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] *Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands*
[b] *CREATES, Aarhus University, Denmark*

A B S T R A C T

We study the performances 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, while the out-of-sample bands reflect not only parameter uncertainty, but also innovation uncertainty. The bands are applicable to a wide range of estimation procedures and a large class of observation driven models with differentiable transition functions. A Monte Carlo study is conducted to investigate time-varying parameter models such as generalized autoregressive conditional heteroskedasticity and autoregressive conditional duration models. Our results show convincing differences between the actual coverages provided by the different methods. We illustrate our findings in a volatility analysis for monthly Standard & Poor's 500 index returns.

© 2016 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Over recent decades, time-varying parameter models have become increasingly popular in empirical economics and finance. The rapid 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. Initially, starting in the 1960s, time-varying parameters for the mean equation in linear Gaussian models were typically handled by the Kalman filter and related methods. For linear Gaussian state space models, the Kalman filter can be used to calculate the conditional means and variances of unobserved time-varying parameters (or linear functions

thereof) in a computationally efficient way; for a detailed treatment, see for example Durbin and Koopman (2012). In this modeling framework, the construction of in-sample confidence bands and out-of-sample forecast bands is straightforward, and is performed on a routine basis, as expressions for the conditional variances of the time-varying parameters are available explicitly. In the case of nonlinear and/or non-Gaussian extensions of state space models, the computation of confidence bands can be somewhat more involved. One example is the stochastic volatility model, for which the analysis is typically based on simulation-based methods; see the discussion by Shephard (2005).

Since the 1980s, new classes of model for time-varying parameters have been developed. Specifically, models for the time-varying conditional variance have received considerable attention in the empirical economics and finance literature. For example, the generalized autoregressive conditional heteroskedasticity (GARCH) model of Bollerslev (1986) and Engle (1982) has led to a range of model

---

* Corresponding author at: Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands.

E-mail addresses: f.blasques@vu.nl (F. Blasques), s.j.koopman@vu.nl (S.J. Koopman), k.a.lasak@vu.nl (K. Łasak), a.lucas@vu.nl (A. Lucas).

formulations for time-varying parameters. In the standard ARCH and GARCH models, the conditional variance is obtained by filtering past observations through a volatility updating equation. The relative simplicity of GARCH models has spurred their widespread adoption by both academics and professionals.

In most empirical studies, the estimated volatility from the GARCH model is presented without in-sample bands reflecting the parameter uncertainty in the volatility updating equation. Similarly, volatility forecasts may feature bands that reflect innovation uncertainty, but they typically ignore the parameter uncertainty. Exact analytical results are not available because the filters are highly nonlinear functions of past observations, and as a result, statistical software rarely provides either in-sample confidence bands or out-of-sample forecast bands to represent the estimates of such time-varying parameters. This argument also applies to other models that are 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, Koopman, and Lucas (2013). All of these models belong to the class of observation-driven models, as opposed to parameter-driven models; see Cox (1981) for a detailed description of these two classes of time series models.

We analyze various different methods of constructing in-sample and out-of-sample bands. For our in-sample bands, we compare two analytical methods and one simulation-based method. All of these bands reflect the parameter uncertainty only. The approximate analytical bands require only simple computations, and are not subject to random fluctuations due to simulation error. These analytical bands can be used when the updating equation is differentiable and the (asymptotic) distribution of the estimator for the static parameters is known. For the computation of forecast bands, we compare three simulation-based procedures. The first method takes only 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 quick to compute. In particular, we argue that the necessary computations are more efficient than the bootstrapped forecast bands proposed by Pascual, Romo, and Ruiz (2006) for GARCH models, for example.

All of the methods that we consider can be implemented readily in software packages. We investigate the coverage probabilities of each of these different approaches in detail over a range of different time-varying parameter models, and 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 time series generated using 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-based in-sample confidence bands and out-of-sample forecast bands all obtain accurate coverage levels. An empirical illustration 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 the 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.

The remainder of this paper is organized as follows. Section 2 introduces the class of observation-driven models. Section 3 introduces different methods of 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 performances 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 on past observations. In these models, the focus is on the specification of the mechanism through which past realizations of the variable of interest affect the current value of the time-varying parameter.

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

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

where density $p_y(\cdot)$ is implied by an "observation equation" for y_t and depends on 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, $\varepsilon_t \sim \text{NID}(0, 1)$, where NID is normally independently distributed with zero mean and unity variance. The time-varying parameter is defined formally as a function $f_t := f_t(y^{1:t-1}, f_1; \boldsymbol{\theta})$ that depends on the past observations $y^{1:t-1} := \{y_1, y_2, \ldots, y_{t-1}\}$, an initial value f_1, and a static parameter vector $\boldsymbol{\theta}$. The updating function for the time-varying parameter can be expressed in different ways. For example, if we consider a linear updating equation consisting of lagged values of y_t and f_t, 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 a (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 (Eq. (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 (Eq. (2)) in the more general form of a stochastic recurrence equation,

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

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

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

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

for $t = 1, \ldots, T$, where $\{\varepsilon_t\}$ is a sequence of NID(0, 1) random variables. Hence, we have that $p_y(y_t|f_t; \boldsymbol{\theta}) = $ NID$(0, f_t)$ in Eq. (1). The GARCH-t model of Bollerslev (1986) is obtained when we replace the normal distribution with the Student's t distribution.

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

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

for $t = 1, \ldots, T$, where $\{\varepsilon_t\}$ is a sequence of independent random variables with the standard exponential distribution Exp(1). Different specifications of this modeling framework for varying durations and intensities are discussed by Grammig and Maurer (2000).

The score driven models of Creal et al. (2013) and Harvey (2013) also belong to the model class represented by Eqs. (1) and (2), where $p_y(y_t|f_t; \boldsymbol{\theta})$ can be any density, with f_t as the time-varying parameter. The specific choice of the function $s(\cdot)$ distinguishes the score driven model from other models, namely

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

for a scaling function. In many time series models that are 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 $S_t(f_t; \boldsymbol{\theta})$ can either be simply set to unity, or alternatively be set to reflect the local curvature in the log conditional density function at time t.[1]

When we consider the observation densities of Examples 1 and 2 and apply the score-driven framework of Creal et al. (2013) with scaling matrix S_t equal to the inverse conditional Fisher information matrix, we obtain exactly the GARCH and ACD models, respectively. Indeed, the group of score-driven models includes 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 Eq. (1), we obtain, not the GARCH-t model, but the model given in Example 3 below, which is discussed further by Creal, Koopman, and Lucas (2011); Creal et al. (2013) and Harvey (2013).

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, and f_t is its squared scale parameter (rather than its variance). 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 in order to obtain estimates of 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 decomposition. 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 of f_t are evaluated via Eq. (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 that are 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 maximizing $\ell(\theta; f_1)$ numerically 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 Eq. (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 weighted functions of f_1 and $y^{1:t-1}$, with the weights determined by $\widehat{\boldsymbol{\theta}}_T$, for $t = 2, \ldots, T$. Thus, 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 with its maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_T$ when computing $\widehat{f}_t$;
- Filtering uncertainty: we consider the updating equation (Eq. (2)) for f_t, but we do not know the true underlying time-varying parameter process;
- Model uncertainty: we consider $p_y(\cdot)$ in Eq. (1), but we do not know the true data generation process for y_t.

Here, we concentrate on developing bands that reflect the parameter uncertainty for $\widehat{f}_t$, $t = 1, \ldots, T$, while ignoring the other two sources of uncertainty. This is still an important step forward in the current literature, where in-sample uncertainty bands around the time-varying parameter are almost never shown. In the next section, we present several ways of constructing these approximate bands, and in Section 5, we investigate the coverage properties of these bands by conducting a Monte Carlo study. In Section 4, we develop out-of-sample bands that also reflect innovation uncertainty.

---

[1] Further details of score-driven models, along with various theoretical and empirical developments, are provided on the website http://www.gasmodel.com.

3. In-sample bands

Conditional on past information and the initial value f_1, and based on the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_T$ of $\boldsymbol{\theta}$, we develop confidence bands for $\widehat{f}_t$ that reflect parameter uncertainty about the true value of $\boldsymbol{\theta}$. The bands are based on the variance of $\widehat{f}_t$, which we denote as $V_t := \mathbb{V}\mathrm{ar}(\widehat{f}_t)$. Thus, the bands reflect the randomness of the estimator $\widehat{\boldsymbol{\theta}}_T$. We propose three different methods for obtaining approximations of 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 Eq. (3) and using the asymptotic variance of $\widehat{\boldsymbol{\theta}}_T$. These analytic methods have the advantage that they are easy to implement in software packages, as they are based on simple calculations. The third method provides the most accurate reflection of parameter uncertainty in $\widehat{f}_t$, but has the disadvantage that the computations are subject to randomness by construction. This method also uses the asymptotic variance of $\widehat{\boldsymbol{\theta}}_T$, 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 naïve, and therefore deliver accurate results only 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 Eq. (2), or when it is sufficiently small. For example, the method is appropriate for the ARCH model of Engle (1982), which is the model in Example 1 with $\beta = 0$.

The *cumulative delta-method bands* take into account the accumulation of parameter uncertainty in the updating process. 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 with not only the accumulation of the uncertainty due to the parameter vector $\boldsymbol{\theta}$, but also the effects of possible nonlinear functional relationships 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 of the methods for computing bands rely on the asymptotic distribution of the estimator $\widehat{\boldsymbol{\theta}}_T$ being known. Typically, this (asymptotic) distribution is normal. Then, the bands can be computed for any estimator $\widehat{\boldsymbol{\theta}}_T$ as long as we can simulate from its asymptotic distribution. 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 Eq. (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 when calculating the approximate variance of $\widehat{f}_{t+1}$. Our proposed simple bands then rely on the approximation

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

for any $\widehat{f}_t$ and around a 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

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

Hence, we can write the approximation in Eq. (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 a given $\widehat{f}_t$ is then obtained by

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

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

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

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*. Replacing $\mathbf{W}$ with a consistent estimate $\widehat{\mathbf{W}}$, we have

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

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

$$V_{t+1} \approx \sum_{i=1}^{q} T^{-1} w_{i,i} \nabla_{i,t}^2 + 2 \sum_{1 \le i < j \le q} T^{-1} w_{i,j} \nabla_{i,t} \nabla_{j,t}, \qquad (8)$$

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

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

These results can also be adapted accordingly for other asymptotic distributions of the maximum likelihood estimator.

The effect of the initial conditions on the in-sample bands is non-trivial and requires more discussion. If f_1 is fixed, there is clearly no effect. However, it is more common for us to let f_1 depend on $\boldsymbol{\theta}$. For example, if we set the initial condition to the unconditional mean, that is $f_1 = \omega/(1 - \alpha - \beta)$, as in a standard GARCH model with $|\alpha + \beta| < 1$. When f_1 is in effect estimated as part of $\boldsymbol{\theta}$, it cannot be estimated consistently, but we still propose to approximate the distribution of $\widehat{f}_1$ as in Eq. (7). In a stationary setting, we only expect the estimation of f_1 to affect the in-sample bands for the first stretch of observations.

Example 4 (Example 1 *Continued*)**.** Consider the ARCH model of Engle (1982), which is defined as the GARCH model of Example 1 with $\beta = 0$ in Eq. (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 the first and second elements of the parameter vector $\widehat{\boldsymbol{\theta}}_T$, respectively. It follows immediately that

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

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

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

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

There are various different approaches that we can use for the computation of the estimate $\widehat{\mathbf{W}}$. Under an assumption of correct specification, we can use the inverse negative Hessian of the log likelihood evaluated at the optimum. Alternatively, we can use a robust covariance matrix estimator, as per Andrews (1991) and White (1980). In our simulations later on, we use the covariance matrix estimator of White (1980). As has been mentioned, alternative estimators can also be used, and may result in a somewhat different behaviors.

In Section 5, we provide evidence that our simple non-cumulative bands are not appropriate for observation-driven models when the filter has strong autoregressive dynamics. This applies to the GARCH model with $\beta \ne 0$,

and in particular to the case where β is a long way from zero.

3.2. In-sample cumulative delta-method bands

The previous approximations did not account for the fact that $\widehat{f}_t$ itself also 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 tracked analytically using the delta-method. This is based on the same linearization as that of $\phi(\cdot)$ in Eq. (4), but where we also explicitly account for the dependence of $\widehat{f}_t$ on $\widehat{\boldsymbol{\theta}}_T$. We redefine $\nabla_{i,t}$ as

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

We assume that the estimator $\widehat{\boldsymbol{\theta}}_T$ is consistent and that the point estimate is sufficiently close to the true parameter $\boldsymbol{\theta}_0$. Hence, when calculating $\nabla_{i,t}$, we take f_t as the value of the filter at time t for the point estimate of $\widehat{\boldsymbol{\theta}}_T$, 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

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

which can be computed in parallel to the recursion for f_t itself. These recursions can be initiated by $\partial f_1 / \partial \theta_i$. The initial conditions may differ depending on the definition of f_1; for example, if f_1 is fixed and independent of $\boldsymbol{\theta}$, than the initial derivative is zero. Different initial conditions apply if f_1 is part of $\boldsymbol{\theta}$ itself, or if f_1 is a function of $\boldsymbol{\theta}$; for example, $f_1 = \omega/(1 - \alpha - \beta)$ in the GARCH context.

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

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

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

Example 5 (Example 1 *Continued*)**.** Consider the GARCH model of Bollerslev (1986), as given by Eq. (2), and with $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$. 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 $t = 1, \dots, T$. Applying Eq. (10) to the GARCH updating function, it follows that

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

$$\nabla_{3,t} = f_t + \beta \frac{\partial f_t}{\partial \beta}, \quad t = 1, \dots, T,$$

where the partial derivatives on the right-hand side are defined via the recursion in Eq. (11). We evaluate f_t and f_{t-1} as $\widehat{f}_t$ and $\widehat{f}_{t-1}$, respectively, and replace the elements of $\boldsymbol{\theta}$ with the 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 computations required are more involved than those required for the simple non-cumulative method used in Example 4. However, the bands are still straightforward to implement and quick 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 using the simulation method below avoid the linearization step altogether. The resulting bands are subject only to the error in approximating the distribution of the estimator using its asymptotic distribution.

3.3. In-sample simulation-based bands

An alternative approach for obtaining in-sample confidence bands for $\widehat{f}_{t+1}$ involves the use of simulation methods. Since the bands reflect only 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 instead draw parameter values from the approximating asymptotic distribution

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

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 Eq. (2) or Eq. (3) with $f_1^i = f_1$. In particular, we consider the filtering recursion for each i, that is

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

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 in $\widehat{\boldsymbol{\theta}}_T$ is characterized by the asymptotic distribution in Eq. (7). Also note that the bands in the simulation method need not be based on V_{t+1}, but rather can 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 become more accurate than those based on the linearization methods. However, simulations can be time-consuming when the sample size T is large and 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 implementation in software packages.

Example 6 (Example 1 *Continued*)**.** Consider the GARCH model with $y_t \sim NID(0, f_t)$ and $f_{t+1} = \omega + \alpha y_t^2 + \beta f_t$, with initial value f_1. The simulation bands can be obtained simply by drawing M parameter vectors from the approximate density of the estimator, as in Eq. (12); then, we obtain $\widehat{\boldsymbol{\theta}}_T^i = (\widehat{\omega}_T^i, \tilde{\alpha}_T^i, \tilde{\beta}_T^i)$ and

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

We can compute in-sample bands for any confidence level from these simulated paths.

It is important to note that the use of an approximate asymptotic distribution may result in us drawing values $\widehat{\boldsymbol{\theta}}_T^i$ that lie outside the admissible 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. Draws from outside the parameter space of interest can be avoided by re-parameterizing the model in such a way as to ensure that Θ is unbounded for the transformed parameters. For example, we can easily ensure that $\omega = \exp(\omega^*)$ is strictly positive by estimating $\omega^* = \log(\omega)$.

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 when constructing 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 constructing out-of-sample forecast bands. The first method considers only innovation uncertainty, while the second and third methods 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 $n \geq 2$, is to first take the point estimate $\widehat{\boldsymbol{\theta}}_T$ and

the resulting filtered value $\widehat{f}_{T+1}$ as given, then extrapolate the filter $f_{T+1}, \ldots, f_{T+n}$ multiple times by drawing multiple innovation paths from the estimated innovation density and 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 of interest 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 $T + 1$; that is,

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

2. Given the observations $y_{T+1}^1, \ldots, y_{T+1}^S$ and the updating equation (Eq. (3)), we obtain filtered values $\widehat{f}_{T+2}^1, \ldots, \widehat{f}_{T+2}^S$, conditional on $\widehat{\boldsymbol{\theta}}_T$ and $\widehat{f}_{T+1}$, using the update

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

3. For each $\widehat{f}_{T+2}^s$, $s = 1, \ldots, S$, repeat 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, it does not take into account the randomness of either the estimator $\widehat{\boldsymbol{\theta}}_T$ or the filtered value $\widehat{f}_{T+1}$.

4.2. Delta-method forecast bands

We incorporate parameter uncertainty into the forecast bands by making use of the asymptotic distribution of the estimator of the parameter. In particular, we follow the approach 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 in Eq. (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. We achieve further computational efficiency by building 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 to the full sample M times 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 the 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 them. 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

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

where the matrix $\widehat{\boldsymbol{\Sigma}}$ 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., $\widehat{\boldsymbol{\Sigma}} = (I_q, \nabla_t)'\widehat{\boldsymbol{W}}(I_q, \nabla_t)$.
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 in Section 3.2, it is important to emphasize that the use of an approximate asymptotic distribution may result in the drawing of values $\widehat{\boldsymbol{\theta}}_T^i$ that lie outside a parameter space of interest Θ. In our Monte Carlo exercise in Section 5, we find that a re-parameterization of the GARCH model is important to ensure the positivity of the conditional variance in our forecast bands. We do this by estimating $\omega^* = \log(\omega)$, $\alpha^* = \log(\alpha/(1 - \alpha - \beta))$, and $\beta^* = \log(\beta/(1-\alpha-\beta))$, such that $\omega = \exp(\omega^*) > 0$, $\alpha = \exp(\alpha^*)/(1 + \exp(\alpha^*) + \exp(\beta^*))$, and $\beta = \exp(\alpha^*)/(1 + \exp(\alpha^*) + \exp(\beta^*))$, with $\alpha, \beta > 0$ and $\alpha + \beta < 1$. This also imposes stationarity on the model. If stationarity cannot be imposed, the simulations may sometimes result in explosive behavior, which has a negative effect on the performance of the bands. Simulations $\widehat{\boldsymbol{\theta}}_T^i$, as well as the delta method approximations, are now based on $\boldsymbol{\theta} = (\omega^*, \alpha^*, \beta^*)$ rather than on $\boldsymbol{\theta} = (\omega, \alpha, \beta)$.

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 the underlying parameter uncertainty accurately. 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 then 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 bands that reflect both parameter and innovation uncertainty.

The key feature of these forecast bands relative to the previous method is their ability to obtain values for $\widehat{f}_{T+1}^i$

---

[2] It is also possible 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 $\boldsymbol{\theta}$.

that reflect the parameter uncertainty about the static parameters $\boldsymbol{\theta}$ more accurately. This additional accuracy comes at a 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 GARCH models, as it does not require the re-estimation of $\widehat{\boldsymbol{\theta}}_T$. This alternative method of obtaining forecast bands can be summarized as follows:

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

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

2. For each $\widehat{\boldsymbol{\theta}}_T^i$, use the observed data $y_1, \ldots, y_T$ and the updating equation (Eq. (3)), and run this 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.

This completes our development of three different simulation based methods for constructing out-of-sample forecast bands for time-varying parameters.

5. Monte Carlo study

5.1. In-sample confidence bands

We verify the performances of the three methods of computing in-sample bands for time-varying parameters in observation-driven models by carrying out a Monte Carlo study. We consider five different models: (i) the GARCH model based on normally distributed innovations; (ii) a 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. If the density $p_y(y_t|f_t; \boldsymbol{\theta})$ is $NID(0, f_t)$, we obtain Model (i); if it is Student's $t(5)$ with variance f_t and degrees of freedom λ, we obtain Model (ii). Model (iii) has $s(y_t, f_t; \boldsymbol{\theta}) = (1 + 3\lambda^{-1})((1 + \lambda^{-1})y_t^2/(1 + \lambda^{-1}y_t^2/f_t) - f_t)$ with $\lambda > 2$ degrees of freedom. 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 consider two different sample sizes, $T = 500, 1000$, and 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 examine how well the three methods presented 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 T.
2. For this generated time series, estimate $\boldsymbol{\theta}$ using the maximum likelihood method, as discussed in Section 2.
3. Calculate the time-varying parameter sequence $\widehat{f}_{t+1}$ and its derivative with respect to $\boldsymbol{\theta}$ using Eqs. (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 coverages 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 time series generated, we determine the in-sample bands using the three methods described in Section 3. As our results are designed for constructing pointwise confidence bands for f_t, we compute (for each t) the number of times that the true f_t lies inside the in-sample band. We then take the average of this number over the T periods and 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 $\beta = 0.8$, but 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 was 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 of the sample sizes considered.

The cumulative delta-method bands are considerably more accurate. The bands are slightly too small, resulting in lower estimated coverages than the nominal levels; however, the differences appear to be tolerable for most practical purposes. The performances for different values of β remain stable, while those for the simple non-cumulative bands deteriorate rapidly if the value of β is increased. The estimated 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 improves slightly if the sample size is increased. Overall, the performance is very stable across the different models.

The coverages reported 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 coverage results presented show that the simulation method is most accurate overall. The coverage is often somewhat higher than the nominal level. This does not appear to be due to the (White, 1980) estimator that we use for $\widehat{\mathbf{W}}$. In fact, using a standard negative inverse Hessian results in even higher coverage levels. Thus, the selection of both the 'best' $\widehat{\mathbf{W}}$ for constructing confidence bands and the 'best' parameterization $\boldsymbol{\theta}$ for the asymptotic normality approximation to be applicable are topics that require further investigation. Despite the higher coverage levels of the simulation-based bands, the improvements in accuracy relative to the cumulative bands are clearly noticeable, but the deviation in accuracy of the cumulative bands appears to be tolerable for empirical applications.

Table 1
Models in the Monte Carlo study.

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

The table presents 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 Eq. (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 $p_y(y_t|f_t;\theta)$ and $s_t=s(y_t,f_t;\theta)$ are given below with reference to the examples discussed in Section 2. Student's $t(0,f_t,\lambda)$ refers to the Student's t density function with a 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 $f_1=1$.

Table 2
In-sample band coverages.

T	β	GARCH			t-GARCH			t-GAS			ACD			LL		
		90	95	99	90	95	99	90	95	99	90	95	99	90	95	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	94.5	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	93.4	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	91.9	96.4	99.4	92.0	96.1	99.0	89.1	94.0	98.0

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 described in Table 1. In the case of the simulation method, we use $M=1000$ simulations. The Monte Carlo study is based on $N=1000$ replications.

This is a relevant finding, as the cumulative bands are easier and quicker to compute than the simulation-based ones. This finding applies to all levels of persistence for the time-varying parameter process and to all of the different models considered. In the 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 using the methods proposed in Section 4. In particular, we consider the GARCH, t-GARCH, t-GAS, ACD and LL models, to assess the accuracies of the different methods for computing forecast bands. For each of the models, we consider two sample sizes, $T=500$ and $T=1000$, and report the actual coverages of the three different forecast bands with nominal coverages of 0.95. Coverage values are reported for k-step-ahead forecasts, with $k\in\{1,\ldots,5,10,20\}$. As was mentioned in Sections 3 and 4, we have avoided negative forecasts of the conditional variances by re-parameterizing the GARCH, t-GARCH and t-GAS models to ensure that the appropriate parameter restrictions are satisfied; for example, by estimating the logarithm of ω rather than ω itself. This alters the derivatives in ∇_t slightly but 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 the parameter uncertainty into account is crucially important for obtaining accurate forecast bands at any horizon. In particular, it is clear that the forecast bands are not wide enough when we do not take parameter uncertainty into account. This results in a low coverage compared to the nominal value of 0.95 for all models and forecast horizons considered. 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 the innovation uncertainty. The coverage improves for larger forecast horizons as the innovation uncertainty becomes progressively more

Table 3
Out-of-sample forecast band coverages.

Model	T	$\widehat{f}_{T+1}$	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

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 that carried out for Table 2, with $\beta = 0.8$ and a 95% coverage level. We consider three methods. The method "*fixed*" is based on the sample estimate of $\widehat{\boldsymbol{\theta}}$ 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 $\widehat{\boldsymbol{\theta}}$ and $\widehat{f}_{T+1}$ from their asymptotic normal distributions. The method "*filtered*" only simulates $\widehat{\boldsymbol{\theta}}$ from the asymptotic normal approximation, then 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.

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 the parameter uncertainty into account typically perform better at longer forecast horizons. With the exception of the t-GAS model, the coverage of the bands improves substantially as k increases. This 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 a repetitive filtering of $\widehat{f}_{T+1}$ perform better than those obtained using the delta method for $k = 1$. However, the two 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 forecast horizons where the innovation uncertainty matters, that is, for $k \geq 2$.

6. An empirical illustration

To illustrate the use and appearance of the different in-sample and out-of-sample bands in an empirical study, we consider a monthly time series of Standard & Poor's (S&P) 500 index returns. Our sample covers the period from 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 the first panel of Fig. 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 Fig. 1.

The standard errors of the filtered $\widehat{f}_t$s are computed by our three methods. We present the results over the last 97 months of the sample in Fig. 2. We clearly confirm that the simple non-cumulative bands are unrealistically small compared to the other two bands. It is interesting to see that the simulation bands and the cumulative delta bands
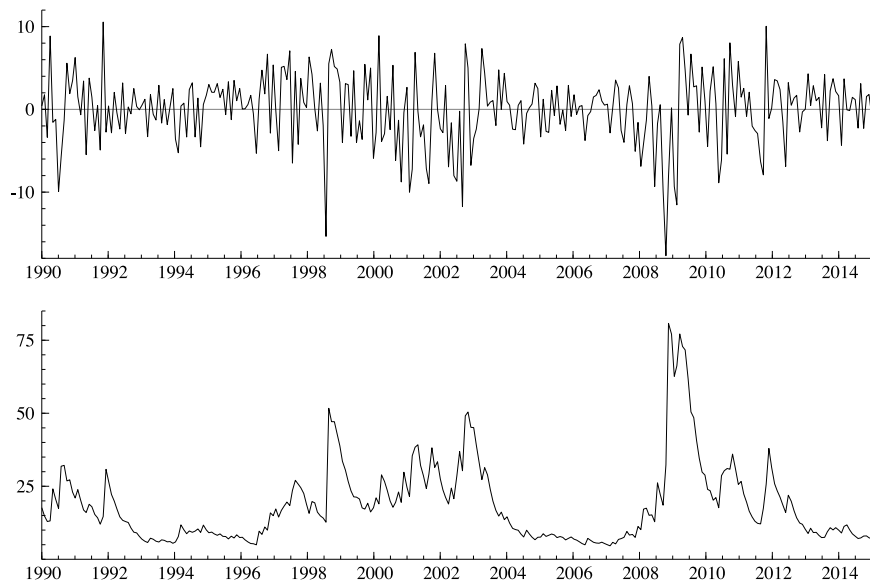
Fig. 1. S&P500 monthly returns from January 1990 until January 2015 (top panel) and the conditional volatility estimates obtained using the standard GARCH model (bottom panel).
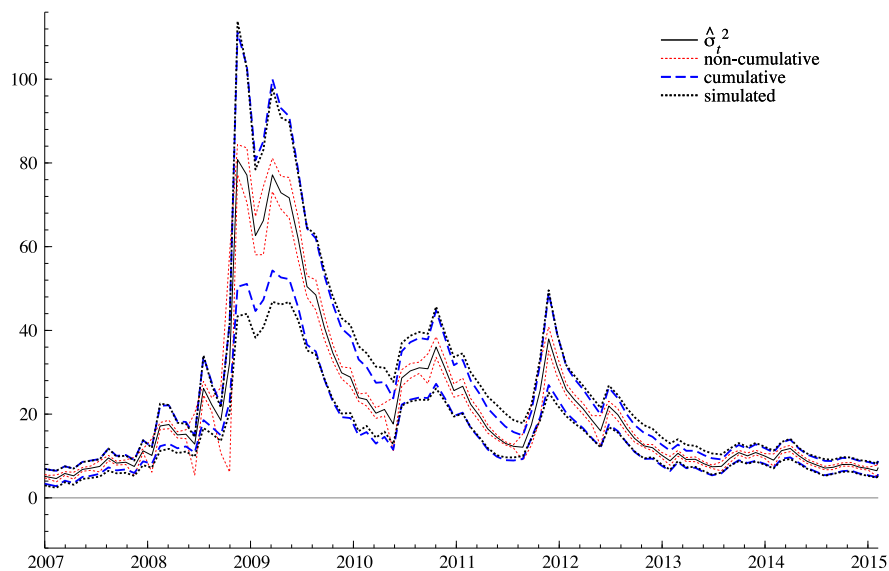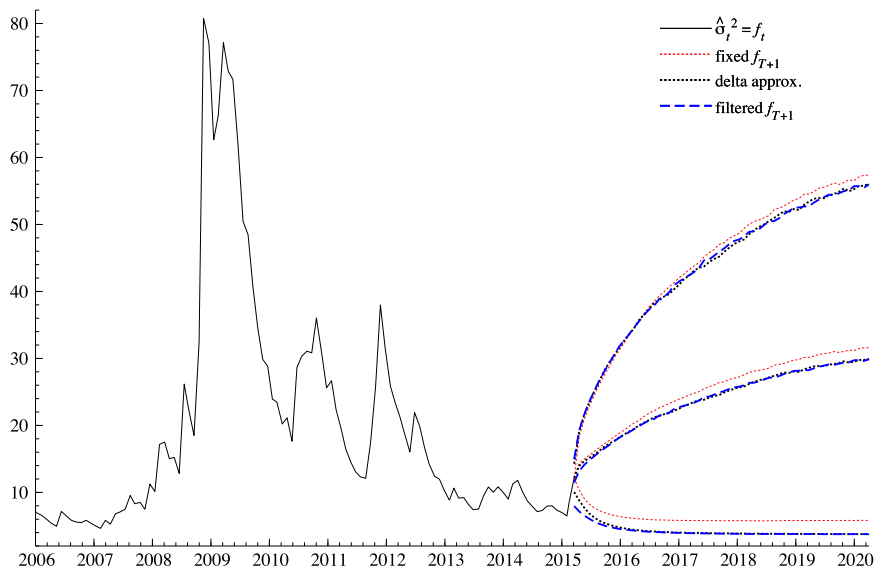

Fig. 2. Conditional volatility estimates and their corresponding in-sample error bands from the GARCH model applied to the S&P500 index monthly returns from January 1990 until January 2015. Note that here we only present the results for the last eight years and one month (97 months), from January 2007 onward. 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).

can hardly be distinguished apart in the sample presented. The largest difference appears 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 at approximating the uncertainty bands around $\widehat{f}_t$, i.e., in terms of accuracy, computational simplicity and speed.

Fig. 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 with a 95% nominal coverage. The fact that the bands

increase in width (and ultimately will converge to their unconditional width) is due to the fact that we construct the forecast at a time of low volatility levels. 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 magnitudes, 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 differences are 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

Fig. 3. Conditional volatility estimates from the GARCH model applied to the S&P500 monthly returns, with their corresponding out-of-sample conditional expectations and 95% forecast bands.

forecast band that stabilizes at a low-volatility level that is permanently higher than those of its two competitors. Our Monte Carlo result indicates that this may result in a poorer coverage.

7. Conclusions

We have discussed three different methods for computing in-sample confidence bands, and three different methods for producing 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 also reflect the uncertainty due to future innovations that determine the values of the unknown time-varying parameter via the future realizations of data. 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, and therefore easy to implement in standard software packages. This would provide useful additional information in empirical analyses, for which 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 the 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 a skewed Student's *t* distribution with time-varying parameters, as per Lucas, Schwaab, and Zhang (2014), or for dynamic discrete data models, as per Rydberg and Shephard (2003).

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*, 817–858.

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

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

Durbin, J., & Koopman, S. J. (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., & Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, *66*, 1127–1162.

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

Harvey, A. C. (2013). *Econometric society monographs, Dynamic models for volatility and heavy tails: with applications to financial and economic time series.* Cambridge: Cambridge University Press.

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

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

Rydberg, T. H., & Shephard, N. (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.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817–838.

Francisco Blasques is Assistant Professor of Econometrics at Vrije Universiteit Amsterdam and junior research fellow at Tinbergen Institute since 2012. He obtained his Ph.D. in 2011 from Maastricht University and he was also a Postdoc in the Department of Finance at Vrije Universiteit Amsterdam. His research interests include various topics in time series econometrics, simulation-based methods, sieve estimation and semi-nonparametric models.

Siem Jan Koopman is Professor of Econometrics at Vrije Universiteit Amsterdam and research fellow at Tinbergen Institute since 1999. He further is a regular Visiting Professor at CREATES, Aarhus University. His Ph.D. is from the London School of Economics (LSE) and dates back to 1992. He had earlier positions at the LSE and CentER, Tilburg University. His research interests include various topics in time series econometrics, financial econometrics and economic forecasting. In particular, his research is focussed on state space models and generalized autoregressive score models.

Katarzyna Łasak is Assistant Professor at the Department of Econometrics and Operations Research of Vrije Universiteit Amsterdam and a research fellow at Tinbergen Institute in the Netherlands. Her Ph.D. is from International Doctorate in Economic Analysis, IDEA Graduate Program, and European Network for Training in Economic Research, ENTER, Universitat Autònoma de Barcelona in Spain. Prior to her appointment at Vrije Universiteit Amsterdam, she was a Postdoctoral Fellow and later an International Fellow at the Center for Research in Econometric Analysis of Time Series, CREATES, Aarhus University. Her research interests include the statistical analysis of time series, financial econometrics and forecasting. She is particularly interested in developing econometrics tools to analyse non-stationarity, long memory, (fractional) co-integration, structural breaks and time varying parameters. Her theoretical developments in the field of time series econometrics have been published in *Journal of Econometrics* and *Journal of Business and Economic Statistics*.

André Lucas is Professor of Finance at Vrije Universiteit Amsterdam. He received his Ph.D. in Econometrics from Erasmus University Rotterdam and has published on financial and time series econometrics as well as risk management in journals such as the *Journal of Business and Economic Statistics, Journal of Econometrics* and *Review of Economics and Statistics*. Together with D. Creal and S.J. Koopman, he propagates the use of generalized autoregressive score dynamics for time varying parameter models. He received a five year VICI research grant for this project from the Dutch national research council (NWO).