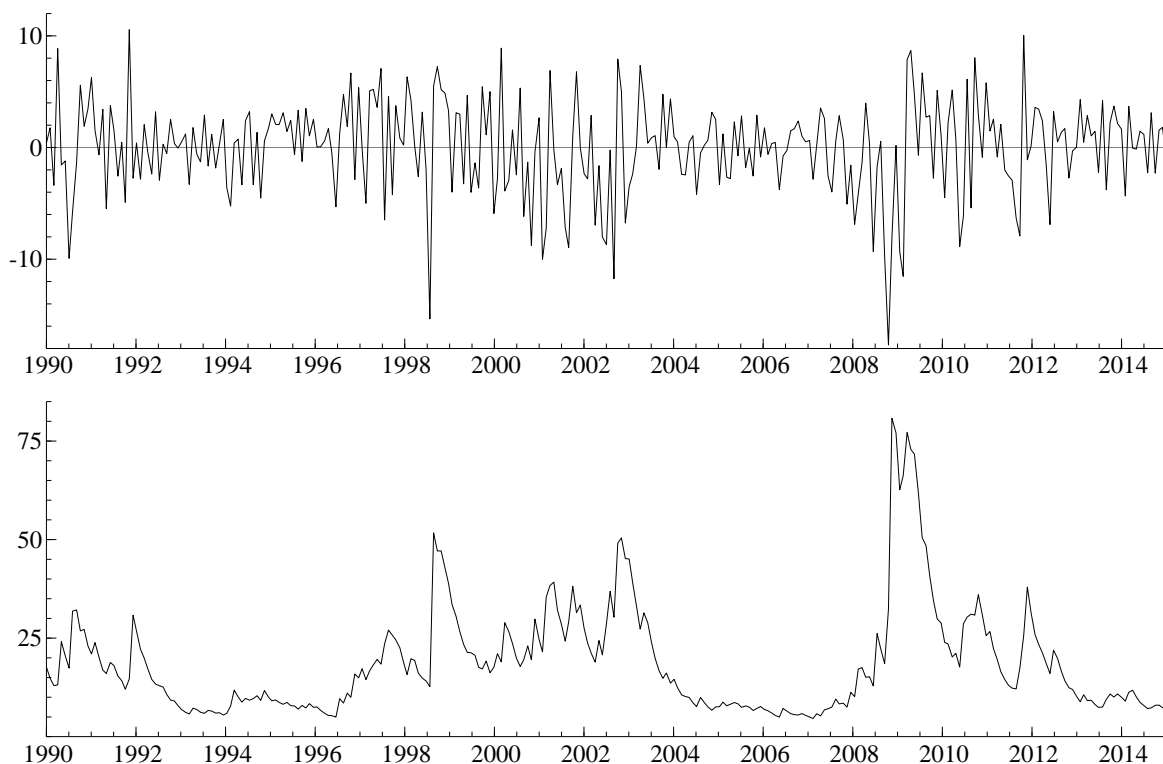
Figure 1: We present the S&P500 monthly returns from January 1990 until January 2015 (top panel) and the conditional volatility estimates obtained from the standard GARCH model (bottom panel).

(S&P) 500 index returns. Our sample covers the period January 1990 until January 2015. We consider the standard GARCH model of Example 1 for illustrative purposes with non-zero conditional mean μ. The returns for our sample are displayed in Figure 1. The maximum likelihood estimates for the coefficients are given by $\widehat{\mu}_T = 0.145$ (0.195), $\widehat{\omega}_T = 0.583$ (0.346), $\widehat{\alpha}_T = 0.171$ (0.053), and $\widehat{\beta}_T = 0.805$ (0.054), with the corresponding standard errors in parentheses, where we set f_1 equal to the sample variance. The filtered conditional variances $\widehat{f}_t$ are presented in the second panel of Figure 1.

The standard errors of the filtered $\widehat{f}_t$s are computed by our three methods. We present the results over the last 8 years of the sample in Figure 2. We clearly corroborate that the simple non-cumulative bands are unrealistically small compared to the other two bands.
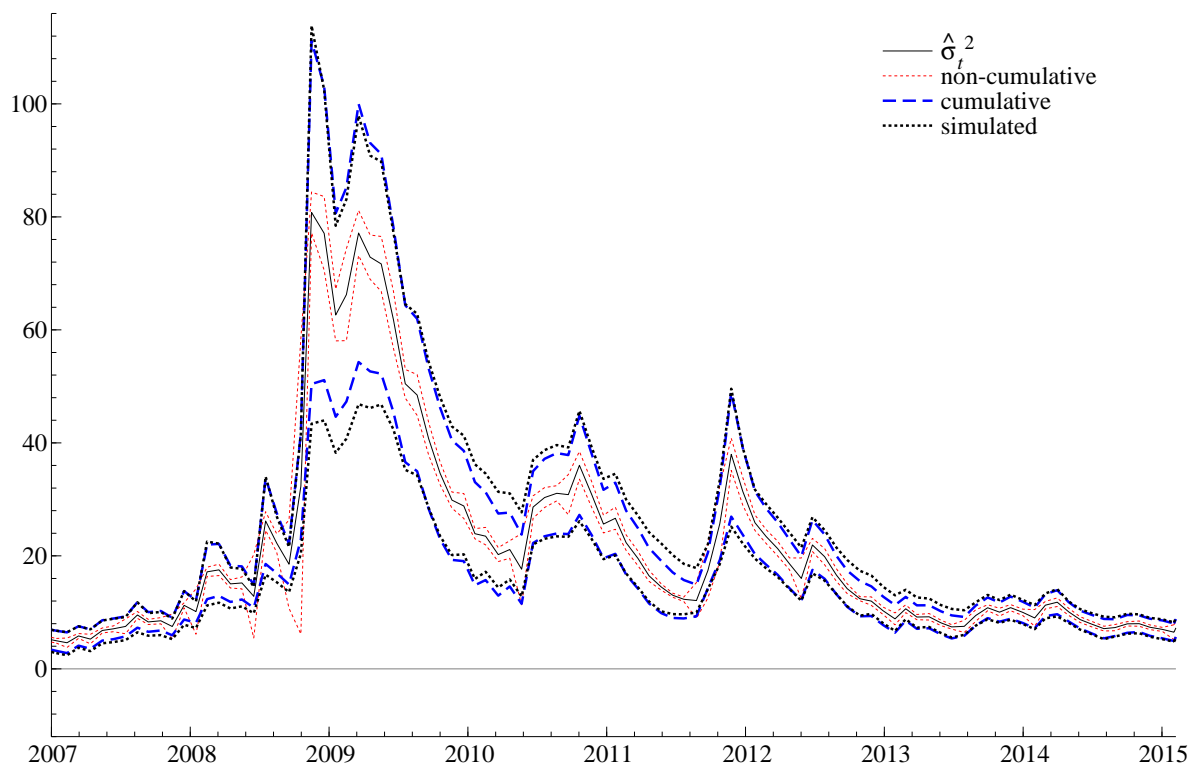
28

Figure 2: We present conditional volatility estimates and their corresponding in-sample error bands from the GARCH model applied to the S&P500 index monthly returns from January 1990 until January 2015; we only present these for the last 8 years (96 months), from January 2007. The in-sample bands are obtained from the non-cumulative simple method (light dotted line), the cumulative delta-method (dashed line) and the simulated method (dark dotted line).

It is interesting to see that the simulation bands and the cumulative delta bands can be hardly distinguished from each other in the presented sample. The largest difference can be noted at the height of the financial crisis, where the lower band of the simulation approach lies substantially below the cumulative lower band. In all other parts of the sample, the analytic cumulative bands appear to be highly successful in approximating the bands of uncertainty around $\widehat{f}_t$, i.e., in terms of accuracy, computational simplicity and speed.

Figure 3 presents the conditional volatility estimates from the GARCH model applied to the S&P500 monthly returns and their corresponding out-of-sample forecast bands
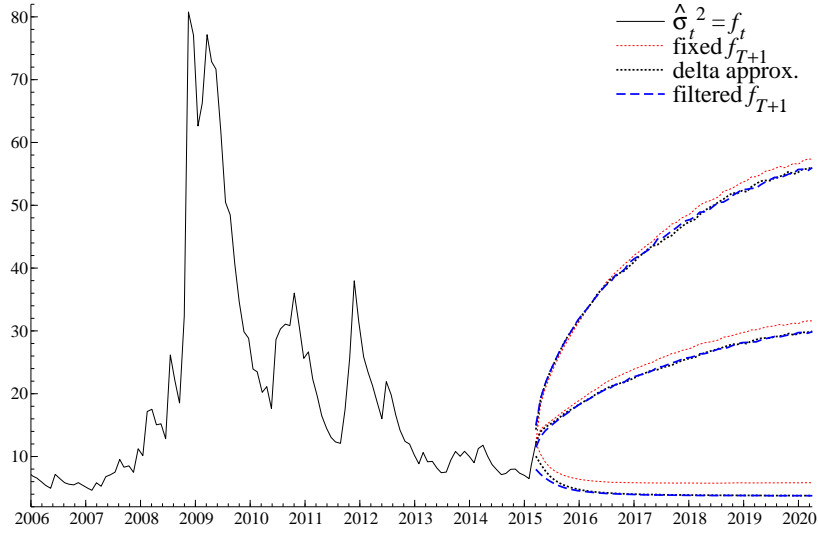
Figure 3: We present conditional volatility estimates from the GARCH model applied to the S&P500 monthly returns and their corresponding out-of-sample conditional expectations and 95% forecast bands.

with 95% nominal coverage. The three different forecast bands produced by the methods studied in Section 5 are similar in many respects. First, at the upper end they all have similar magnitude and increase sharply over the first years of the total forecast horizon. Second, at the lower end they decay to low volatility levels over the forecast horizon. The main visible difference is that the band that does not incorporate parameter uncertainty (i) has a degenerate band at $T + 1$, (ii) predicts a higher median volatility over the forecast horizon, and (iii) produces a forecast band that stabilizes at a low-volatility level that is permanently higher than its two competitors. According to our Monte Carlo result, this may result in poorer coverage.

7. Conclusions

We have reviewed three different methods for the computation of in-sample confidence bands and three different methods for out-of-sample forecast bands for time-varying parameters in a general class of observation driven time series models. The in-

sample bands reflect the uncertainty due to parameter estimation. The forecast bands in addition reflect the uncertainty due to future innovations of the time-varying parameter. In a Monte Carlo study we showed that the simulation-based methods are most accurate for obtaining the correct nominal coverage levels. Analytical non-cumulative simple bands are not sufficiently accurate and should not be used in practice if the time-varying parameter possesses persistent autoregressive dynamics. Interestingly, our analytic approximate bands based on a simple recursive delta-method approximation appear to reflect parameter uncertainty quite well, both in a controlled Monte Carlo study and in our empirical illustration. The delta-method is straightforward and fast. Therefore it can easily be implemented in standard software packages. This would provide useful additional information in empirical analyses, where confidence bands around the time-varying parameters are usually not presented. The out-of-sample forecast bands can be computed by efficient simulation methods that account for uncertainty due to future innovations and parameters. Although we have illustrated the methods for well-known models, the different approaches for computing in-sample and out-of-sample bands can also be used for other observation driven models from the literature, such as the skewed Student's t distribution with time-varying parameters as in Lucas, Schwaab, and Zhang [2014] and for dynamic discrete data models as in Rydberg and Shephard [2003].

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics 29*(4), 552–563.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Davis, R., W. Dunsmuir, and S. Streett (2003). Observation-driven models for Poisson counts. *Biometrika 90*, 777–790.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica 50*, 987–1008.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics 17*(5), 425–446.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127–1162.

Grammig, J. and K. O. Maurer (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal 3*, 16–38.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge: Cambridge University Press.

Lucas, A., B. Schwaab, and X. Zhang (2014). Measuring credit risk in a large banking system: econometric modeling and empirics. *Journal of Business and Economic Statistics 32*, 271–284.

Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics and Data Analysis 50*, 2293–2312.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics 1*(1), 2.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.