Kless, Philipp Christian

Doctoral Thesis

New bootstrap methods for financial and economic time series

PhD Series, No. 197

Provided in Cooperation with:
University of Copenhagen, Department of Economics

Suggested Citation: Kless, Philipp Christian (2019) : New bootstrap methods for financial and economic time series, PhD Series, No. 197, University of Copenhagen, Department of Economics, Copenhagen

This Version is available at:
https://hdl.handle.net/10419/240546

FACULTY OF SOCIAL SCIENCES
Department of Economics
University of Copenhagen

PhD Thesis

Philipp Christian Kless

New Bootstrap Methods for Financial and Economic Time Series

Academic advisor: Anders Rahbek

Date of submission: 28/02/2019

Acknowledgments

There are numerous people who have helped me during the last three years in writing my thesis. Without them, I would have never made it.

First and foremost, I thank my supervisor, Anders Rahbek. His guidance and advice has always been outstanding. I am also thankful for all the discussions we had about econometrics over the years. Thank you for pushing me so far.

In addition, I want to say thank you to Rasmus Søndergaard Pedersen for many helpful conversations related to my research, and to Heino Bohn Nielsen for broadening my understanding of OxMetrics.

I would also like to thank the Department of Economics at the University of Copenhagen for funding my studies and providing a roof over my head. I am also grateful for my fellow PhD students who made life a lot easier. A special thanks goes to Anne, Nick, Marcus, and Simon.

My research visit at the Department of Economics at University of California, San Diego, was a truly great experience. I want to thank Graham Elliott and his staff at the department for making all this possible.

Lastly, I would like to thank my family and friends for their unconditional support through all these years. Especially, I say thank you to Helene and Ruth who never failed to make me smile after a day at the office.

Philipp Christian Kless
København
February 2019

Summary

The bootstrap is a promising simulation tool that can help to solve complicated statistical problems with no tractable solution. Specifically, the fundamental idea of the bootstrap is to use re-sampling methods to approximate otherwise unknown properties of an estimator.

This thesis investigates bootstrap methods for financial and economic time series to do forecasting. The results are presented in three self-contained parts which include theory, simulations, and empirics for the implemented bootstrap method.

In the first part, "*Smoothed Bootstrap Forecasts for Autoregressive Conditionally Heteroscedastic Models*", we consider forecasting a general class of ARCH(q) models using the smoothed bootstrap. Like the i.i.d. bootstrap, we first obtain the estimated residuals based on quasi maximum likelihood estimation. However, instead of drawing directly from the empirical distribution function of the residuals, as for the i.i.d. bootstrap, we draw bootstrap innovations from a kernel smoothed density. We provide a full asymptotic analysis which demonstrates the asymptotic validity of our forecasts based on the smoothed bootstrap. An important property of the smoothed bootstrap is that the bootstrap innovations have a well-behaved density. In particular, the proof of the bootstrap validity considerably simplifies when the bootstrap innovations possess such a density. We also perform a Monte Carlo experiment to investigate the finite sample performance of the smoothed bootstrap. We find that prediction intervals constructed from the smoothed bootstrap forecasts have the correct coverage in different stylized settings. Moreover, our simulations show that the smoothed bootstrap forecasts are robust to the choice of the kernel function and its bandwidth. Finally, a small empirical illustration confirms that our smoothed bootstrap also performs well with real data.

In the second part, "*Bootstrap Forecasts for the Poisson Autogressive Model*", we present a parametric bootstrap scheme to forecast the Poisson autoregressive (PAR) model. More precisely, our bootstrap simulates the analytically unknown multi-step ahead probability mass function (pmf) via recursive one-step ahead predictions of future counts using the Poisson distribution and the estimated model parameters. By repeating this step a large number of times, we numerically approximate the pmf of future counts. We prove that our bootstrap forecasts are asymptotically valid. We study the finite sample properties of our forecasts by means of a Monte Carlo experiment. This experiment supports our theoretical results, that is, prediction intervals based on the forecasted pmf according to our bootstrap have the correct coverage on average. Finally, in two empirical applications, we demonstrate that our parametric bootstrap also improves the forecasting performance of the PAR model for stock transaction data and monthly US default count data.

Lastly, in the third chapter, "*Estimation Uncertainty in GARCH Option Prices*", we investigate the impact of estimation uncertainty on GARCH option prices via two novel bootstrap algorithms. First, we design a bootstrap algorithm that allows us to assess

the impact of parameter uncertainty on GARCH option prices. For this bootstrap, we assume that the conditional distribution of the innovations is normal under the equivalent martingale measure, $\mathbb{Q}$. Next, we introduce estimation uncertainty by varying the GARCH model parameters in each bootstrap repetition while we keep the innovations fixed. In each bootstrap repetition, we vary the model parameters by drawing them from the asymptotic distribution of the proposed estimator. This design allows us to make sure that the only variation in the bootstrap stems from parameter uncertainty. By means of simulations, we find that the impact of estimation uncertainty is higher for options at the money. In addition, we introduce a second bootstrap which is more general. The second bootstrap allows for conditional distributions under $\mathbb{Q}$ with more skewness and heavier tails than the normal distribution. In an empirical application of the second bootstrap, we show that the uncertainty contained in option prices leads to variation in the metric used to evaluate the pricing performance of competing models. As a result, we find that no single GARCH specification dominates in terms of average relative pricing error when taking parameter uncertainty into account.

Danish Summary

Bootstrap metoden er et lovende simuleringsværktøj, der kan hjælpe med at løse komplicerede statistiske problemer der ikke har nogen teoretisk løsning. Den fundamentale ide med bootstrap er at genbruge data til at approksimere ellers ukendte egenskaber af estimatoren.

Denne afhandling undersøger bootstrap metoder til at lave forudsigelser for finansielle og økonomiske tidsserier. Resultaterne er præsenteret i tre uafhængige kapitler, som alle indeholder teori, simuleringer og empirisk test for den anvendte bootstrap metode.

I første kapitel, "*Smoothed Bootstrap Forecasts for Autoregressive Conditionally Heteroscedastic Models*", undersøger vi forudsigelser af en generel klasse af ARCH(q) modeller ved at bruge smoothed bootstrap. Ligesom ved i.i.d. bootstrap finder vi først de estimerede residualer ved hjælp af quasi maximum likelihood estimation. Fremfor at trække direkte fra den empiriske fordelingsfunktion af residualer som ved i.i.d bootstrap, trækker vi bootstrap innovationer fra en kernel smoothed tæthed. Vi præsenterer fuld asymptotisk teori, som demonstrerer den asymptotiske validitet af vores forudsigelser baseret på smoothed bootstrap. En vigtig egenskab ved vores smoothed bootstrap er at bootstrap innovationerne har en tæthedsfunktion, der opfører sig pænt. Beviset for validiteten af smoothed bootstrap simplificeres med denne tæthedsfunktion. Vi laver også en Monte Carlo simulation for at undersøge adfærden af smoothed bootstrap i små datasæt. Vi finder, at prædiktions-intervallerne som vi konstruerer med smoothed bootstrap har den korrekte præcision i en række stiliserede situationer. Derudover viser vores simulationer, at forudsigelserne fra smoothed bootstrap er robuste i forhold til valg af kernel Function og dets båndbredde. Endelig bekræfter en empirisk anvendelse, at vores smoothed bootstrap fungerer godt empirisk.

I anden kapitel, "*Bootstrap Forecasts for the Poisson Autogressive Model*", præsenterer vi en parametrisk bootstrap metode til at forudsige den velkendte Poisson Autoregressive (PAR) model. Mere præcist simulerer vi den analytisk ukendte fremtidige sandsynlighedsmassefunktion via rekursive et-trins forudsigelser hvor vi bruger Poissonfordelingen og de estimerede model parametre. Ved at gentage dette mange gange approksimerer vi numerisk den fremtidige sandsynlighedsmassefunktion. Vi viser at vores bootstrap forudsigelser er asymptotisk korrekte. Derudover studerer vi egenskaberne af vores forudsigelser med en Monte Carlo simulation. Simulationen understøtter det teoretiske resultat, idet intervallerne for forudsigelserne baseret på bootstrap gennemsnitligt er korrekte. Afslutningsvist, viser vi at vores paramestriske boostrap forbedrer forudsigelser af PAR modellen for data om aktiehandel og amerikanske konkurser.

I tredje kapitel, "*Estimation Uncertainty in GARCH Option Prices*", ved hjælp af to forskellige bootstrap metoder, undersøger vi betydningen af estimationsusikkerhed på optionspriser udregnet ved brug af GARCH modeller. Optionspriser er beregnet ved hjælp

af Monte Carlo simulationer i kombination med model parametre estimeret på historisk tidsrække data. At estimere optionspriser medfører estimationsusikkerhed i optionspriser. Vi præsenterer en bootstrap metode, som gør det muligt at evaluere parameterusikkerheden i GARCH optionspriser med normalfordelte innovationer under $\mathbb{Q}$. Vi tager højde for estimationsusikkerhed ved at bruge estimatorernes asymptotiske fordeling. I et numerisk eksperiment finder vi, at estimationsusikkerhed er højere for optioner at the money. Vi foreslår en anden bootstrap metode, der tillader en betinget fordeling med mere skævhed og tungere haler end normalfordelingen under $\mathbb{Q}$. I en empirisk anvendelse viser vi betydningen af parameterusikkerhed på optionspriser beregnet ved forskellige GARCH modeller. Derudover, viser vi at usikkerheden indeholdt i optionspriserne medfører variation i hvilket mål, der bruges til at evaluere performance af forskellige modeller. Derfor finder vi, at ingen GARCH model dominerer på alle parametre når man tager højde for estimationsusikkerhed.

Contents

Part I

Smoothed Bootstrap Forecasts for Autoregressive Conditionally Heteroscedastic Models

Smoothed Bootstrap Forecasts for Autoregressive Conditionally Heteroscedastic Models*

Philipp Christian Kless,
Department of Economics,
University of Copenhagen, Denmark.

Abstract

We consider forecasting a general class of ARCH(q) models using the smoothed bootstrap. Like the i.i.d. bootstrap, we first obtain the estimated residuals based on quasi maximum likelihood estimation. However, instead of drawing directly from the empirical distribution function of the residuals, as for the i.i.d. bootstrap, we draw bootstrap innovations from a kernel smoothed density. We provide a full asymptotic analysis which demonstrates the asymptotic validity of our forecasts based on the smoothed bootstrap. An important property of the smoothed bootstrap is that the bootstrap innovations have a well-behaved density. In particular, the proof of the bootstrap validity considerably simplifies when the bootstrap innovations possess such a density. We also perform a Monte Carlo experiment to investigate the finite sample performance of the smoothed bootstrap. We find that prediction intervals constructed from the smoothed bootstrap forecasts have the correct coverage in different stylized settings. Moreover, our simulations show that the smoothed bootstrap forecasts are robust to the choice of the kernel function and its bandwidth. Finally, a small empirical illustration confirms that our smoothed bootstrap also performs well with real data.

Keywords: ARCH; Smoothed Bootstrap; Forecasting

*I thank Anders Rahbek, Rasmus Søndergaard Pedersen, Heino Bohn Nielsen, and Brendan Beare for valuable comments. Part of this research was done while I was visiting University of California, San Diego, during Spring 2017. I thank Graham Elliott for his hospitality. This chapter pursues some ideas first discussed in my master's thesis. However, this chapter is completely re-written and contains substantial new research results.

1 Introduction

The h-day ahead conditional distribution of the return and volatility process of (G)ARCH type models is in general unknown. One possible way to obtain the conditional distributions is through bootstrap simulations. For instance, the standard i.i.d. bootstrap has been widely studied in empirical applications when it is used to forecast (G)ARCH models; cf. Robio (1999), Reeves (2005), Pascual, Romo, and Ruiz (2006), and Shimizu (2010), among others.

In this article, we study if the smoothed bootstrap can be an alternative to the standard i.i.d. bootstrap for forecasting (G)ARCH type models. For the i.i.d. bootstrap, we draw bootstrap innovations with replacement from the empirical distribution of the standardized residuals, while we draw from a kernel smoothed distribution for the smoothed bootstrap. The idea for the smoothed bootstrap dates back to Silverman and Young (1987) and has not been applied much in the forecasting literature.

More precisely, we explore the smoothed bootstrap and its application to dynamic volatility models when forecasting a general class of ARCH(q) models. This class includes, among others, asymmetric and threshold ARCH models. In regards of the smoothed bootstrap, we use a non-parametric density estimator to approximate the distribution of the model innovations. For this non-parametric estimator, we implement a kernel density estimator based on the residuals. Like the i.i.d. bootstrap, a smoothed bootstrap has the advantage that it avoids any particular parametric assumption on the innovation distribution when forecasting.

We further contribute to the literature by providing a full asymptotic analysis of the smoothed bootstrap when forecasting with ARCH type models. First, we show that our proposed kernel estimator based on the residuals is consistent. Second, we then prove that the in-sample bootstrap process is geometric ergodic. Third, we show that the bootstrap parameters are consistently estimated by QML on the in-sample bootstrap data. Finally, we establish that our bootstrap simulations based on recursive one-step ahead forecasts mimic the unknown h-day ahead conditional distribution of the return and volatility series.

The bootstrap has been often studied in the existing forecast literature. Pascual, Romo, and Ruiz (2006), henceforth PRR, propose an i.i.d. bootstrap scheme for a GARCH(1,1) model to simulate h-step ahead prediction intervals that also incorporate an additional component of uncertainty due to parameter estimation. In particular, they show through a Monte Carlo experiment that their prediction intervals for the return and volatility series of their model have the correct coverage. Trucíos and Hotta (2016) adapt the bootstrap algorithm of PRR to obtain prediction intervals for EGARCH and GJR-GARCH models. Chen, Gel, Balakrishna, and Abraham (2011) exploit the linear representation of the GARCH process to utilize the sieve bootstrap to construct predictions intervals. Their

approach has the advantage that computational costs are reduced by up to 100 times compared to other bootstrap procedures for the GARCH model. To the best of our knowledge, Reeves (2005) is the only contribution also providing an asymptotic analysis of the i.i.d. bootstrap used to generate prediction intervals for a AR(1)-GARCH(1,1) model. In contrast to Reeves (2005), our approach is based on a smoothed bootstrap for a general class of ARCH models. At the same time, the bootstrap gets more and more attention in the forecasting literature because it can simulate the complete probability distribution of possible outcomes which is helpful in making economic decision; see, among others, Aastveit, Foroni, and Ravazzolo (2017), and Tay and Wallis (2000). For instance, describing the uncertainty of volatility forecasts is highly relevant for trading and pricing volatility derivatives; cf. Vorbrink (2014).

To present numerical evidence of the performance of our smoothed bootstrap device, we conduct a Monte Carlo experiment in which we calculate prediction intervals based on our bootstrap forecasts. The results demonstrate that our device delivers prediction intervals with correct coverage on average in different stylized settings. Moreover, we show that the numerical results for the smoothed bootstrap are robust to different choices of the kernel function and its bandwidth. Finally, in a short empirical illustration, we demonstrate that our smoothed bootstrap performs at par with the standard i.i.d. bootstrap of PRR(2006).

This article is organized as follows: In Section 2, we introduce the model, its main assumptions and some standard asymptotic results for the QML estimator. Next, in Section 3, we provide details about the smoothed bootstrap forecasts. In Section 4, we give a full asymptotic theory for the smoothed bootstrap. Section 5 reports the results of our Monte Carlo experiment. Section 6 considers different kernel functions and corresponding bandwidths. Next, Section 7 contains a small empirical illustration while we conclude in Section 8. Additional simulations, empirical results along with all proofs can be found in the Appendix.

NOTATION: We use P^* and E^* respectively, to denote the probability and expectation, conditional on the original sample. With $\xrightarrow{w*}_p$ we denote weak convergence in probability; that is $Y_T^* \xrightarrow{w^*}_p Y$ means that, as the sample size T diverges, $|P^*(Y_T^* \leq x) - P(Y \leq x)| \xrightarrow{p} 0$ at all continuity points. Moreover, for a given sequence Y_T^* computed from the bootstrap data, $Y_T^* - Y = o_p^*(1)$, in probability, or $Y_T^* \xrightarrow{p^*}_p Y$, means that for any $\epsilon > 0$, $P^*(||Y_T^* - Y|| > \epsilon) \xrightarrow{p} 0$ as $T \to \infty$. Similarly, $Y_T^* = O_p^*(1)$, in probability, means that, for every $\epsilon > 0$, there exists a constant $M > 0$ such that, for all large T, $P(P^* ||Y_T^*|| > M) < \epsilon)$.

2 Model, Assumptions and Estimation

We consider a general class of ARCH models as in Kristensen and Rahbek (2005). That is, assume that the observed process $(Y_t, X_{t-1})_{t=1}^T$ with $Y_t \in \mathbb{R}$ and $X_{t-1} \in \mathbb{R}^m$ is described

by the following ARCH model,

$$\begin{aligned} Y_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \alpha' X_{t-1}, \end{aligned} \tag{2.1}$$

where $X_0 \coloneqq (x_0, \ldots, x_{1-q})$ is fixed, $(z_t)_{t=1}^T$ is *i.i.d.* $(0,1)$, $\omega > 0$, $\alpha \in \mathbb{R}_+$ and $X_t \coloneqq h(Y_t, \ldots, Y_{t-q+1}) \in \mathbb{R}_+^m$ is a measurable function of $(Y_t, \ldots, Y_{t-q+1})$ for some finite, fixed $q \geq 1$. The parameters of the models are given by $\theta \coloneqq (\omega, \alpha') \in \Theta \subset \mathbb{R}_+^{m+1}$.

Possible parameterizations for $\alpha' X_{t-1}$ that fall into our class are the standard linear ARCH(q) model given by $\alpha' X_{t-1} = \sum_{i=1}^q \alpha_i Y_{t-i}^2$; see Engle (1982). Another important example is the GJR-ARCH(q) model of Glosten, Jagannathan, and Runkle (1993) which can be expressed as $\alpha' X_{t-1} = \sum_{i=1}^q \alpha_{1i} \mathbb{I}_{(Y_{t-i}<0)} Y_{t-i}^2 + \alpha_{2i} \mathbb{I}_{(Y_{t-i}\geq 0)} Y_{t-i}^2$. More parametrization, which are part of our framework in eq. (2.1), are listed in Table 1 in Kristensen and Rahbek (2005); see also Section 5 in Kristensen and Rahbek (2005).

Using eq. (2.1), the likelihood function for (G)ARCH models is readily available, and therefore θ can be directly estimated via Gaussian QMLE. That is, we define the Gaussian QML estimator by

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min}\, L_T(\theta), \tag{2.2}$$

where $L_T(\theta) = \frac{1}{T} \sum_{t=1}^T \log(\sigma_t^2(\theta)) + \frac{Y_t^2}{\sigma_t^2(\theta)}$ with $\sigma_t^2(\theta) = \omega + \alpha' X_{t-1}$.

The asymptotic properties of $\hat{\theta}$ are well-understood in the literature. Here, we follow Kristensen and Rahbek (2005) and assume that the following conditions for $(Y_t, X_{t-1})_{t=1}^T$ and $(z_t)_{t=1}^T$ hold throughout the remainder of our paper.

Assumption 2.1 *(i)* (Y_t, X_{t-1}) *is geometric ergodic such that for the true parameter* $\theta_0 \in \Theta$ *a stationary solution exists with (ii)* $E[|\log(X_{i,t})|] < \infty$, $1 \leq i \leq m$.

Assumption 2.2 *The i.i.d.* $(0,1)$ *random variables* z_t *have a symmetric Lebesgue density* p_z *which is everywhere positive and Lipschitz continuous on* $\mathbb{R}$. *Moreover,* $E\left[(z_t^2 - 1)^2\right] < \infty$.

In regard of the parameter space, Θ, and to ensure identification of the model parameters, we impose the next assumption.

Assumption 2.3 *(i) The parameter space is given by* $\Theta = [\underline{\omega}, \overline{\omega}] \times [\underline{\alpha}_i, \overline{\alpha}_i] \times \ldots \times [\underline{\alpha}_m, \overline{\alpha}_m]$ *for some* $0 < \underline{\omega} < \overline{\omega} < \infty$ *and* $0 < \underline{\alpha}_i < \overline{\alpha}_i < \infty$ *for* $i = 1, \ldots, m$. *(ii) for any* $\gamma \in \mathbb{R}^m \backslash \{0\}$ *and* $g \in \mathbb{R}$, $P(\gamma' X_{t-1} \neq g) > 0$.

Given Assumption 2.1, the strong LLN applies to (Y_t, X_{t-1}) which is needed for the asymptotic analysis. Later, we state explicit conditions such that Assumption 2.1 holds

for different choices of $\alpha' X_{t-1}$ in eq. (2.1). Assumption 2.2 is slightly stronger compared to Kristensen and Rahbek (2005). We require in addition that the innovations have Lipschitz continuous density. This slightly stronger assumption is imposed because it allows us to apply Markov Chain theory to establish geometric ergodicity of the in-sample bootstrap ARCH(q) process. Finally, Assumption 2.3 contains standard conditions which, in (i), define the parameter space and, in (ii), ensure identification.

Under the above assumptions, it follows by Kristensen and Rahbek (2005) that

$$\sqrt{T}\left(\hat{\theta}-\theta_0\right) = O_P(1) \tag{2.3}$$

if θ_0 is in the interior of Θ. This result will be used in Section 4 to demonstrate the validity of the smoothed bootstrap. Note that it is the rate of convergence we use to show our results and not the actual limiting distribution.

3 Smoothed Bootstrap Forecasts

In the following, we are interested in the h-step ahead conditional distribution of the return process, $P(Y_h \leq x|\mathcal{F}_T)$, where $\mathcal{F}_T$ contains all information up to and including T; respectively, the h-step ahead conditional distribution of the volatility process, $P(\sigma_h^2 \leq x|\mathcal{F}_T)$ for a forecasting horizon $h > T$. In general, these two expressions are analytically unknown. One way to estimate them is through bootstrap based simulations.

For instance, we can generate B bootstrap replicates of future return and volatility paths and use their empirical distribution functions to estimate the unknown h-step ahead conditional distributions. One popular method to generate these replicates is the standard i.i.d. bootstrap. More precisely, for the i.i.d. bootstrap, we sample with replacement from the empirical distribution function of the centered residuals to generate bootstrap innovations to simulate the h-step ahead conditional distributions. That is, we estimate the centered residuals as

$$\hat{z}_{t,c} \;=\; \hat{z}_t - \hat{z}, \tag{3.1}$$

where $\hat{z}_t = Y_t/\left(\hat{\omega}+\hat{\alpha}' X_{t-1}\right)$ and $\hat{z} = T^{-1}\sum_{t=1}^{T}\hat{z}_t$. Then we sample bootstrap innovations, z_t^*, as $z_t^* \sim$ i.i.d.$\hat{F}_T$, conditional on the original sample, where $\hat{F}_T$ is the empirical distribution function of the centered residuals.

The smoothed bootstrap we explore in this article is different in one important aspect. More precisely, after having obtained the centered residuals, we additionally use a kernel density estimator,

$$\hat{p}_{\hat{z}}(y) = \frac{1}{Tu}\sum_{t=1}^{T} K\left(\frac{y-\hat{z}_{t,c}}{u}\right), \tag{3.2}$$

where $K(\cdot)$ is a kernel function and u its bandwidth, to estimate the probability density function of z_t, $p_z(y)$. Then, we use $z_t^* \sim$ i.i.d.$\hat{p}_{\hat{z}}(y)$, conditional on the data, to generate our bootstrap replications to simulate the unknown conditional distributions. As we discuss later, drawing from $\hat{p}_{\hat{z}}(y)$ has several implications on the bootstrap data generating process. Most importantly, it simplifies key steps in the proofs of bootstrap validity.

In our paper, the smoothed bootstrap is combined with the bootstrap scheme in PRR. This scheme is attractive mainly because of two aspects. First, it is independent of the distribution of the model innovations since we draw bootstrap innovations from $\hat{p}_{\hat{z}}(y)$. Second, bootstrap forecasts also reflect parameter uncertainty since the bootstrap parameters are re-estimated in each bootstrap repetition. Both aspects are crucial for generating prediction intervals with correct coverage especially in cases where the sample is small or the error term follows a non-normal distribution; see also, Thombs and Schucany (1990), and Blasques, Koopman, Lasak, and Lucas (2015).

The smoothed bootstrap generates forecasts through the following algorithm:

Smoothed Bootstrap Algorithm:

(i) Obtain $\hat{\theta} = (\hat{\omega}, \hat{\alpha}')$ in eq. (2.1) on the original sample, $(Y_t)_{t=1}^T$, using Gaussian QML estimation.

(ii) Compute the centered residuals, $\hat{z}_{t,c}$, as described in eq. (3.1).

(iii) Estimate the probability density function, $\hat{p}_{\hat{z}}(y)$, as given in eq. (3.2).

(iv) Construct the in-sample bootstrap sample values, $(Y_t^*, \sigma_t^{*2})_{t=1}^T$, recursively from

$$\sigma_t^{*2} = \hat{\omega} + \hat{\alpha}' \hat{X}_{t-1}^*, \; Y_t^* = z_t^* \sigma_t^* \tag{3.3}$$

initialized at $X_0^* = (X_{m-1}, \ldots, X_0)'$, and with T bootstrap errors, $(z_t^*)_{t=1}^T$, drawn i.i.d. from $\hat{p}_{\hat{z}}(y)$.

(v) Obtain $\hat{\theta}^* = \left(\hat{\omega}^*, \hat{\alpha}^{*\prime}\right)$ in eq. (2.1) on the bootstrap sample, $(Y_t^*)_{t=1}^T$, using Gaussian QML estimation.

(vi) Construct the bootstrap out-of-sample values, $(Y_h^*, \sigma_h^{*2})_{h=1}^H$, recursively from

$$\sigma_h^{*2} = \hat{\omega}^* + \hat{\alpha}^{*\prime} X_{h-1}^*, \; Y_h^* = z_h^* \sigma_h^* \tag{3.4}$$

initialized at $X_0^* = (X_{T-j}, \ldots, X_T)'$, $j = m-1, \ldots, 0$ and with H bootstrap errors, $(z_h^*)_{h=1}^H$ drawn i.i.d. from $\hat{p}_{\hat{z}}(y)$.

(vii) Repeat step (iv) to (vi) and obtain a set of bootstrap replicates, $\left(Y_h^{*(b)}, \sigma_h^{*2(b)}\right)_{h=1}^H$, for $b = 1, \ldots, B$.

(viii) Calculate $P^*\left(Y_h^* \leq x | X_0^* = y\right) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(Y_h^{*(b)} \leq x\right)$ and $P^*\left(\sigma_h^{*2} \leq x | X_0^* = y\right) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(\sigma_h^{*2(b)} \leq x\right)$ for $x \in \mathbb{R}$.

Remark 3.1 *Step (i), (ii), and (v) are similar to the approach of PRR. In the first two steps, we estimate the model parameters; and then we obtain the centered residuals. In step (v), we re-estimate the bootstrap parameters in each bootstrap repetition and thereby allow our forecasts to reflect parameter uncertainty.*

Remark 3.2 *In step (iii), we introduce a non-parametric kernel estimator such that we can draw bootstrap innovations from $\hat{p}_{\hat{z}}(y)$ instead of $\hat{F}_T$. Assumptions for the kernel $K(\cdot)$ and its bandwidth u will be detailed in Section 4.*

4 Asymptotic Properties of the Smoothed Bootstrap

In this section we prove that our forecasts, based on a smoothed bootstrap, mimic the correct out-of-sample distribution of the return and volatility series in the limit.

The following discussion uses the ARCH(1) model of Engle (1982) to ease notation and focus on the asymptotic arguments; thus, we set $\alpha' X_{t-1} = \alpha Y_{t-1}^2$ in eq. (2.1). Then, according to our bootstrap algorithm the parameter estimates are obtained by Gaussian QML estimation and defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min}\, L_T(\theta), \tag{4.1}$$

where $L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log\left(\sigma_t^2(\theta)\right) + \frac{Y_t^2}{\sigma_t^2(\theta)}$ with $\sigma_t^2(\theta) = \omega + \alpha Y_{t-1}^2$. In addition, Y_0 is given as initial value. Let $\theta_0 = (\omega_0, \alpha_0)'$ denote the true parameter values.

Corollary 4.1 *Set $\alpha' X_{t-1} = \alpha Y_{t-1}^2$ in eq. (2.1) and suppose that $\alpha < 1$ such that Assumption 2.1 holds. Further, suppose that Assumption 2.2 and Assumption 2.3 hold. Then, $\sqrt{T}\left(\hat{\theta} - \theta_0\right) = O_P(1)$.*

Before discussing asymptotic theory for our bootstrap, we impose assumptions on the kernel estimator, $\hat{p}_{\hat{z}}(y)$, in eq. (3.2).

Assumption 4.1 *Assume for the kernel $K(\cdot)$ and its bandwidth $u > 0$ in eq. (3.2) that*

(a) $K(\cdot)$ is a measurable function $K(y)$, $y \in \mathbb{R}$, with (i) $K(\cdot)$ symmetric, positive and Lipschitz continuous, (ii) $\int K(y)\,dy = 1$, (iii) $\int |K(y)|\,dy < \infty$, (iv) $|y|\,|K(y)| \to 0$ as $|y| \to \infty$, (v) $\sup_{y \in \mathbb{R}} |K(y)| < \infty$, (vi) $\int K^2(y)\,dy < \infty$, and (vii) $\int |y|^r K(y)\,dy < \infty$ for $r \geq 2$.

(b) As $T \to \infty$, (i) $u \to 0$, and (ii) $T^{1/2-\epsilon}u^2 \to \infty$ for some small $\epsilon > 0$.

The assumptions for $K(\cdot)$ in part (a) are standard for non-parametric density estimation; e.g. see Pagan and Ullah (1999), Chapter 2. For instance, consider Assumption 4.1(a) (ii) which ensures that the kernel function is a probability density function. Most of the standard choices for a kernel function, such as the Gaussian kernel, fulfill Assumption 4.1(a). Observe that in Assumption 4.1(b) (ii) we require a slower rate of convergence for the bandwidth u compared to the standard i.i.d. case. A slower rate is needed because the kernel estimator is based on estimated residuals; see also Gao and Song (2008).

The next two results establish conditions which we use to show that the in-sample bootstrap process, $(Y_t^*)_{t=1}^T$, is geometrically ergodic, conditional on the data.

Lemma 4.1 *Suppose that $\alpha_0 < 1$ and that Assumption 4.1 holds. Then,*

$$\sup_{y \in \mathbb{R}} |\hat{p}_{\hat{z}}(y) - p_z(y)| = o_p\left(T^{-\epsilon}\right) \text{ for some } \epsilon > 0.$$

According to the result in the above lemma, $\hat{p}_{\hat{z}}$ is uniform consistent estimator for $p_z(y)$ when it is based on a non-parametric kernel density estimator. Note that this result ensures that we can show that the in-sample bootstrap process is geometric ergodic; see also Franke, Kreiss, Mammen, and Neumann (2002) and Franke, Neumann, and Stockis (2004).

Lemma 4.2 *Suppose that $\alpha_0 < 1$ and that Assumption 4.1 holds. Moreover, there exists an appropriate sequence of sets $\Omega_T \subseteq \mathbb{R}^{T+1}$ with $P((Y_0, \ldots, Y_T) \notin \Omega_T) = o(1)$ for $T \to \infty$ such that for $(Y_0, \ldots, Y_T) \in \Omega_T$,*

$$\hat{\sigma}(y) E^*[|z_t^*|] \leq C_1 + C_2|y| \text{ for all } y \in \mathbb{R} \text{ and some } C_1 < \infty, C_2 < 1,$$

where $\hat{\sigma}(y) = (\hat{\omega} + \hat{\alpha} y^2)^{1/2}$.

Remark 4.1 *C_1 and C_2 in Lemma 4.2 are functions of $(Y_0, \ldots, Y_T)$ such that the restrictions $C_1 < \infty$; respectively $C_2 < 1$ hold with probability tending to one.*

The result in the previous lemma ensures, along with the result in Lemma 4.1, that the in-sample bootstrap process is geometric ergodic, conditional on the data. We state this result next.

Lemma 4.3 *Suppose that $\alpha_0 < 1$ and that Assumption 4.1 holds. Then, $(Y_t^*)_{t=1}^T$ is geometrically ergodic and β-mixing, conditional on the original sample, $(Y_0, \ldots, Y_T) \in \Omega_T$.*

Remark 4.2 *Lemma 4.3 complements the findings in Theorem 2 of Franke, Kreiss, and Mammen (2002); respectively in Theorem 2 of Franke, Neumann, and Stockis (2004). Both papers show this result for their non-parametric bootstrap process.*

The proposed bootstrap is only relevant in our setting if it mimics properties of the original process. More precisely, we want that the stationary distribution of $(Y_t^*)_{t=1}^T$ is close to that of $(Y_t)_{t=1}^T$. This result requires that the estimators $\hat{p}_{\hat{z}}(y)$ and $\hat{\sigma}(y)$ are consistent. Specifically, we need that $\hat{\sigma}(y)$ converges uniformly in y. To ensure uniform convergence, we restrict the estimator $\hat{\sigma}$ to regions where the stationary density is above some threshold, and hence enough observations are available.

Before formalizing the closeness between the stationary distributions of $(Y_t)_{t=1}^T$ and $(Y_t^*)_{t=1}^T$, we describe these additional properties for $\hat{p}_{\hat{z}}(y)$ and $\hat{\sigma}(y)$.

Lemma 4.4 *Suppose that $\alpha_0 < 1$ and that Assumption 4.1 holds. Let $\{\mathcal{Y}_T\}$ be a sequence of sets with $\ldots \subseteq \mathcal{Y}_T \subseteq \mathcal{Y}_{T+1} \subseteq \ldots$ with $\mathcal{Y}_T = \left[-T^\delta, T^\delta\right]$ for some $\delta > 0$. Then*

(a) $P(Y_t \in \mathcal{Y}_T^c) = O(T^{-\nu})$ *for some* $\nu > 0$.

Moreover,

(b) $\sup_{y \in \mathcal{Y}_T} |\hat{\sigma}(y) - \sigma(y)| = O_p(T^{-\nu})$,

(c) $\inf_{y \in \mathbb{R}} \hat{\sigma}(y) \geq C > 0$,

(d) $\int |\hat{p}_{\hat{z}}(y) - p_z(y)|\, dy = O_p(T^{-\nu})$.

Furthermore, we need some additional mild conditions on $p_z(y)$ for the next result; see Franke, Neumann, and Stockis (2004).

Assumption 4.2 *The density $p_z(y)$ has the following properties:*

(a) $\int |p_z(y) - p_z(y+a)|\, dy = O(a)$,

(b) $\int |p_z(y) - p_z(y/(1+a))|\, dy = O(a)$.

Next, we formalize the result for the stationary distribution of the in-sample bootstrap process, π_T^*, and the original process, π.

Theorem 4.1 *Suppose that $\alpha_0 < 1$ and that Assumption 4.1 and Assumption 4.2 hold. Then, for $\nu > 0$ and arbitrarily large $\Lambda < \infty$, it holds that*

$$\sup_B \left\{ \left(\lambda(B)\, T^{-\nu} + T^{-\Lambda} \right)^{-1} |\pi(B) - \pi_T^*(B)| \right\} \leq C,$$

conditional on the original sample, $(Y_0, \ldots, Y_T) \in \Omega_T$, where $\lambda(\cdot)$ denotes the Lebesgue measure and B is a measurable subset of $\mathbb{R}$.

Remark 4.3 *The above theorem implies that π_T^* approximates π with a rate equal to some power of T^{-1}. This approximation is used to prove the validity of in-sample confidence bands based on the bootstrap; see Franke, Neumann, and Stockis (2004).*

Before stating the result which describes the asymptotic behavior of the parameters estimated on the bootstrap data, we describe the properties of the bootstrap residuals.

Lemma 4.5 *Suppose that the assumptions of Theorem 4.1 hold. Then,* $E^*[z_t^*] = o_p(1)$, $E^*[z_t^{*2}] = 1+o_p(1)$, *and* $E^*[z_t^{*4}] = O_p(1)$, *conditional on the original sample,* $(Y_0, \ldots, Y_T) \in \Omega_T$. *Moreover, it holds that*

$$|P^*(z_h^* \leq x) - P(z_h \leq x)| \xrightarrow{p} 0.$$

Next, we define the Gaussian bootstrap QML estimator of the ARCH(1) model in eq. (3.3) by

$$\hat{\theta}^* = \underset{\theta \in \Theta}{\arg\min}\, L_T^*(\theta), \tag{4.2}$$

where $L_T^*(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log(\sigma_t^{*2}(\theta)) + \frac{Y_t^{*2}}{\sigma_t^{*2}(\theta)}$ with $\sigma_t^{*2}(\theta) = \omega + \alpha Y_{t-1}^{*2}$. In addition, Y_0 is given as initial value. Then, as a consequence of Lemma 4.5 and Lemma 4.1, we formalize the asymptotic properties of $\hat{\theta}^*$ in the next lemma.

Lemma 4.6 *Suppose that the assumptions of Theorem 4.1 hold. Then,* $\sqrt{T}\left(\hat{\theta}^* - \theta_0\right) = O_P^*(1)$, *conditional on the original sample, for* $(Y_0, \ldots, Y_T) \in \Omega_T$.

Finally, we are ready to state the main result of this paper in the following theorem.

Theorem 4.2 *Suppose that the assumptions of Theorem 4.1 hold and let* $(Y_h^*, \sigma_h^{*2})_{h=1}^H$ *be generated as detailed in the algorithm given in Section 3. Then, conditional on the original sample,* $(Y_0, \ldots, Y_T) \in \Omega_T$, *it holds that*

$$|P^*(Y_h^* \leq x | Y_0^* = y) - P(Y_h \leq x | Y_0 = y)| \xrightarrow{p} 0.$$

Moreover,

$$\left|P^*\left(\sigma_h^{*2} \leq x | Y_0^* = y\right) - P\left(\sigma_h^2 \leq x | Y_0 = y\right)\right| \xrightarrow{p} 0.$$

Remark 4.4 *In practice, we use* $Y_0 = Y_0^* = Y_T$ *to initialize the algorithm. This choice of fixing the last observation in each bootstrap replications also ensure that the bootstrap forecasts are conditional on the original sample.*

Remark 4.5 *We conjecture that the proposed smoothed bootstrap as well as the theory can be extended to include GARCH(p,q)-type models. This analysis, however, is beyond the scope of this article and left for future research.*

5 Finite sample properties

We investigate the finite sample properties of our smoothed bootstrap forecasts by means of a Monte Carlo experiment. In the following, we consider the linear ARCH(1) model of the general ARCH(q) model class. Moreover, we consider two different assumptions for the z_t's: (i) $z_t \sim i.i.d.N(0,1)$, and (ii) $z_t \sim i.i.d.t_5(0,1)$, where $t_5(0,1)$ is the Student's t distribution with 5 degrees of freedom, mean zero and variance one. The ARCH parameter α takes values in $\{0.5, 0.8\}$ while $\omega = 0.1$. The sample size is $T \in \{500, 1000\}$ and the forecasting horizon, H, takes values in $\{1, 2, 5, 10, 20\}$. We implement a Gaussian kernel for the smoothed bootstrap with a bandwidth that minimizes an L^2 distance between the density estimate and the true density as proposed in Silverman (1986):

$$\begin{aligned} K(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \ -\infty < x < \infty, \\ u &= 1.03 \hat{\tau} T^{-0.2}, \end{aligned}$$

where $\hat{\tau}$ is the empirical standard deviation of the variables $\hat{z}_{t,c}$. Implementing a Gaussian kernel implies that we can generate the smoothed bootstrap innovations, z_t^*, using following approach:

$$z_t^* = z_t^\dagger + u\eta_t, \tag{5.1}$$

where $z_t^\dagger \sim i.i.d.\hat{F}_T$, u is the bandwidth as before, and $\eta_t \sim i.i.d.N(0,1)$; see Silverman (1986).

We verify the performance of our smoothed bootstrap by checking if the forecasts are correctly calibrated. Correct calibration means that the density forecasts coincides with the true density of the predicted variable; cf. Aastveit, Foroni, and Ravazzolo (2017). We study correct calibration by comparing the actual coverage of prediction intervals for the out-of-sample return and volatility process based on our smoothed bootstrap forecasts with a chosen nominal coverage level. That is, we repeat the following steps:

1. Generate a time series $(Y_t)_{t=1}^{T+H}$ using the ARCH(1) model.
2. Use the Smoothed Bootstrap Algorithm to forecast the conditional distribution of the return and volatility process from Step 1.
3. Based on the forecast from Step 2, construct prediction intervals with nominal coverage of 95%.

For each repetition we calculate the coverage and length of our bootstrapped prediction interval, where the length of a prediction interval is defined as the difference between the 97.5% and 2.5% quantile. We use $B = 999$ to obtain the forecasts of the distributions.

We repeat the above three steps by M times and then compute the mean and standard errors of our measures. In addition, we also compare our smoothed bootstrap forecasts to the i.i.d. bootstrap forecasts of PRR. In this case, the bootstrap innovations are i.i.d. draws from $\hat{F}_T$.

The results of our Monte Carlo study are summarized by Table 5.1 and 5.2. We make the following observations: First, we note that both bootstrap procedures perform equally well, that is, both procedures deliver the correct coverage on average. This observation is closely related to innovation uncertainty and the similar performance of both bootstrap procedures is best explained by Figure 5.1. This graph shows a QQ-plot of a random sample of z_t^* with size $T = 500$ obtained via smoothed bootstrap; respectively i.i.d. bootstrap where the assumed distribution is $t_5(0, 1)$. We observe that both bootstrap methods approximate the Student's t distribution equally well which is one reason for the similar performance in terms of average coverage. Hence, in practice, it seems that it does not matter if we use the smoothed or i.i.d. bootstrap. The same observations is also made in Franke, Neumann, and Stockis (2004). Moreover, both bootstrap algorithms also incorporate estimation uncertainty which is another explanation for their good performance; see Pascual, Romo, and Ruiz (2006) and Blasques, Koopman, Lasak, and Lucas (2015), among others. Second, we see in Table 5.1 that the actual coverage of the bootstrap intervals for the one-step ahead conditional volatility process is slightly below the nominal coverage level when we consider the Student's t distribution. At this short horizon the only variation in σ_1^* is due to $\hat{\theta}^*$ since Y_T is fixed in each bootstrap repetition. This source of randomness is apparently not sufficient to model the conditional distribution of σ_1 appropriately. Then, as the forecast horizon increases and innovation uncertainty becomes relatively more important than parameter uncertainty, the bootstrap prediction intervals yield the correct coverage on average. The tables where $\alpha = 0.8$ allow for similar observations and are relegated to Appendix C.

As an alternative, we also consider the situation where the data generating process is a GJR-ARCH(1) process with $(\omega, \alpha, \gamma) = (0.1, 0.1, 0.35)$. As before, we let $T \in \{500, 1000\}$ and z_t be $N(0, 1)$. The properties of both bootstrap procedures are qualitatively the same as for the ARCH(1) case; see Appendix C.

6 Sensitivity of the smoothed bootstrap to the choice of the kernel and bandwidth

In the previous Monte Carlo experiment we used a Gaussian kernel with a plug-in bandwidth. A natural question is how the form of the kernel function and the choice of the bandwidth affect the smoothed bootstrap.

Properties of the kernel density estimate depend on the smoothness of the underlying true density. From non-parametric density estimation it is known that the Epanechnikov

Table 5.1: Prediction intervals for the volatility process with 95% coverage level, where $\alpha = 0.5$. For the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Method		h				
				1	2	5	10	20
N(0,1)	500	PRR	Coverage	94.700 (0.224)	94.050 (0.169)	94.160 (0.127)	94.320 (0.106)	94.295 (0.095)
			Above	3.600 (0.186)	3.300 (0.130)	3.340 (0.090)	3.320 (0.073)	3.285 (0.058)
			Below	1.700 (0.129)	2.650 (0.114)	2.500 (0.094)	2.360 (0.079)	2.420 (0.072)
			Length	0.085 (0.176)	0.054 (0.854)	0.063 (0.857)	0.062 (0.583)	0.060 (0.337)
		Smoothed		94.900 (0.220)	94.050 (0.169)	94.040 (0.129)	94.200 (0.110)	94.175 (0.099)
				3.200 (0.176)	2.900 (0.119)	3.040 (0.085)	2.980 (0.070)	3.000 (0.057)
				1.900 (0.137)	3.050 (0.126)	2.920 (0.103)	2.820 (0.087)	2.825 (0.079)
				0.086 (0.181)	0.567 (0.925)	0.663 (0.871)	0.661 (0.631)	0.642 (0.432)
	1000	PRR		93.000 (0.255)	94.000 (0.171)	94.780 (0.115)	94.860 (0.093)	94.650 (0.085)
				4.800 (0.214)	3.800 (0.132)	2.940 (0.081)	2.820 (0.062)	3.010 (0.051)
				2.200 (0.147)	2.200 (0.116)	2.280 (0.088)	2.320 (0.075)	2.340 (0.069)
				0.054 (0.083)	0.050 (0.485)	0.059 (0.417)	0.058 (0.264)	0.057 (0.182)
		Smoothed		93.300 (0.250)	94.150 (0.171)	94.560 (0.120)	94.630 (0.096)	94.455 (0.088)
				4.100 (0.198)	3.350 (0.125)	2.880 (0.083)	2.740 (0.060)	2.840 (0.050)
				2.600 (0.159)	2.500 (0.124)	2.560 (0.093)	2.630 (0.080)	2.705 (0.074)
				0.055 (0.084)	0.514 (0.511)	0.602 (0.429)	0.598 (0.264)	0.593 (0.190)
$t_5(0,1)$	500	PRR		90.500 (0.293)	92.450 (0.199)	93.500 (0.143)	93.980 (0.116)	94.255 (0.100)
				8.300 (0.276)	5.850 (0.173)	4.460 (0.110)	4.030 (0.078)	3.800 (0.062)
				1.200 (0.109)	1.700 (0.103)	2.040 (0.096)	1.990 (0.086)	1.945 (0.078)
				0.115 (0.218)	0.630 (1.027)	0.704 (0.977)	0.696 (0.720)	0.684 (0.534)
		Smoothed		91.000 (0.286)	92.700 (0.197)	93.760 (0.142)	94.020 (0.118)	94.140 (0.105)
				7.100 (0.257)	5.050 (0.162)	3.880 (0.104)	3.590 (0.076)	3.510 (0.061)
				1.900 (0.137)	2.250 (0.117)	2.360 (0.102)	2.390 (0.093)	2.350 (0.085)
				0.115 (0.219)	0.644 (1.083)	0.715 (0.948)	0.727 (0.789)	0.713 (0.602)
	1000	PRR		91.000 (0.286)	93.350 (0.180)	94.680 (0.120)	94.830 (0.103)	95.020 (0.091)
				6.700 (0.250)	4.700 (0.148)	3.360 (0.088)	3.060 (0.066)	2.895 (0.048)
				2.300 (0.150)	1.950 (0.111)	1.960 (0.086)	2.110 (0.084)	2.085 (0.079)
				0.091 (0.215)	0.623 (1.103)	0.702 (0.953)	0.691 (0.744)	0.670 (0.419)
		Smoothed		91.000 (0.286)	93.150 (0.180)	94.560 (0.120)	94.770 (0.102)	94.935 (0.090)
				6.600 (0.248)	4.600 (0.146)	3.260 (0.085)	3.020 (0.066)	2.850 (0.047)
				2.400 (0.153)	2.250 (0.115)	2.180 (0.089)	2.210 (0.083)	2.215 (0.080)
				0.091 (0.215)	0.646 (1.271)	0.712 (1.045)	0.693 (0.628)	0.678 (0.456)

Table 5.2: Prediction intervals for the return process with 95% coverage level, where $\alpha = 0.5$. For the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Method		h				
				1	2	5	10	20
N(0,1)	500	PRR	Coverage	94.700 (0.224)	95.000 (0.161)	94.680 (0.120)	94.710 (0.092)	94.590 (0.068)
			Above	2.400 (0.153)	2.050 (0.099)	2.660 (0.078)	2.620 (0.058)	2.640 (0.043)
			Below	2.900 (0.168)	2.950 (0.122)	2.660 (0.080)	2.670 (0.059)	2.770 (0.044)
			Length	1.651 (0.663)	1.749 (0.490)	1.740 (0.309)	1.734 (0.199)	1.734 (0.161)
		Smoothed		94.600 (0.226)	95.000 (0.160)	94.820 (0.115)	94.900 (0.088)	94.760 (0.067)
				2.100 (0.143)	2.050 (0.099)	2.600 (0.075)	2.530 (0.055)	2.565 (0.042)
				3.300 (0.179)	2.950 (0.120)	2.580 (0.078)	2.570 (0.058)	2.675 (0.043)
				1.665 (0.667)	1.771 (0.484)	1.773 (0.315)	1.763 (0.214)	1.759 (0.166)
	1000	PRR		94.200 (0.234)	94.750 (0.177)	95.040 (0.115)	94.700 (0.089)	94.675 (0.068)
				2.700 (0.162)	2.250 (0.108)	2.200 (0.071)	2.640 (0.057)	2.685 (0.042)
				3.100 (0.173)	3.000 (0.129)	2.760 (0.082)	2.660 (0.059)	2.640 (0.043)
				1.641 (0.552)	1.737 (0.365)	1.735 (0.208)	1.734 (0.137)	1.731 (0.119)
		Smoothed		94.200 (0.234)	94.800 (0.175)	95.080 (0.115)	94.920 (0.086)	94.895 (0.066)
				2.600 (0.159)	2.200 (0.110)	2.220 (0.072)	2.580 (0.057)	2.600 (0.042)
				3.200 (0.176)	3.000 (0.129)	2.700 (0.082)	2.500 (0.057)	2.505 (0.042)
				1.648 (0.555)	1.749 (0.365)	1.750 (0.213)	1.747 (0.143)	1.744 (0.115)
$t_5(0,1)$	500	PRR		95.400 (0.209)	94.600 (0.168)	95.000 (0.115)	94.550 (0.089)	94.645 (0.067)
				2.300 (0.150)	2.750 (0.116)	2.400 (0.073)	2.720 (0.056)	2.745 (0.044)
				2.300 (0.150)	2.650 (0.116)	2.600 (0.078)	2.730 (0.058)	2.610 (0.041)
				1.613 (0.611)	1.686 (0.416)	1.690 (0.269)	1.694 (0.228)	1.690 (0.208)
		Smoothed		95.200 (0.214)	94.700 (0.168)	95.420 (0.110)	94.920 (0.086)	94.865 (0.066)
				2.300 (0.150)	2.750 (0.116)	2.200 (0.069)	2.550 (0.052)	2.625 (0.044)
				2.500 (0.156)	2.550 (0.114)	2.380 (0.073)	2.530 (0.056)	2.510 (0.040)
				1.629 (0.610)	1.714 (0.425)	1.716 (0.282)	1.717 (0.230)	1.719 (0.209)
	1000	PRR		95.500 (0.207)	95.350 (0.160)	95.140 (0.116)	95.080 (0.085)	94.940 (0.061)
				1.800 (0.133)	2.200 (0.107)	2.480 (0.078)	2.550 (0.056)	2.625 (0.040)
				2.700 (0.162)	2.450 (0.115)	2.380 (0.076)	2.370 (0.054)	2.435 (0.039)
				1.613 (0.669)	1.700 (0.447)	1.704 (0.293)	1.699 (0.202)	1.696 (0.165)
		Smoothed		95.100 (0.216)	95.000 (0.163)	95.260 (0.115)	95.180 (0.085)	95.015 (0.060)
				1.600 (0.125)	2.150 (0.106)	2.280 (0.076)	2.410 (0.055)	2.540 (0.040)
				3.300 (0.179)	2.850 (0.120)	2.460 (0.075)	2.410 (0.054)	2.445 (0.038)
				1.618 (0.677)	1.709 (0.446)	1.714 (0.284)	1.709 (0.193)	1.712 (0.169)

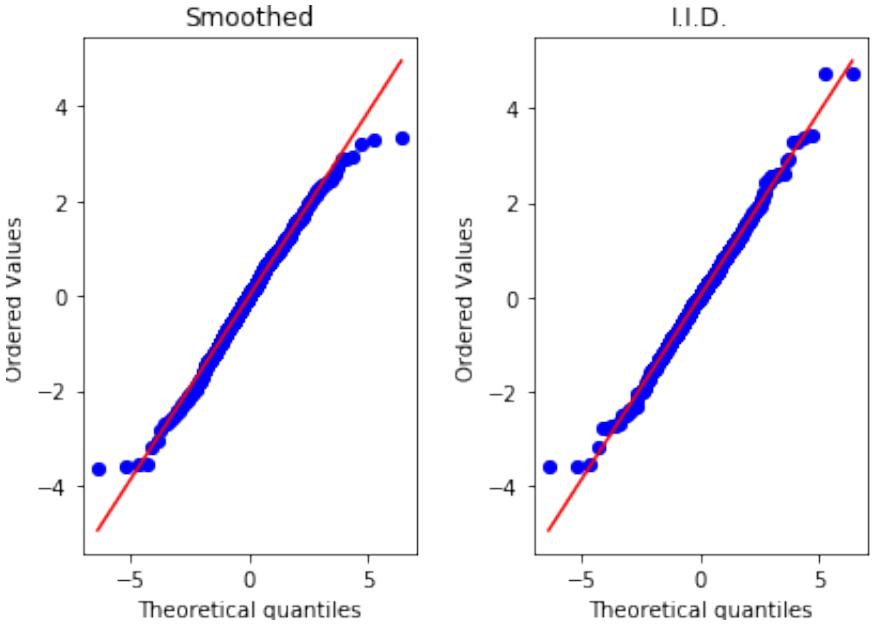

Figure 5.1: QQ-plot of z_t^*'s obtained via smoothed bootstrap (left) and i.i.d. bootstrap (right) for sample size $T = 500$, where the underlying distribution is $t_5(0, 1)$.

kernel is L^2 optimal for all smooth distributions; see Silverman (1986). The Epanechnikov kernel is

$$K(x) = \frac{3}{4}\left(1 - x^2\right), -1 \leq x \leq 1. \tag{6.1}$$

For choosing a bandwidth for the Epanechinkov kernel, there exist different optimality criteria. For instance, Hall and Wand (1988) obtain an optimal window by minimizing an L^1 and L^2 distance between the true and estimated distribution. They obtain optimal bandwidths of $u_E^{L^2} = 2.345\hat{\tau}T^{-0.2}$ under the L^2 criterion, and $u_E^{L^1} = 2.279\hat{\tau}T^{-0.2}$ under the L^1 criterion, where $\hat{\tau}$ is again the empirical standard deviation of $\hat{z}_{t,c}$. Both optimality considerations assume that the true density is normal. Sampling from the Epanechnikov kernel is discussed in Silverman (1986) and hence the z_t^*'s can be generated as follows:

1. Generate three uniform $[-1, 1]$ random variates V_1, V_2, V_3. If $|V_3| \geq |V_2|$ and $|V_3| \geq |V_1|$, set $\eta_t = V_2$; otherwise set $\eta_t = V_3$.

2. $z_t^* = z_t^\dagger + u\eta_t$, where $z_t^\dagger \sim i.i.d.\hat{F}_T$ and $u \in \{u_E^{L^1}, u_E^{L^2}\}$.

To assess the impact of the kernel function and its bandwidth on average coverage, we follow the same Monte Carlo approach as outlined in Section 5. Table 6.1 suggests that the choice of the Kernel and its bandwidth is not decisive when forecasting the conditional distribution of the volatility process. The same conclusion holds true for the forecasts of the conditional distribution of the return process; cf. Appendix C. This observation is explained by studying Figure 6.1. In this figure, we plot a QQ-plot of bootstrap innovations, z_t^*, drawn from a Epanechnikov kernel with bandwidth $u_E^{L^1}$; respectively $u_E^{L^2}$ for

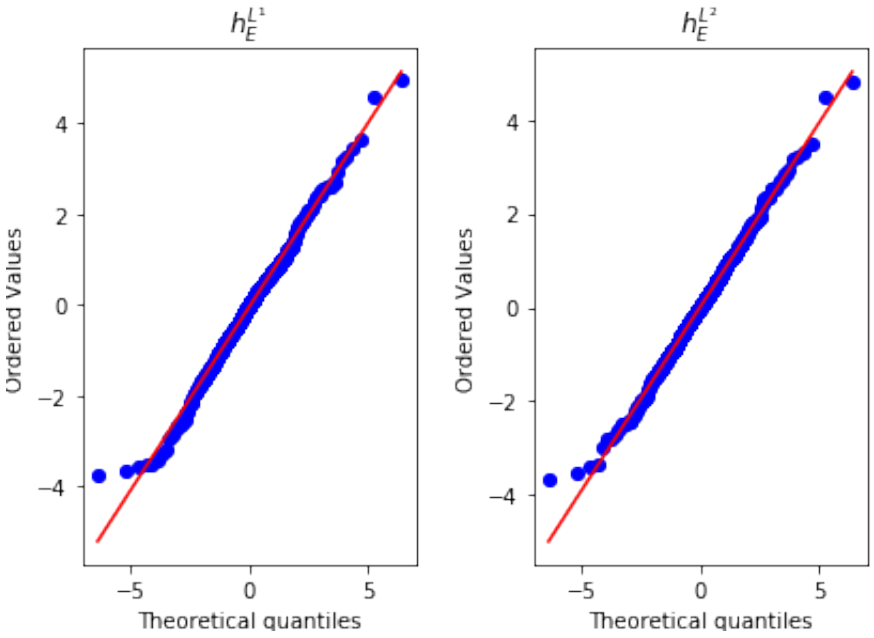

Figure 6.1: QQ-plot of z_t^*'s obtained via smoothed bootstrap using the Epanechnikov kernel with bandwidth $h_E^{L^1}$ (left); respectively $h_E^{L^1}$ (right) for sample size $T = 500$ where the underlying distribution is $t_5(0, 1)$.

sample size $T = 500$ where the assumed distribution is Student's t with 5 degrees of freedom. We observe that this kernel function for both bandwidths approximates the underlying distribution well for both bandwidths such that innovation uncertainty is taken into account.

7 Empirical illustration

Finally, we illustrate the performance of our smoothed bootstrap by replicating the empirical results in PRR. In the paper of PRR, the authors construct prediction intervals for the Spanish IBEX-35. Rather than using an ARCH(q) model, we proceed as in PRR and apply a GARCH(1,1)[1] model. The model parameters are recovered from daily closing prices of the IBEX-35 ranging from January 2, 1996 to March 3, 2000 such that we have 1042 observations in total. As in PRR, we first filter the original series to remove the effect of extraordinary events that occurred in Spain during the considered period; cf. PRR. The results of this filtering step are available in Appendix D. For the rest of this illustration, we will treat the residuals of the filtering step, denoted Y_t, as the underlying return series. As seen in Table D.2, the return series distribution is leptokurtotic and slightly skewed. Moreover, the levels of the series are not autocorrelated while the squared series displays autocorrelation as typical for financial data. Next, we fit a GARCH(1,1) model to Y_t and obtain the centered residuals, $\hat{\epsilon}_t$. The estimation output is provided in Table D.3 and sample moments of the centered residuals are given in Table D.4. The

[1]Note that our theory can be extended to the GARCH(1,1) case.

Table 6.1: Prediction intervals for the volatility process with 95% coverage level, where $\alpha = 0.5$. We use the Epanechnikov kernel and for the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Bandwidth		h				
				1	2	5	10	20
N(0,1)	500	$u_E^{L^1}$	Coverage	94.100 (0.236)	93.150 (0.194)	92.780 (0.149)	92.970 (0.128)	92.860 (0.117)
			Above	1.100 (0.104)	1.600 (0.091)	2.100 (0.072)	2.090 (0.060)	2.180 (0.048)
			Below	4.800 (0.214)	5.250 (0.175)	5.120 (0.134)	4.940 (0.116)	4.960 (0.105)
			Length	0.088 (0.184)	0.618 (1.013)	0.751 (0.951)	0.767 (0.800)	0.757 (0.539)
		$u_E^{L^2}$		94.800 (0.222)	93.550 (0.189)	93.100 (0.147)	93.180 (0.127)	93.030 (0.115)
				1.100 (0.104)	1.600 (0.091)	2.180 (0.074)	2.230 (0.062)	2.295 (0.050)
				4.100 (0.198)	4.850 (0.169)	4.720 (0.130)	4.590 (0.113)	4.675 (0.102)
				0.087 (0.183)	0.610 (1.000)	0.738 (0.922)	0.751 (0.786)	0.739 (0.512)
	1000	$u_E^{L^1}$		92.800 (0.258)	93.250 (0.186)	93.620 (0.134)	93.370 (0.113)	93.210 (0.102)
				2.300 (0.150)	2.350 (0.106)	2.120 (0.072)	2.220 (0.055)	2.385 (0.047)
				4.900 (0.216)	4.400 (0.160)	4.260 (0.118)	4.410 (0.102)	4.405 (0.092)
				0.055 (0.084)	0.542 (0.533)	0.652 (0.455)	0.657 (0.291)	0.658 (0.214)
		$u_E^{L^2}$		93.200 (0.252)	93.550 (0.182)	93.840 (0.131)	93.570 (0.110)	93.445 (0.100)
				2.400 (0.153)	2.400 (0.107)	2.160 (0.073)	2.270 (0.056)	2.440 (0.047)
				4.400 (0.205)	4.050 (0.154)	4.000 (0.115)	4.160 (0.100)	4.115 (0.090)
				0.055 (0.084)	0.538 (0.530)	0.645 (0.448)	0.649 (0.286)	0.649 (0.211)
$t_5(0,1)$	500	$u_E^{L^1}$		91.800 (0.274)	92.400 (0.205)	92.760 (0.156)	92.960 (0.135)	93.060 (0.126)
				5.000 (0.218)	3.950 (0.144)	3.160 (0.093)	2.990 (0.069)	2.900 (0.055)
				3.200 (0.176)	3.650 (0.151)	4.080 (0.132)	4.050 (0.119)	4.040 (0.112)
				0.114 (0.212)	0.695 (1.232)	0.815 (1.132)	0.833 (0.883)	0.828 (0.732)
		$u_E^{L^2}$		91.600 (0.277)	92.350 (0.205)	92.780 (0.156)	93.000 (0.134)	93.135 (0.124)
				5.300 (0.224)	4.200 (0.147)	3.360 (0.096)	3.140 (0.071)	3.030 (0.056)
				3.100 (0.173)	3.450 (0.147)	3.860 (0.129)	3.860 (0.117)	3.835 (0.109)
				0.114 (0.213)	0.689 (1.233)	0.802 (1.101)	0.817 (0.861)	0.810 (0.713)
	1000	$u_E^{L^2}$		92.500 (0.263)	93.800 (0.178)	94.160 (0.130)	94.050 (0.115)	94.160 (0.106)
				4.000 (0.196)	3.100 (0.123)	2.520 (0.077)	2.440 (0.057)	2.390 (0.042)
				3.500 (0.184)	3.100 (0.136)	3.320 (0.110)	3.510 (0.104)	3.450 (0.099)
				0.091 (0.213)	0.673 (1.221)	0.786 (1.240)	0.756 (0.752)	0.734 (0.491)
		$u_E^{L^1}$		92.500 (0.263)	93.850 (0.177)	94.300 (0.128)	94.190 (0.114)	94.330 (0.104)
				4.100 (0.198)	3.150 (0.124)	2.540 (0.077)	2.470 (0.058)	2.425 (0.042)
				3.400 (0.181)	3.000 (0.135)	3.160 (0.107)	3.340 (0.102)	3.245 (0.096)
				0.091 (0.212)	0.668 (1.210)	0.779 (1.230)	0.748 (0.749)	0.725 (0.483)

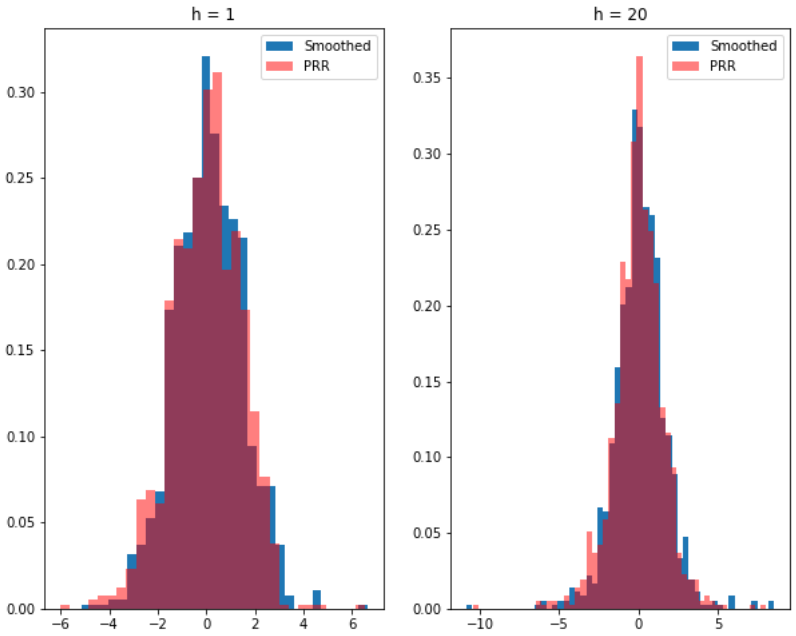

Figure 7.1: Histogram of one- and 20-step ahead predictions of returns based on PRR and smoothed bootstrap.

distribution of the residuals is still leptokurtotic and slightly skewed, however, there is no autocorrelation left in the residuals; cf. Figure D.2.

Next, we apply our smoothed bootstrap and the approach of PRR to create prediction intervals based on the simulated distribution of the return and volatility series. The forecasting period runs from March 4, 2000 to March 31, 2000. In Figure 7.1, we plot the the estimated bootstrap densities for $h = 1$ and $h = 20$. We note that the future densities are also asymmetric as observed with the returns before. Then, we use these densities to plot 95% prediction intervals for the period $h = 1, \ldots, 20$; cf. Figure 7.2. In addition, we plot the histograms for the bootstrap predictions of the volatility for $h = 1, 2, 10, 20$ into the future. Again, the shape of the future volatility is highly asymmetric. Overall, we can conclude that our smoothed bootstrap performs at par with the method of PRR. This results is also true when we use the Epanechnikov kernel with the two previously discussed bandwidths. The graphs for this kernel are available in Appendix D.

8 Conclusion

In this article, we have studied the asymptotic properties of the smoothed bootstrap scheme for forecasting volatility models when applied to a large class of ARCH(q) models. Specifically, we show that our smoothed bootstrap forecasts are asymptotically valid. In regards of theory, we exploit that the smoothed bootstrap residuals are obtained using a

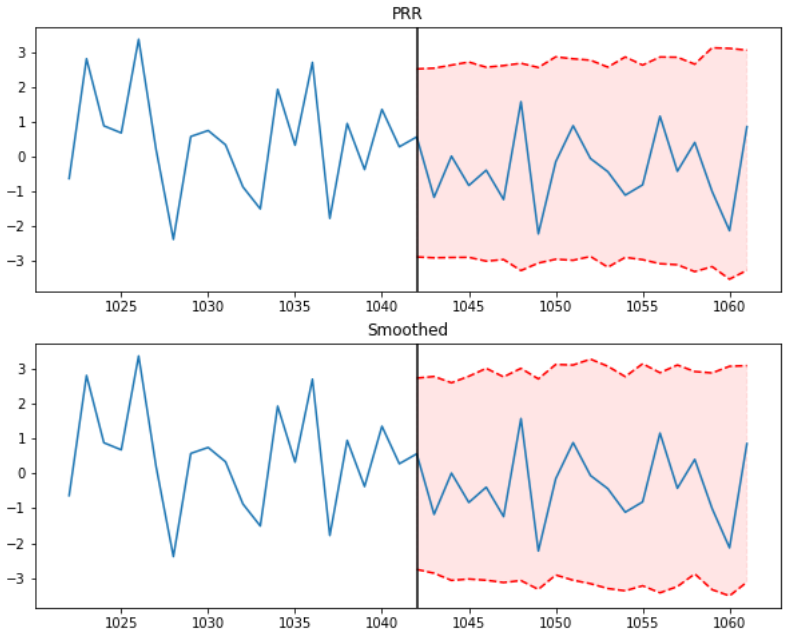

Figure 7.2: 20-step ahead prediction intervals of future returns based on PRR (upper) and smoothed bootstrap (lower).

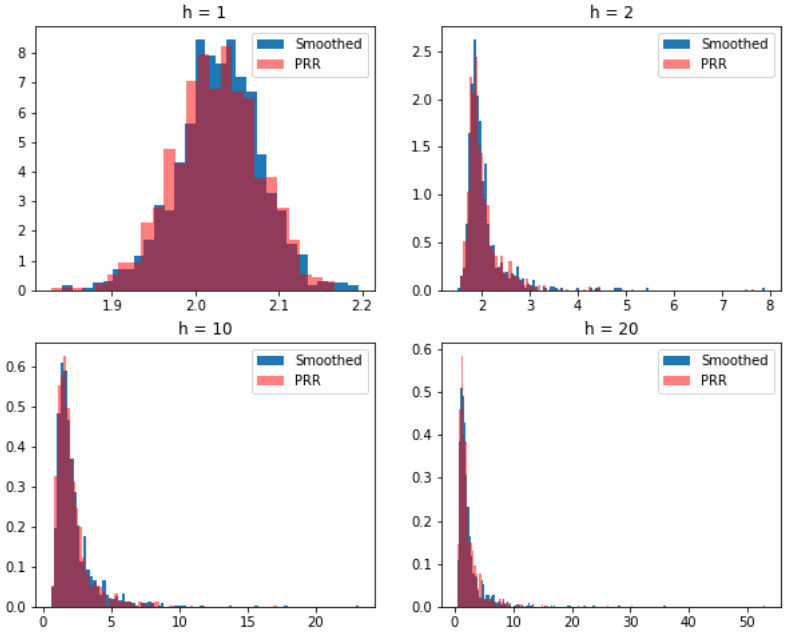

Figure 7.3: Histogram of one- and 20-step ahead predictions of volatility based on PRR and smoothed bootstrap.

non-parametric kernel estimator.

As a result the smoothed bootstrap process posses a density which simplifies the theoretical arguments considerably. More precisely, a well-behaved density of the bootstrap process allows us to rely on results in Franke, Neumann, and Stockis (2004) to show that our in-sample bootstrap process is geometric ergodic and approximates the underlying process well, conditional on the original sample. Given these results, we argue that the QML estimates based on the in-sample bootstrap process are consistent, conditional on the original sample. We thereby contribute to the theoretical understanding of bootstrap methods for (G)ARCH type process. In addition, we contribute to the forecasting literature for volatility models by providing a theoretically motivated alternative to existing bootstrap methods.

A Monte Carlo experiment confirms our theoretical results, that is, prediction intervals based on our smoothed bootstrap forecast of the conditional distribution of the volatility model are well calibrated. Moreover, the simulations show that our new bootstrap procedure delivers the same performance as existing bootstrap procedures while at the same time being theoretically justified. Our results also confirm that forecast of conditional distributions of volatility models should be corrected for parameter and innovation uncertainty. In addition, we show that our smoothed bootstrap is robust to the choice of the kernel function and its bandwidth. Finally, we demonstrate that our smoothed bootstrap performs at part with the method of PRR(2006) in a small empirical illustration.

For future research, it would be interesting to extend our arguments to the more general class of (G)ARCH type models.

Appendix

A Proofs

Proof of Corollary 4.1. The result is due to Theorem 3 in Kristensen and Rahbek (2005). □

Proof of Lemma 4.1. To show the result, consider the triangle inequality and write

$$\begin{aligned} \sup_{y\in\mathbb{R}} \left|\hat{p}_{\hat{z}}\left(y\right) - p_z\left(y\right)\right| &\leq \sup_{y\in\mathbb{R}} \left|\hat{p}_z\left(y\right) - p_z\left(y\right)\right| + \sup_{y\in\mathbb{R}} \left|\hat{p}_{\hat{z}}\left(y\right) - \hat{p}_z\left(y\right)\right|. \\ &=: T_1 + T_2. \end{aligned} \tag{A.1}$$

First, $T_1 = o\left(1\right)$ as $T \to \infty$ by Assumption 2.2 and Assumption 4.1. This holds because of the consistency of the kernel density estimator with i.i.d. observations; cf. Theorem 2.8 in Pagan and Ullah (1999).

For T_2 we use Assumption 4.1 in combination with $\hat{z}_{t,c} = \hat{z}_t - \hat{z}$ to see that

$$\begin{aligned} T_2 &= \sup_{y\in\mathbb{R}} \left| \frac{1}{Tu} \sum_{t=1}^{T} K\left(\frac{y-\hat{z}_{t,c}}{u}\right) - K\left(\frac{y-z_t}{u}\right) \right| \\ &\leq \frac{1}{Tu^2} \sum_{t=1}^{T} k\left|\hat{z}_t - \hat{z} - z_t\right| \\ &\leq \frac{1}{u^2} k\left|\hat{z}\right| + \frac{1}{Tu^2} \sum_{t=1}^{T} k\left|\hat{z}_t - z_t\right| \\ &=: T_3 + T_4, \end{aligned} \tag{A.2}$$

where we used the Lipschitz property of $K(\cdot)$ with $k \geq 0$ for the first inequality. The second inequality is due to the triangle inequality.

Next, we exploit the definition of $\hat{z}$ in (ii) of Algorithm 1 to write T_3 as

$$\begin{aligned} T_3 &= \frac{1}{u^2} k \left| \frac{1}{T} \sum_{t=1}^{T} \hat{z}_t \right| \\ &= \frac{1}{u^2} k \left| \frac{1}{T} \sum_{t=1}^{T} (\hat{z}_t - z_t) + \frac{1}{T} \sum_{t=1}^{T} z_t \right| \\ &\leq \frac{1}{Tu^2} \sum_{t=1}^{T} k\left|\hat{z}_t - z_t\right| + \frac{1}{u^2} k \left| \frac{1}{T} \sum_{t=1}^{T} z_t \right| \\ &=: T_4 + T_5. \end{aligned} \tag{A.3}$$

Further, for T_4 it holds that

$$\begin{aligned} T_4 &= \frac{1}{Tu^2} \sum_{t=1}^{T} k \left| y_t \frac{\hat{\sigma}_t - \sigma_t}{\hat{\sigma}_t \sigma_t} \right| \\ &= \frac{1}{Tu^2} \sum_{t=1}^{T} k \left| y_t \frac{(\hat{\sigma}_t - \sigma_t)(\hat{\sigma}_t + \sigma_t)}{\hat{\sigma}_t \sigma_t (\hat{\sigma}_t + \sigma_t)} \right| \\ &\leq \frac{1}{\sqrt{T}u^2} \sqrt{T}(\hat{\omega} - \omega_0) \frac{1}{T} \sum_{t=1}^{T} k\left|z_t\right| \frac{1}{\hat{\sigma}_t (\hat{\sigma}_t + \sigma_t)} \\ &\quad + \frac{1}{\sqrt{T}u^2} \sqrt{T}(\hat{\alpha} - \alpha_0) \frac{1}{T} \sum_{t=1}^{T} k\left|z_t\right| \frac{Y_{t-1}^2}{\hat{\sigma}_t (\hat{\sigma}_t + \sigma_t)} \\ &\leq \frac{1}{T^{1/2-\epsilon} u^2 T^{\epsilon}} \sqrt{T}(\hat{\omega} - \omega_0) \frac{1}{T} \sum_{t=1}^{T} k\left|z_t\right| \frac{1}{\underline{\omega}} \\ &\quad + \frac{1}{T^{1/2-\epsilon} u^2 T^{\epsilon}} \sqrt{T}(\hat{\alpha} - \alpha_0) \frac{1}{T} \sum_{t=1}^{T} k\left|z_t\right| \frac{1}{\underline{\alpha}} \\ &= o_p(T^{-\epsilon}), \end{aligned} \tag{A.4}$$

for some $0 < \epsilon < \frac{1}{2}$ since $T^{1/2-\epsilon}u^2 \to \infty$ by Assumption 4.1(b) (ii); $\sqrt{T}(\hat{\omega} - \omega_0) = O_p(1)$ and $\sqrt{T}(\hat{\alpha} - \alpha_0) = O_p(1)$ by Corollary 4.1; $\frac{1}{T}\sum_{t=1}^{T} k|z_t| \frac{1}{\underline{\omega}} = O_p(1)$ and $\frac{1}{T}\sum_{t=1}^{T} k|z_t| \frac{1}{\underline{\alpha}} = O_p(1)$ by the LLN and Assumption 2.3. Finally, we turn to T_5 to note that

$$T_5 = \frac{1}{T^{1/2-\epsilon}u^2T^{-\epsilon}} k \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t \right| = o_p\left(T^{-\epsilon}\right), \tag{A.5}$$

with $0 < \epsilon < \frac{1}{2}$, $\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t\right| = O_p(1)$ since the CLT applies due to Assumption 2.2 and $T^{1/2-\epsilon}u^2 \to \infty$ by Assumption 4.1(b) (ii).

Proof of Lemma 4.2. First, we note that

$$\begin{aligned} \hat{\sigma}(y) E^*[|z_t^*|] &\leq \hat{\omega}^{1/2} E^*[|z_t^*|] + \hat{\alpha}_1^{1/2} E^*[|z_t^*|]\,|y| \\ &= C_1 + C_2|y|, \end{aligned} \tag{A.6}$$

where $C_1 = \hat{\omega}^{1/2}E^*[|z_t^*|]$ and $C_2 = \hat{\alpha}_1^{1/2}E^*[|z_t^*|]$. Using a change of variable, we find or the term $E^*[|z_t^*|]$ that

$$\begin{aligned} E^*[|z_t^*|] &= \frac{1}{T}\sum_{t=1}^{T} \int_{-\infty}^{\infty} |\hat{z}_{t,c} + u\psi| K(\psi)\, d\psi \\ &\leq \frac{1}{T}\sum_{t=1}^{T} |\hat{z}_{t,c}| + u\int_{-\infty}^{\infty} |\psi| K(\psi)\, d\psi \\ &\leq \frac{1}{T}\sum_{t=1}^{T} |\hat{z}_t| - |\hat{z}| + o(1) \\ &= \frac{1}{T}\sum_{t=1}^{T} |\hat{z}_t| + o_p(1), \end{aligned} \tag{A.7}$$

where the second term in the second line is $o(1)$ due to Assumption 4.1. Further, $|\hat{z}| = \left|\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t\right| = o_p(1)$ by Lemma B.1(a). Moreover, by Lemma B.1(b), we see that $\frac{1}{T}\sum_{t=1}^{T}|\hat{z}_t| = \frac{1}{T}\sum_{t=1}^{T}|z_t| + o_P(1)$ such that $E^*[|z_t^*|] \to_p E[|z_t|]$. Thus, for T large enough, we get that $C_1 \leq \overline{\omega}^{1/2}E[|z_t|] < \infty$ since $\hat{\omega} \to_p \omega_0 < \overline{\omega}$. Next, we conclude that $E[|z_t|] \leq (E[|z_t|])^2 \leq E[|z_t^2|] = E[z_t^2] = 1$, where the second inequality follows by Jensen's inequality. Hence, $E^*[|z_t^*|] \to_p E[|z_t|] \leq 1$ and $\hat{\alpha} \to_p \alpha_0$, for T large enough, such that $C_2 < \alpha_0^{1/2}E[|z_t|] < 1$. □

Proof of Lemma 4.3. To show that $(Y_t^*)_{t=1}^T$ is geometrically ergodic, we verify that conditions (B2) and (B3) of Theorem 2 in Franke, Neumann, and Stockis (2004) hold. First, since we assumed that $\alpha_0 < 1$, we have that (B2) holds if $\sigma(y) = (\omega_0 + \alpha_0 y^2)^{1/2}$ is bounded away from zero and if the density p_z is continuous and everywhere positive. The

former holds under Assumption 2.3 while the later holds under Assumption 2.2. Next, condition (B3) of Franke, Neumann, and Stockis (2004) is satisfied due to Lemma 4.1. And hence the result follows. □

Proof of Lemma 4.4. To show (a), we first describe the tail behavior of $(Y_t)_{t=1}^T$ by Theorem 2.1 in Mikosch and Stărică (2000). We then use this behavior to show that $P(Y_t \in \mathcal{Y}_T^c)$ is bounded. First, note that the assumptions of Theorem 2.1 in Mikosch and Stărică (2000) hold under the assumptions made in Corollary 4.1. Then the equation

$$E\left(\alpha_0 z_t^2\right)^{\kappa/2} = 1 \tag{A.8}$$

has a unique positive solution. Let κ satisfy eq. (A.8). Then

$$P(|Y_t| > y) \sim c y^{-\kappa} \text{ as } y \to \infty, \tag{A.9}$$

where

$$c = E|z_t|^{\kappa} \frac{E\left(\left(\omega_0 + \alpha_0 z_t^2 \sigma_t^2\right)^{\kappa/2} - \left(\alpha_0 z_t^2 \sigma_t^2\right)^{\kappa/2}\right)}{(\kappa/2)\, E\left(\alpha_0 z_t^2\right)^{\kappa/2} \ln\left(\alpha_0 z_t^2\right)} \tag{A.10}$$

such that

$$P\left(|Y_t| > T^{\delta}\right) \sim c T^{-\kappa\delta}. \tag{A.11}$$

Hence, as a result of eq. (A.11), it follows that

$$\begin{aligned} P(Y_t \in \mathcal{Y}_T^c) &= P\left(|Y_t| > T^{\delta}\right) \\ &= O\left(T^{-\nu}\right), \end{aligned} \tag{A.12}$$

where we chose $\kappa\delta > \nu > 0$.

To prove (b), we use the mean value theorem and note that

$$\begin{aligned} \sup_{y \in \mathcal{Y}_T} |\hat{\sigma}(y) - \sigma(y)| &= \sup_{y \in \mathcal{Y}_T} \left| \frac{1}{2\sqrt{\tilde{\omega} + \tilde{\alpha} y^2}} \left[(\hat{\omega} - \omega_0) + (\hat{\alpha} - \alpha_0) y^2 \right] \right| \\ &\leq \frac{1}{\underline{\omega}} \frac{1}{\sqrt{T}} \sqrt{T} |\hat{\omega} - \omega_0| + \frac{1}{\underline{\alpha}} \frac{1}{\sqrt{T}} \sqrt{T} |\hat{\alpha} - \alpha_0| T^{2\delta} \\ &\leq O_p\left(T^{-1/2}\right) + O_p\left(T^{-1/2}\right) T^{2\delta} \\ &= O_p\left(T^{-\nu}\right), \end{aligned} \tag{A.13}$$

where $\tilde{\theta}_1 = (\tilde{\omega}, \tilde{\alpha}) \in \left(\hat{\theta}, \theta_0\right)$ to obtain the first line. Observe that the first inequality holds under Assumption 2.3. The third line follows by the result of Corollary 4.1. Finally,

the last equality is obtained with setting $\frac{1}{4} - \delta > \nu > 0$ and for $\alpha_0 > 0$ which holds by Assumption 2.3.

Next, to establish (c), we can use $\underline{\omega}$ from Assumption 2.3 to bound $\inf_{y \in \mathbb{R}} \hat{\sigma}(y)$ away from zero.

To prove (d), we write

$$\begin{aligned} \int |\hat{p}_{\hat{z}}(y) - p_z(y)|\, dy &= \int_{[-T^\delta, T^\delta]} |\hat{p}_{\hat{z}}(y) - p_z(y)|\, dy \\ &\quad + \int_{[-T^\delta, T^\delta]^\complement} |\hat{p}_{\hat{z}}(y) - p_z(y)|\, dy. \end{aligned} \tag{A.14}$$

For the first term we find that

$$\begin{aligned} \int_{[-T^\delta, T^\delta]} |\hat{p}_{\hat{z}}(y) - p_z(y)| &\leq \sup_{y \in \mathbb{R}} |\hat{p}_{\hat{z}}(y) - p_z(y)| \times 2T^\delta \\ &= O_p\left(T^{-\epsilon}\right) 2T^\delta \\ &= O_p\left(T^{-\nu}\right) \end{aligned} \tag{A.15}$$

for some $\epsilon - \delta > \nu > 0$, where we used that $\sup_{y \in \mathbb{R}} |\hat{p}_{\hat{z}}(y) - p_z(y)| = O_p(T^{-\epsilon})$ from Lemma (4.1). Furthermore, for second term in eq. (A.14) we obtain that

$$\int_{[-T^\delta, T^\delta]^\complement} |\hat{p}_{\hat{z}}(y) - p_z(y)|\, dy \leq \int_{[-T^\delta, T^\delta]^\complement} \hat{p}_{\hat{z}}(y)\, dy + \int_{[-T^\delta, T^\delta]^\complement} p_z(y)\, dy, \tag{A.16}$$

where for the latter term it holds that $\int_{[-T^\delta, T^\delta]^\complement} p_z(y)\, dy = P\left(|z_t| > T^\delta\right) = o(1)$. For the former term we write

$$\begin{aligned} \int_{[-T^\delta, T^\delta]^\complement} \hat{p}_{\hat{z}}(y)\, dy &= 1 - \int_{[-T^\delta, T^\delta]} \hat{p}_{\hat{z}}(y)\, dy \\ &\leq 1 - \int_{[-T^\delta, T^\delta]} p_z(y)\, dy + \int_{[-T^\delta, T^\delta]} |p_y(y) - \hat{p}_{\hat{z}}(y)|\, dy \\ &= 1 - \left(1 + \int_{[-T^\delta, T^\delta]^\complement} p_z(y)\, dy\right) + \int_{[-T^\delta, T^\delta]} |p_y(y) - \hat{p}_{\hat{z}}(y)|\, dy \\ &= o_p(1), \end{aligned} \tag{A.17}$$

where the first equality holds since $\hat{p}_{\hat{z}}$ is a density. And the last line follows by similar arguments given before. □

Proof of Theorem 4.1. To show the result, we verify that the conditions of Theorem 3 in Franke, Neumann, and Stockis (2004) hold. First, (B1) is fulfilled if $\alpha_0 < 1$ which is assumed in Corollary 4.1. Second, if (B1) holds, then condition (B2) in Franke, Neu-

mann, and Stockis (2004) is satisfied if $\sigma(y)$ is bounded away from zero which holds by Assumption 2.2. And if the density of the innovations is continuous and everywhere positive which is ensured by Assumption 2.3. Third, due to Lemma 4.1(a) and (b), we have that (B3)(i) and (ii) hold. In addition, (B4)(ii)-(iv) are fulfilled by the results in Lemma 4.4. Finally, (B4)(v) is assumed in Assumption 4.2. Thus, conditions (B1) to (B4) in Franke, Neumann, and Stockis (2004) hold and the result follows. □

Proof of Lemma 4.5 To prove the first part, using a change of variable, we find that

$$\begin{aligned} E^*[z_t^*] &= \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{\infty}(\hat{z}_{t,c}+u\psi)K(\psi)\,d\psi \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{z}_{t,c}+u\int_{-\infty}^{\infty}\psi K(\psi)\,d\psi \\ &= o_p(1) \end{aligned} \tag{A.18}$$

where $T^{-1}\sum_{t=1}^{T}\hat{z}_{t,c}=0$ by construction and $u\int_{-\infty}^{\infty}\psi K(\psi)\,d\psi = o(1)$ by Assumption 4.1(b)(i).

To calculate the second moment, again using a change of variable, we obtain that

$$\begin{aligned} E^*\left[z_t^{*2}\right] &= \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{\infty}(\hat{z}_{t,c}+u\psi)^2K(\psi)\,d\psi \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{z}_{t,c}^2+u^2\int_{-\infty}^{\infty}\psi^2K(\psi)\,d\psi+2\left(\frac{1}{T}\sum_{t=1}^{T}\hat{z}_{t,c}\right)u\int_{-\infty}^{\infty}\psi K(\psi)\,d\psi \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{z}_{t,c}^2+o(1), \\ &= \frac{1}{T}\sum_{t=1}^{T}(\hat{z}_t-\hat{z})^2+o(1) \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{z}_t^2-\left(\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t\right)+o(1) \\ &= 1+o_p(1), \end{aligned} \tag{A.19}$$

where the last two terms in the second line are $o(1)$ by similar arguments as before. And the last line is due to Lemma B.1(a), respectively (c).

Next, to obtain the fourth moment, using the same approach as before, we get that

$$\begin{aligned} E^*\left[z_t^{*4}\right] &= \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{\infty}\left(\hat{z}_{t,c}+u\psi\right)^4 K\left(\psi\right)d\psi \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{z}_{t,c}^4+o(1) \\ &= O_p(1), \end{aligned} \tag{A.20}$$

where the second and third equality follow by similar arguments as before. And the last equation holds under Lemma B.1(d).

Finally,

$$\begin{aligned} \left|P^*\left(z_h^*\leq x\right)-P\left(z_h\leq x\right)\right| &= \left|\int_{-\infty}^{x}\left(\hat{p}_{\hat{z}}\left(y\right)-p_z\left(y\right)\right)dy\right| \\ &\leq \int_{-\infty}^{x}\left|\hat{p}_{\hat{z}}\left(y\right)-p_z\left(y\right)\right|dy \\ &\leq \int_{-\infty}^{\infty}\left|\hat{p}_{\hat{z}}\left(y\right)-p_z\left(u\right)\right|dy \\ &= o_P\left(1\right) \end{aligned} \tag{A.21}$$

by the same arguments used in the proof of Lemma 4.4. □

Proof of Lemma 4.6. By Lemma 4.1, it holds that z_t^*'s have the same properties as the model innovations, conditional on the original sample. Due to Lemma 4.3, the bootstrap process, $(Y_t^*)_{t=1}^T$, is geometric ergodic, conditional on the data. Thus we have that the bootstrap analogues of Assumption 2.1, 2.2 and 2.3 hold, conditional on the data. Consequently, Corollary 4.1 holds, conditional on the original sample, which is enough to establish the result. □

Proof of Theorem 4.2. For $h=1$, we use a Taylor expansion and write

$$\begin{aligned} Y_1^* &= \left(\sqrt{\omega_0+\alpha_0 y^2}+\frac{1}{2\sqrt{\tilde{\omega}+\tilde{\alpha}_1 y^2}}\left(\left(\hat{\omega}^*-\omega_0\right)+\left(\hat{\alpha}^*-\alpha_0\right)y^2\right)\right)z_1^* \\ &= \sqrt{\omega_0+\alpha_0 y^2}z_1^*+o_p^*\left(1\right), \end{aligned} \tag{A.22}$$

where $\tilde{\theta}=(\tilde{\omega},\tilde{\alpha})\in\left(\hat{\theta}^*,\theta_0\right)$. Then, $\left|P^*\left(Y_1^*\leq x|X_0^*=y\right)-P\left(Y_1\leq x|X_0=y\right)\right|\xrightarrow{p}0$ since $z_1^*\xrightarrow{w^*}_p z_1$ due to Lemma 4.5.

For $h = 2$, we see that

$$\begin{aligned} Y_2^* &= \sqrt{\omega_0 + \alpha_0 Y_1^{*2}} z_2^* + \frac{1}{2\sqrt{\tilde{\omega} + \tilde{\alpha}_1 Y_1^{*2}}} \left((\hat{\omega}^* - \omega_0) + (\hat{\alpha}^* - \alpha_0) Y_1^{*2} \right) z_2^* \\ &= \sqrt{\omega_0 + \alpha_0 Y_1^{*2}} z_2^* + o_p^*(1), \end{aligned} \tag{A.23}$$

where $Y_1^{*2} = \omega_0 z_1^{*2} + \alpha_0 z_1^{*2} y^2 + o_p^*(1)$. Now, by Lemma 4.5, $(z_1^*, z_2^*) \xrightarrow{w^*}_p (z_1, z_2)$ such that convergence in terms of simultaneous distribution implies $|P^*(Y_2^* \leq x | X_0^* = y) - P(Y_2 \leq x | X_0 = y)| \xrightarrow{p} 0$ by the bootstrap version of Slutsky's lemma.

In the case of $h \geq 3$, we write

$$\begin{aligned} Y_h^* &= \sqrt{\omega_0 + \alpha_0 Y_{h-1}^{*2}} z_h^* + \frac{1}{2\sqrt{\tilde{\omega} + \tilde{\alpha}_1 Y_{h-1}^{*2}}} \left((\hat{\omega}^* - \omega_0) + (\hat{\alpha}^* - \alpha_0) Y_{h-1}^{*2} \right) z_h^* \\ &= \sqrt{\omega_0 + \alpha_0 Y_{h-1}^{*2}} z_h^* + o_p^*(1), \end{aligned} \tag{A.24}$$

where $Y_h^* = \omega_0 \sum_{k=0}^{h-1} \alpha_0^k \prod_{m=0}^{k} z_{h-m}^{*2} + \alpha_0^h \left(\prod_{m=0}^{h-1} z_{h-m}^{*2} \right) y^2 + o_p^*(1)$. Then, with $(z_1^*, \ldots, z_h^*) \xrightarrow{w^*} (z_1, \ldots, z_h)$ such that convergence in terms of simultaneous distribution implies that $|P^*(Y_h^* \leq x | X_0^* = y) - P(Y_h \leq x | X_0 = y)| \xrightarrow{p} 0$ by the bootstrap version of Slutsky's lemma.

For σ_h^{*2} we get with $h = 1$ that

$$\begin{aligned} \sigma_1^{*2} &= \hat{\omega}^* + \hat{\alpha}^* y^2 \\ &= \left(\omega_0 + o_p^*(1) \right) + \left(\alpha_0 + o_p^*(1) \right) y^2 \\ &= \omega_0 + \alpha_0 y^2 + o_p^*(1), \end{aligned} \tag{A.25}$$

where the second quality holds due to Lemma 4.6. Next, for $h \geq 2$, we write

$$\begin{aligned} \sigma_h^{*2} &= \hat{\omega}^* + \sum_{k=1}^{h-1} \hat{\alpha}^{*k} \prod_{m=1}^{k} z_{h-m}^{*2} + \hat{\alpha}^* \left(\prod_{m=1}^{h-1} z_{h-m}^{*2} \right) y^2 \\ &= \omega_0 + \sum_{k=1}^{h-1} \alpha_0 \prod_{m=1}^{k} z_{h-m}^{*2} + \alpha_0 \left(\prod_{m=1}^{h-1} z_{h-m}^{*2} \right) y^2 + o_p^*(1). \end{aligned} \tag{A.26}$$

Again, with $(z_1^*, \ldots, z_h^*) \xrightarrow{w^*}_p (z_1, \ldots, z_h)$ such that convergence in terms of simultaneous distribution implies that $|P^*(\sigma_h^{*2} \leq x | X_0^* = y) - P(\sigma_h^2 \leq x | X_0 = y)| \xrightarrow{p} 0$ by the bootstrap version of Slutsky's lemma.

□

B Lemmas for the residuals

Lemma B.1 *Suppose that the assumptions of Corollary 4.1 hold. Then, as* $T \to \infty$,

(a) $\left|\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t\right| = \left|\frac{1}{T}\sum_{t=1}^{T}z_t\right| + o_P(1)$,

(b) $\frac{1}{T}\sum_{t=1}^{T}|\hat{z}_t| = \frac{1}{T}\sum_{t=1}^{T}|z_t| + o_P(1)$,

(c) $\left|\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t^2\right| = \left|\frac{1}{T}\sum_{t=1}^{T}z_t^2\right| + o_P(1)$,

(d) $\left|\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t^4\right| = \left|\frac{1}{T}\sum_{t=1}^{T}z_t^4\right| + o_P(1)$.

Proof of Lemma B.1. To show (a), we use the definition of $\hat{z}_t$ and Y_t to see that

$$\begin{aligned}
\left|\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}z_t\frac{\sigma_t}{\hat{\sigma}_t}\right| \\
&\leq \left|\frac{1}{T}\sum_{t=1}^{T}z_t\right| + \left|\frac{1}{T}\sum_{t=1}^{T}\frac{z_t}{2\left(\tilde{\omega}+\tilde{\alpha}Y_{t-1}^2\right)^{1/2}\left(\hat{\omega}+\hat{\alpha}Y_{t-1}^2\right)^{1/2}}\left[\left(\hat{\omega}-\omega_0\right)+\left(\hat{\alpha}-\alpha_0\right)Y_{t-1}^2\right]\right| \\
&\leq \left|\frac{1}{T}\sum_{t=1}^{T}z_t\right| + \frac{1}{\underline{\omega}}\left|\hat{\omega}-\omega_0\right|\left|\frac{1}{T}\sum_{t=1}^{T}z_t\right| + \frac{1}{\underline{\alpha}}\left|\hat{\alpha}-\alpha_0\right|\left|\frac{1}{T}\sum_{t=1}^{T}z_t\right| \\
&= \left|\frac{1}{T}\sum_{t=1}^{T}z_t\right| + o_P(1), \qquad \text{(B.1)}
\end{aligned}$$

where we use the mean value theorem with $\tilde{\theta} = (\tilde{\omega}, \tilde{\alpha}) \in \left(\hat{\theta}, \theta_0\right)$ to obtain the first inequality. The second inequality follows under Assumption 2.3. And the last line is due to Assumption 2.2 and Corollary 4.1.

For (b), we use the definition of $\hat{z}_t$ and Y_t to find that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}|\hat{z}_t| &= \frac{1}{T}\sum_{t=1}^{T}\left|z_t\frac{\sigma_t}{\hat{\sigma}_t}\right| \\
&\leq \frac{1}{T}\sum_{t=1}^{T}|z_t| + \frac{1}{\underline{\omega}}\left|\hat{\omega}-\omega_0\right|\frac{1}{T}\sum_{t=1}^{T}|z_t| + \frac{1}{\underline{\alpha}}\left|\hat{\alpha}-\alpha_0\right|\frac{1}{T}\sum_{t=1}^{T}|z_t| \\
&= \frac{1}{T}\sum_{t=1}^{T}|z_t| + o_P(1), \qquad \text{(B.2)}
\end{aligned}$$

where we used the same arguments to obtain the second line as in eq. (B.1). The last line follows with Corollary 4.1 and under Assumption 2.2 and Assumption 2.3.

To show (c), we use again the definition of $\hat{z}_t$ and Y_t to see that

$$\begin{aligned}
\left|\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t^2\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}\frac{\sigma_t^2}{\hat{\sigma}_t^2}z_t^2\right| \\
&\leq \left|\frac{1}{T}\sum_{t=1}^{T}z_t^2\right| + \left|\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\sigma}_t^2-\sigma_t^2}{\hat{\sigma}_t^2}z_t^2\right| \\
&\leq \left|\frac{1}{T}\sum_{t=1}^{T}z_t^2\right| + \frac{1}{\underline{\omega}}\left|\hat{\omega}-\omega_0\right|\left|\frac{1}{T}\sum_{t=1}^{T}z_t^2\right| + \frac{1}{\underline{\alpha}}\left|\hat{\alpha}-\alpha_0\right|\left|\frac{1}{T}\sum_{t=1}^{T}z_t^2\right| \\
&= \left|\frac{1}{T}\sum_{t=1}^{T}z_t^2\right| + o_P(1), \qquad \text{(B.3)}
\end{aligned}$$

where the second inequality follows by using the definition of $\hat{\sigma}_t^2$ and σ_t. Finally, the last equality holds because of Corollary 4.1, Assumption 2.2, and Assumption 2.3.

Finally, for (d), we see that

$$\begin{aligned}
\left|\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t^4\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}z_t^2\frac{\sigma_t^2}{\hat{\sigma}_t^2}z_t^2\frac{\sigma_t^2}{\hat{\sigma}_t^2}\right| \\
&\leq \left|\frac{1}{T}\sum_{t=1}^{T}z_t^4\right| + o_P(1)
\end{aligned} \qquad \text{(B.4)}$$

where the result follows by the same arguments used in deriving eq. (B.3). □

C Additional simulation results

Table C.1: Prediction intervals for the volatility process with 95% coverage level, where the DGP is a GJR-ARCH model with $(\omega, \alpha, \gamma) = (0.1, 0.1, 0.35)$. For the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Method		h				
				1	2	5	10	20
N(0,1)	500	PRR	Coverage	94.000 (0.237)	94.500 (0.169)	94.460 (0.128)	94.560 (0.109)	94.630 (0.093)
			Above	4.500 (0.207)	4.000 (0.141)	3.680 (0.095)	3.410 (0.068)	3.430 (0.053)
			Below	1.500 (0.122)	1.500 (0.099)	1.860 (0.088)	2.030 (0.087)	1.940 (0.078)
			Length	0.060 (0.069)	0.256 (0.198)	0.258 (0.110)	0.256 (0.087)	0.259 (0.087)
		Smoothed		94.600 (0.226)	94.800 (0.162)	94.500 (0.127)	94.490 (0.111)	94.520 (0.098)
				3.800 (0.191)	3.500 (0.130)	3.400 (0.090)	3.230 (0.066)	3.185 (0.051)
				1.600 (0.125)	1.700 (0.103)	2.100 (0.094)	2.280 (0.091)	2.295 (0.084)
				0.060 (0.070)	0.260 (0.194)	0.268 (0.111)	0.266 (0.091)	0.264 (0.089)
	1000	PRR		93.600 (0.245)	93.950 (0.179)	93.880 (0.136)	93.910 (0.115)	94.155 (0.104)
				4.300 (0.203)	3.550 (0.130)	3.260 (0.086)	3.110 (0.061)	2.930 (0.044)
				2.100 (0.143)	2.500 (0.124)	2.860 (0.110)	2.980 (0.101)	2.915 (0.094)
				0.043 (0.047)	0.244 (0.136)	0.249 (0.077)	0.247 (0.062)	0.247 (0.062)
		Smoothed		94.100 (0.236)	93.950 (0.182)	93.720 (0.140)	93.780 (0.118)	93.955 (0.109)
				3.700 (0.189)	3.150 (0.124)	3.080 (0.083)	2.890 (0.058)	2.785 (0.043)
				2.200 (0.147)	2.900 (0.135)	3.200 (0.117)	3.330 (0.106)	3.260 (0.099)
				0.044 (0.047)	0.247 (0.144)	0.254 (0.077)	0.251 (0.063)	0.253 (0.064)

Table C.2: Prediction intervals for the return process with 95% coverage level, where the DGP is a GJR-ARCH model with $(\omega, \alpha, \gamma) = (0.1, 0.1, 0.35)$. For the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Method		h				
				1	2	5	10	20
N(0,1)	500	PRR	Coverage	95.200 (0.214)	94.150 (0.178)	94.680 (0.116)	94.800 (0.082)	94.505 (0.062)
			Above	2.300 (0.150)	2.500 (0.111)	2.440 (0.072)	2.440 (0.053)	2.715 (0.039)
			Below	2.500 (0.156)	3.350 (0.135)	2.880 (0.079)	2.760 (0.057)	2.780 (0.043)
			Length	1.407 (0.306)	1.435 (0.142)	1.435 (0.091)	1.439 (0.087)	1.437 (0.090)
		Smoothed		95.100 (0.216)	94.700 (0.169)	95.020 (0.112)	95.040 (0.078)	94.840 (0.059)
				2.200 (0.147)	2.250 (0.106)	2.300 (0.070)	2.340 (0.051)	2.560 (0.037)
				2.700 (0.162)	3.050 (0.130)	2.680 (0.078)	2.620 (0.056)	2.600 (0.042)
				1.418 (0.308)	1.453 (0.144)	1.454 (0.092)	1.451 (0.088)	1.454 (0.086)
	1000	PRR	Coverage	95.300 (0.212)	95.050 (0.161)	95.120 (0.109)	94.770 (0.079)	94.860 (0.057)
			Above	2.400 (0.153)	2.200 (0.105)	2.240 (0.069)	2.610 (0.050)	2.600 (0.036)
			Below	2.300 (0.150)	2.750 (0.120)	2.640 (0.081)	2.620 (0.056)	2.540 (0.039)
			Length	1.412 (0.278)	1.444 (0.117)	1.441 (0.073)	1.441 (0.070)	1.443 (0.071)
		Smoothed		95.400 (0.209)	95.300 (0.157)	95.100 (0.108)	94.880 (0.077)	94.915 (0.056)
				2.200 (0.147)	2.000 (0.100)	2.120 (0.066)	2.460 (0.049)	2.530 (0.036)
				2.400 (0.153)	2.700 (0.119)	2.780 (0.082)	2.660 (0.057)	2.555 (0.039)
				1.417 (0.282)	1.454 (0.119)	1.453 (0.072)	1.452 (0.072)	1.452 (0.073)

Table C.3: Prediction intervals for the volatility process with 95% coverage level, where $\alpha = 0.8$. For the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Method		h				
				1	2	5	10	20
N(0,1)	500	PRR	Coverage	94.600 (0.226)	93.900 (0.171)	94.080 (0.127)	94.420 (0.105)	94.315 (0.094)
			Above	4.200 (0.201)	3.500 (0.133)	3.280 (0.093)	3.070 (0.079)	3.105 (0.068)
			Below	1.200 (0.109)	2.600 (0.113)	2.640 (0.092)	2.510 (0.075)	2.580 (0.067)
			Length	0.196 (0.917)	2.241 (11.138)	3.227 (16.834)	3.243 (13.697)	3.009 (10.488)
		Smoothed		95.000 (0.218)	93.900 (0.173)	93.960 (0.130)	94.300 (0.107)	94.190 (0.097)
				3.400 (0.181)	3.050 (0.126)	2.960 (0.087)	2.840 (0.076)	2.905 (0.066)
				1.600 (0.125)	3.050 (0.124)	3.080 (0.100)	2.860 (0.080)	2.905 (0.071)
				0.197 (0.935)	2.290 (11.121)	3.310 (15.810)	3.676 (17.275)	3.421 (14.252)
	1000	PRR		93.100 (0.253)	94.150 (0.168)	94.980 (0.110)	95.070 (0.089)	94.825 (0.080)
				4.800 (0.214)	3.800 (0.132)	2.880 (0.082)	2.710 (0.064)	2.900 (0.058)
				2.100 (0.143)	2.050 (0.111)	2.140 (0.080)	2.220 (0.067)	2.275 (0.060)
				0.132 (0.680)	1.835 (7.371)	2.611 (11.328)	2.689 (8.254)	2.560 (6.463)
		Smoothed		93.600 (0.245)	94.250 (0.167)	94.820 (0.113)	94.940 (0.090)	94.640 (0.082)
				3.700 (0.189)	3.150 (0.121)	2.720 (0.083)	2.540 (0.062)	2.760 (0.056)
				2.700 (0.162)	2.600 (0.122)	2.460 (0.085)	2.520 (0.071)	2.600 (0.063)
				0.132 (0.693)	1.860 (7.465)	2.649 (10.410)	2.761 (9.047)	2.617 (5.670)
$t_5(0,1)$	500	PRR		90.000 (0.300)	92.050 (0.204)	93.160 (0.144)	93.740 (0.116)	93.900 (0.104)
				8.700 (0.282)	5.800 (0.169)	4.200 (0.104)	3.690 (0.078)	3.585 (0.064)
				1.300 (0.113)	2.150 (0.117)	2.640 (0.106)	2.570 (0.091)	2.515 (0.083)
				0.180 (0.581)	1.678 (5.338)	2.012 (5.364)	1.964 (3.880)	1.837 (2.351)
		Smoothed		90.800 (0.289)	92.600 (0.199)	93.260 (0.147)	93.660 (0.121)	93.725 (0.109)
				7.500 (0.263)	5.100 (0.159)	3.900 (0.103)	3.460 (0.077)	3.450 (0.063)
				1.700 (0.129)	2.300 (0.124)	2.840 (0.112)	2.880 (0.098)	2.825 (0.089)
				0.179 (0.581)	1.683 (5.151)	2.038 (4.837)	2.027 (3.532)	1.962 (2.898)
	1000	PRR		91.900 (0.273)	93.700 (0.176)	94.620 (0.118)	94.830 (0.102)	94.815 (0.090)
				6.700 (0.250)	4.500 (0.147)	3.320 (0.091)	2.950 (0.071)	2.890 (0.052)
				1.400 (0.117)	1.800 (0.106)	2.060 (0.084)	2.220 (0.081)	2.295 (0.076)
				0.175 (0.819)	2.077 (9.489)	2.597 (11.083)	2.536 (10.834)	2.256 (6.535)
		Smoothed		91.700 (0.276)	93.450 (0.179)	94.420 (0.123)	94.500 (0.107)	94.575 (0.092)
				6.500 (0.247)	4.400 (0.142)	3.200 (0.089)	2.920 (0.072)	2.825 (0.052)
				1.800 (0.133)	2.150 (0.117)	2.380 (0.092)	2.580 (0.087)	2.600 (0.080)
				0.173 (0.802)	2.179 (10.896)	2.645 (12.148)	2.490 (9.123)	2.154 (5.031)

Table C.4: Prediction intervals for the return process with 95% coverage level, where $\alpha = 0.8$. For the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Method		h				
				1	2	5	10	20
N(0,1)	500	PRR	Coverage	94.600 (0.226)	94.800 (0.165)	94.560 (0.128)	94.600 (0.106)	94.450 (0.085)
			Above	2.400 (0.153)	2.350 (0.106)	2.880 (0.085)	2.830 (0.066)	2.795 (0.051)
			Below	3.000 (0.171)	2.850 (0.120)	2.560 (0.081)	2.570 (0.065)	2.755 (0.050)
			Length	2.144 (1.872)	2.447 (1.921)	2.490 (1.660)	2.480 (1.246)	2.456 (0.966)
		Smoothed		94.800 (0.222)	95.050 (0.161)	94.960 (0.122)	94.880 (0.102)	94.830 (0.081)
				2.100 (0.143)	2.150 (0.104)	2.580 (0.081)	2.650 (0.064)	2.575 (0.048)
				3.100 (0.173)	2.800 (0.117)	2.460 (0.076)	2.470 (0.061)	2.595 (0.048)
				2.162 (1.892)	2.483 (1.944)	2.550 (1.697)	2.538 (1.330)	2.510 (0.997)
	1000	PRR		94.200 (0.234)	94.750 (0.182)	95.280 (0.119)	94.810 (0.096)	94.650 (0.078)
				2.700 (0.162)	2.250 (0.111)	2.260 (0.074)	2.620 (0.059)	2.680 (0.046)
				3.100 (0.173)	3.000 (0.133)	2.460 (0.080)	2.570 (0.060)	2.670 (0.047)
				2.121 (1.563)	2.417 (1.547)	2.463 (1.184)	2.452 (0.841)	2.430 (0.614)
		Smoothed		94.400 (0.230)	94.900 (0.176)	95.260 (0.121)	94.900 (0.098)	94.815 (0.078)
				2.500 (0.156)	2.200 (0.110)	2.280 (0.077)	2.570 (0.060)	2.615 (0.046)
				3.100 (0.173)	2.900 (0.127)	2.460 (0.080)	2.530 (0.061)	2.570 (0.046)
				2.129 (1.548)	2.433 (1.471)	2.495 (1.221)	2.480 (0.892)	2.456 (0.634)
$t_5(0,1)$	500	PRR		95.600 (0.205)	94.500 (0.175)	94.880 (0.128)	94.510 (0.100)	94.415 (0.077)
				2.300 (0.150)	2.950 (0.122)	2.500 (0.077)	2.780 (0.060)	2.880 (0.049)
				2.100 (0.143)	2.550 (0.119)	2.620 (0.084)	2.710 (0.064)	2.705 (0.046)
				1.952 (1.267)	2.117 (1.043)	2.130 (0.749)	2.117 (0.528)	2.112 (0.404)
		Smoothed		95.600 (0.205)	94.800 (0.167)	95.200 (0.125)	94.740 (0.098)	94.645 (0.076)
				2.200 (0.147)	2.600 (0.113)	2.280 (0.072)	2.650 (0.058)	2.740 (0.048)
				2.200 (0.147)	2.600 (0.118)	2.520 (0.082)	2.610 (0.062)	2.615 (0.045)
				1.966 (1.245)	2.153 (1.049)	2.165 (0.734)	2.160 (0.551)	2.156 (0.409)
	1000	PRR		95.600 (0.205)	94.500 (0.175)	94.880 (0.128)	94.510 (0.100)	94.415 (0.077)
				2.300 (0.150)	2.950 (0.122)	2.500 (0.077)	2.780 (0.060)	2.880 (0.049)
				2.100 (0.143)	2.550 (0.119)	2.620 (0.084)	2.710 (0.064)	2.705 (0.046)
				1.952 (1.267)	2.117 (1.043)	2.130 (0.749)	2.117 (0.528)	2.112 (0.404)
		Smoothed		95.600 (0.205)	94.800 (0.167)	95.200 (0.125)	94.740 (0.098)	94.645 (0.076)
				2.200 (0.147)	2.600 (0.113)	2.280 (0.072)	2.650 (0.058)	2.740 (0.048)
				2.200 (0.147)	2.600 (0.118)	2.520 (0.082)	2.610 (0.062)	2.615 (0.045)
				1.966 (1.245)	2.153 (1.049)	2.165 (0.734)	2.160 (0.551)	2.156 (0.409)

Table C.5: Prediction intervals for the return process with 95% coverage level, where $\alpha = 0.5$. We use the Epanechnikov kernel and for the bootstrap we set $B = 999$. Standard errors are given in parenthesis and based on $M = 1000$.

z_t	T	Bandwidth		h				
				1	2	5	10	20
N(0,1)	500	$u_E^{L^1}$	Coverage	95.700 (0.203)	95.750 (0.153)	95.500 (0.112)	95.570 (0.086)	95.670 (0.063)
			Above	2.100 (0.143)	1.850 (0.100)	2.380 (0.075)	2.310 (0.054)	2.180 (0.039)
			Below	2.200 (0.147)	2.400 (0.109)	2.120 (0.072)	2.120 (0.055)	2.150 (0.040)
			Length	1.718 (0.689)	1.856 (0.510)	1.870 (0.351)	1.872 (0.249)	1.869 (0.199)
		$u_E^{L^2}$		95.100 (0.216)	95.500 (0.166)	95.920 (0.105)	95.560 (0.080)	95.425 (0.063)
				2.300 (0.150)	1.950 (0.102)	1.840 (0.066)	2.240 (0.053)	2.295 (0.039)
				2.600 (0.147)	2.550 (0.109)	2.240 (0.072)	2.200 (0.055)	2.280 (0.040)
				1.681 (0.564)	1.803 (0.381)	1.813 (0.221)	1.811 (0.147)	1.811 (0.126)
	1000	$u_E^{L^1}$		95.100 (0.216)	95.500 (0.166)	95.920 (0.105)	95.560 (0.080)	95.425 (0.063)
				2.300 (0.150)	1.950 (0.102)	1.840 (0.066)	2.240 (0.053)	2.295 (0.039)
				2.600 (0.147)	2.550 (0.109)	2.240 (0.072)	2.200 (0.055)	2.280 (0.040)
				1.681 (0.564)	1.803 (0.381)	1.813 (0.221)	1.811 (0.147)	1.811 (0.126)
		$u_E^{L^2}$		95.000 (0.218)	95.400 (0.167)	95.780 (0.107)	95.490 (0.081)	95.355 (0.063)
				2.300 (0.150)	1.950 (0.102)	1.900 (0.068)	2.270 (0.054)	2.325 (0.039)
				2.700 (0.147)	2.650 (0.109)	2.320 (0.073)	2.240 (0.055)	2.320 (0.041)
				1.677 (0.563)	1.796 (0.379)	1.805 (0.218)	1.803 (0.146)	1.802 (0.125)
$t_5(0,1)$	500	$u_E^{L^1}$		95.700 (0.203)	95.200 (0.165)	95.720 (0.107)	95.460 (0.081)	95.460 (0.062)
				1.900 (0.137)	2.400 (0.109)	2.140 (0.067)	2.270 (0.050)	2.330 (0.041)
				2.400 (0.173)	2.400 (0.116)	2.140 (0.075)	2.270 (0.053)	2.210 (0.037)
				1.673 (0.630)	1.784 (0.458)	1.805 (0.330)	1.810 (0.270)	1.812 (0.241)
		$u_E^{L^2}$		95.600 (0.205)	95.050 (0.167)	95.620 (0.108)	95.360 (0.083)	95.355 (0.063)
				2.000 (0.140)	2.550 (0.112)	2.200 (0.068)	2.310 (0.050)	2.385 (0.041)
				2.400 (0.153)	2.400 (0.114)	2.180 (0.071)	2.330 (0.054)	2.260 (0.038)
				1.667 (0.629)	1.773 (0.453)	1.792 (0.326)	1.797 (0.267)	1.798 (0.237)
	1000	$u_E^{L^1}$		95.400 (0.209)	95.500 (0.156)	95.460 (0.113)	95.480 (0.083)	95.440 (0.059)
				1.500 (0.122)	1.850 (0.097)	2.140 (0.073)	2.230 (0.055)	2.295 (0.040)
				3.100 (0.173)	2.650 (0.116)	2.400 (0.075)	2.290 (0.053)	2.265 (0.037)
				1.651 (0.693)	1.766 (0.497)	1.765 (0.307)	1.760 (0.204)	1.765 (0.182)
		$u_E^{L^2}$		95.600 (0.205)	95.050 (0.167)	95.620 (0.108)	95.360 (0.083)	95.355 (0.063)
				2.000 (0.140)	2.550 (0.112)	2.200 (0.068)	2.310 (0.050)	2.385 (0.041)
				2.400 (0.153)	2.400 (0.114)	2.180 (0.071)	2.330 (0.054)	2.260 (0.038)
				1.667 (0.629)	1.773 (0.453)	1.792 (0.326)	1.797 (0.267)	1.798 (0.237)

D Empirical illustration

Table D.1: MA(1) filter with dummies

constant	γ	D1	D2	D3
0.125	0.0561	-5.374	6.201	-7.147
(0.046)	(1.398)	(1.400)	(1.409)	0.0561

Robust standard errors are given in parentheses. D_1 takes value 1 on March 4, 1996. D_2 takes value 1 on September 23, 1998. D_3 takes value 1 on September 10, 1998.

Table D.2: Sample moments of residuals from MA(1) filter with dummies

Sample Size	Mean	SD	Skewness	Excess Kurtosis	Max	Min
1042	0.006	1.399	-0.270^*	3.100^*	6.052	-7.147

*Significant values at 5% level.

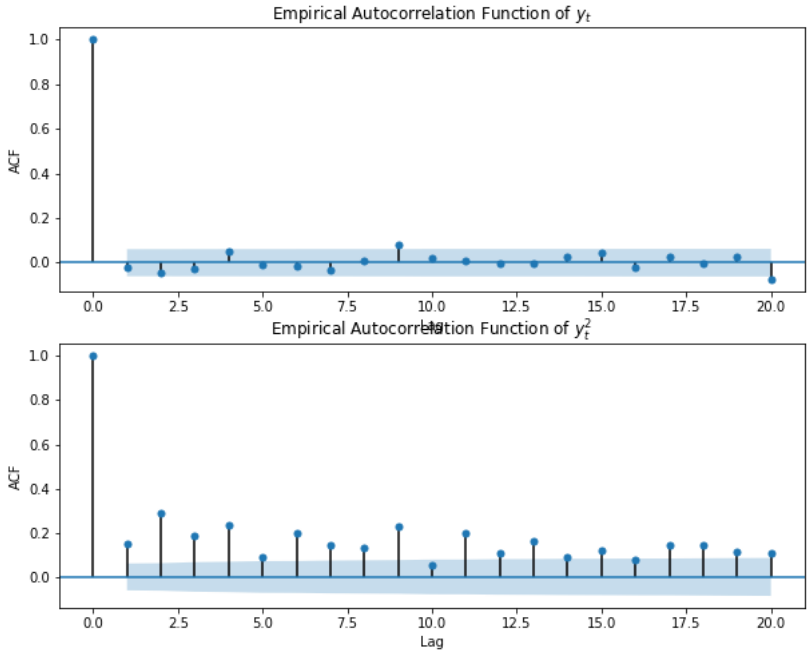

Figure D.1: Empirical autocorrelation functions of Y_t; respectively Y_t^2.

Table D.3: GARCH(1,1)

ω	α	β	$\alpha + \beta$
0.028	0.132	0.861	0.993
(0.014)	(0.025)	(0.024)	

Robust standard errors are given in parentheses.

Table D.4: Sample moments of centered residuals

Sample Size	Mean	SD	Skewness	Excess Kurtosis	Max	Min
1042	0.000	1.006	-0.176^*	0.741^*	4.615	-4.295

*Significant values at 5% level.

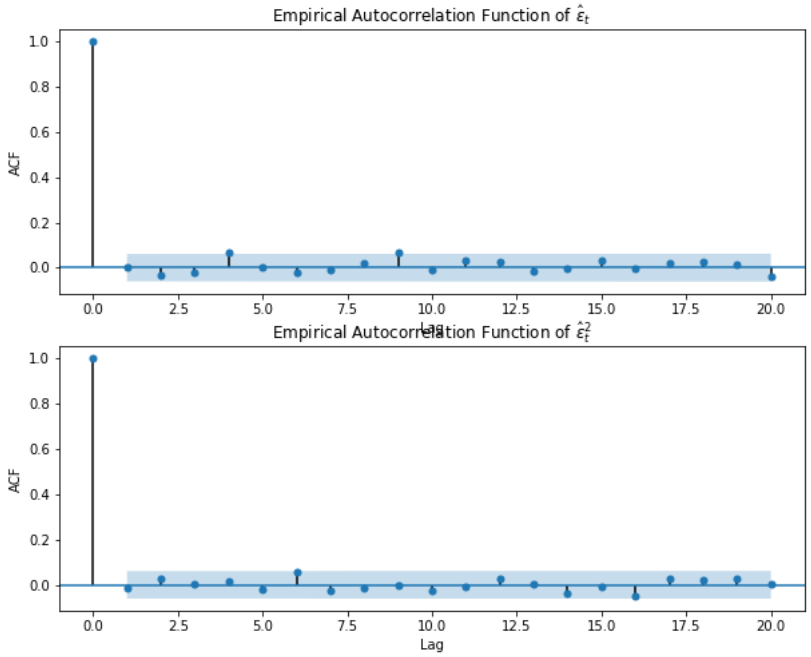

Figure D.2: Empirical autocorrelation function of $\hat{\epsilon}_t$; respectively $\hat{\epsilon}_t^2$.

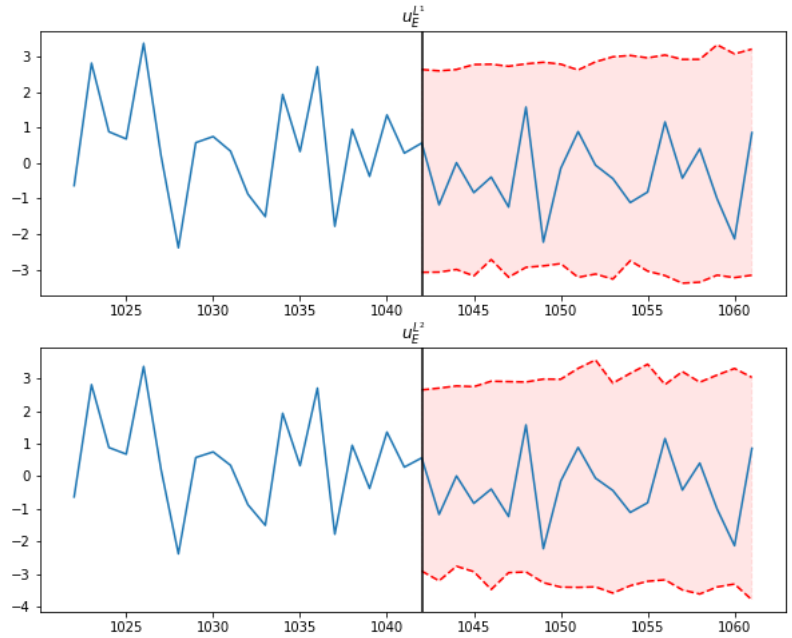

Figure D.3: 20-step ahead prediction intervals of future returns based on Epanechnikov kernel with bandwidth $u_E^{L^1}$ (upper); respectively $u_E^{L^2}$ (lower).

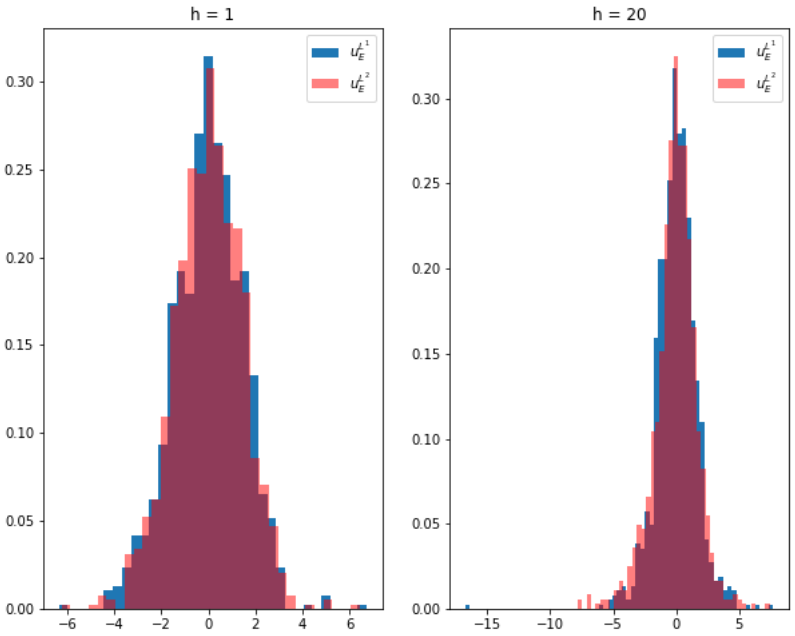

Figure D.4: Histogram of one- and 20-step ahead predictions of returns based on Epanechnikov kernel with bandwidth $u_E^{L^1}$; respectively $u_E^{L^2}$.

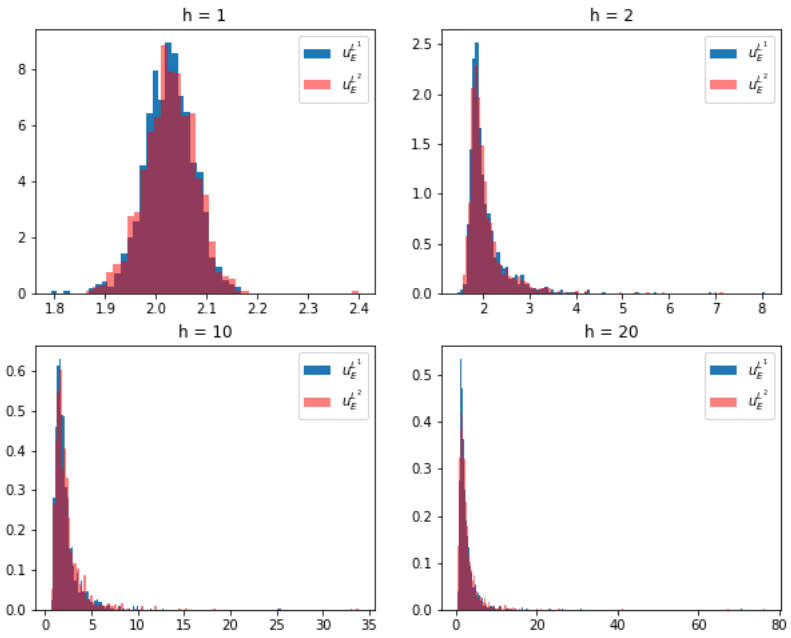

Figure D.5: Histogram of one- and 20-step ahead predictions of volatility based on Epanechnikov kernel with bandwidth $u_E^{L^1}$; respectively $u_E^{L^2}$.

Part II

Bootstrap Forecasts for the Poisson Autoregressive Model

Bootstrap Forecasts for the Poisson Autoregressive Model[†]

Philipp Christian Kless,
Department of Economics,
University of Copenhagen, Denmark.

Abstract

We present a parametric bootstrap scheme to forecast the Poisson autoregressive (PAR) model. More precisely, our bootstrap simulates the analytically unknown multi-step ahead probability mass function (pmf) via recursive one-step ahead predictions of future counts using the Poisson distribution and the estimated model parameters. By repeating this step a large number of times, we numerically approximate the pmf of future counts. We prove that our bootstrap forecasts are asymptotically valid. We study the finite sample properties of our forecasts by means of a Monte Carlo experiment. This experiment supports our theoretical results, that is, prediction intervals based on the forecasted pmf according to our bootstrap have the correct coverage on average. Finally, in two empirical applications, we demonstrate that our parametric bootstrap also improves the forecasting performance of the PAR model for stock transaction data and monthly US default count data.

Keywords: Poisson autoregression; Bootstrap; Forecasting

[†]I thank Anders Rahbek, Rasmus Søndergaard Pedersen, and Michael H. Neumann for helpful comments.

1 Introduction

In this paper, we discuss a fully parametric bootstrap to approximate the analytically unknown h-step ahead distribution of the Poissson autoregressive (PAR) model. This model class assumes that the conditional distribution of the observations follows a Poisson distribution where the intensity parameter varies over time. Applications of this model span various academic fields ranging from economics to medicine and focus on predicting future counts. For instance, Agosto, Cavaliere, Kristensen, and Rahbek (2016) predict time series counts of corporate defaults while Angelini and De Angelis (2017) implement the PAR model to predict the outcome of football matches to develop a profitable betting strategy. Moreover, Liboschik, Fokianos, and Fried (2017) use the PAR model to design a prediction-based monitoring procedure for infectious disease surveillance.

We contribute to the literature by proposing a simulation based forecasting approach for the PAR model. More precisely, we discuss a fully parametric bootstrap scheme. This parametric bootstrap can be applied to approximate the analytically unknown h-step ahead conditional distribution of the count series. Our method simulates multi-step ahead forecasts via recursive one-step ahead predictions of future counts using the Poisson distribution and the estimated intensity parameter. By creating a large number of these iterated multi-step ahead forecasts, we numerically approximate the probability mass function (pmf) of future counts. This simulated pmf is then used to create prediction intervals for the count time series. Our second contribution is that we develop asymptotic theory for the discussed bootstrap approach. That is, we give explicit conditions such that our simulated forecasts converge weakly to the unknown distribution of the future counts. We also evaluate our bootstrap approach in Monte Carlo simulations as well as in two empirical exercises with real data.

First, we study the finite-sample properties of our bootstrap based pmf forecasts in a controlled Monte Carlo setup. For this Monte Carlo experiment, we assume that the DGP is a PAR model. Then, we compare the actual coverage of our prediction intervals, based on the simulated pmf, with a chosen nominal coverage level such as 5%. For this setup, we show that bootstrap based prediction intervals have a better coverage than a closed form forecast method proposed in Agosto, Cavaliere, Kristensen, and Rahbek (2016).

Second, in two empirical exercises, we compute rolling one-step ahead prediction intervals based on our simulated pmf via the bootstrap and the closed from approach of Agosto, Cavaliere, Kristensen, and Rahbek (2016). The first data cover transactions of a stock per minute and is the same as in Fokianos, Rahbek, and Tjøstheim (2009). The second data contains monthly US default counts and is first analyzed in Agosto, Cavaliere, Kristensen, and Rahbek (2016). For both data sets, we find that the bootstrap prediction intervals always outperform the method of Agosto, Cavaliere, Kristensen, and Rahbek (2016) in terms of coverage.

In terms of existing literature, the PAR model is widely applied and therefore its stochastic properties are well understood. For instance, Ferland, Latour, and Oraichi (2006) give conditions such that the count process is stationary while mixing properties and weak dependence are studied in Neumann (2011) and Doukhan, Fokianos, and Tjøstheim (2012), among others. Moreover, the asymptotic behavior of its (quasi) maximum likelihood estimator ((Q)MLE) is also well established in the literature; cf. Fokianos, Rahbek, and Tjøstheim (2009), Ahmad and Francq (2016), and Agosto, Cavaliere, Kristensen, and Rahbek (2016).

Our parametric bootstrap approach for the PAR model complements existing studies. For instance, Hudecová, Hušková, and Meintanis (2015) propose a test that a given set of counts can be formulated as a particular series of counts with a given conditional distribution. Their test has power against the alternative of a different conditional distribution and even against the alternative of a different model class. In a Monte Carlo power study, they show that the bootstrap versions of the tests perform well. Moreover, Fokianos and Fried (2010) introduce tests to detect intervention effects that generate various types of outliers in count data. In particular, they apply a parametric bootstrap procedure based on the maximum of the different score test statistics to test for an intervention when the type and the time of the intervention are unknown. The practical relevance of the bootstrapped based test of Fokianos and Fried (2010) is confirmed using simulated and real data examples. For an alternative intervention testing procedure using the bootstrap see also Liboschik, Kerschke, Fokianos, and Fried (2016). Another application of the bootstrap is provided by Christou and Fokianos (2015a) who test the linearity of the mean process when a non-linear model contains nuisance parameters that are not identified under the null. The lack of identification affects their score based test such that classical asymptotic theory does not apply. To obtain critical values in this situation, they use a parametric bootstrap scheme based on the Poisson distribution. The empirical results demonstrate that their bootstrap works well. Finally, Fokianos and Neumann (2013) study a class of goodness-of-fit tests for parametric count time series models with power against local alternatives. In an effort to approximate the asymptotic null distribution of the statistical test, the authors advocate a parametric bootstrap method and prove its asymptotic validity. Their empirical applications demonstrates that their bootstrap works well in practice.

The article is organized as follows: In Section 2, we introduce the PAR model, discuss its properties and give some standard asymptotic results for the maximum likelihood estimator. Moreover, we discuss standard forecasting with the model. Next, in Section 3, we introduce our forecasting algorithm and prove its asymptotic validity. In Section 4, we report the results of our Monte Carlo experiment while Section 5 contains the empirical applications. Section 6 concludes the paper. Finally, the Appendix contains additional simulation results, empirical results, and all the proofs of the paper.

2 The Poisson autoregressive model

Suppose that Y_t is a time series of counts, e.g. the number of transactions per minute for a given stock. In the following, we wish to model the dynamics of this process in terms of its own past, $\{Y_{t-1}, Y_{t-2}, \ldots\}$. We achieve this by modeling Y_t as a conditional Poisson distribution with time-varying intensity parameter, λ_t, expressed a linear function of past counts and its own past. That is, we consider following model:

$$Y_t|\mathcal{F}_{t-1} \sim \text{Poisson}\left(\lambda_t\right), \ t = 1, \ldots, T, \tag{2.1}$$

where $\mathcal{F}_t$ represents the σ-field generated by $\{Y_0, \ldots, Y_t, \lambda_0, \ldots, \lambda_t\}$, that is, $\mathcal{F}_t = \sigma\left(Y_s, s \leq t\right)$ and Poisson (λ) is a Poisson distribution with parameter λ. We close the model by choosing a linear specification for λ_t,

$$\lambda_t = \omega + \alpha Y_{t-1} + \beta \lambda_{t-1}, \tag{2.2}$$

where $\omega > 0$ and $\alpha, \beta \in \mathbb{R}_+$ such that $\lambda_t > 0$ since Y_t is a non-negative integer. In the following, we refer this model by PAR(1,1). The parameters of the model are given by $\theta \coloneqq (\omega, \alpha, \beta)' \in \Theta \subset \mathbb{R}_+$. We let $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$ denote the true data-generating parameter value. In addition, we assume that Y_0 is fixed.

The chosen specification allows us to model the dynamics of the number of counts in terms of past counts which is captured by αY_{t-1} and $\beta \lambda_{t-1}$. This approach is similar to the standard (G)ARCH process applied to describe the evolution of the conditional variance of a continuously distributed variable; see Bollerslev (1986). Due to this similarity, the Poisson autoregressive model is also referred to as an integer valued (G)ARCH process in the literature; see Ferland, Latour, and Oraichi (2006).

2.1 Properties

In this section, we state sufficient conditions for the PAR model to be stationary, mixing and ergodic. These stochastic properties have been studied in detail in the literature; see, among others, Doukhan and Wintenberger (2008), Fokianos, Rahbek, and Tjøstheim (2009), Neumann (2011), Doukhan, Fokianos, and Tjøstheim (2012, 2013), and Agosto, Cavaliere, Kristensen, and Rahbek (2016). Most of the literature uses the concept of τ-dependence, henceforth weak dependence, as defined in Doukhan and Wintenberger (2008). Weak dependence is a stability concept for discrete-valued Markov chains which implies stationarity and ergodicity. Importantly, establishing weak dependence allows us to apply a LLN and a CLT to the studied process.

To establish weak dependence, we write the model in eq. (2.1) in terms of an i.i.d. sequence of Poisson processes with unit intensity. More precisely, for each t, let $N_t(\cdot)$ be

a Poisson process of unit intensity. Then, for any $u > 0$, the number of events $N_t(u)$ is distributed as a Poisson random variable with intensity u such that we can re-write eq. (2.1) as

$$Y_t = N_t(\lambda_t), \tag{2.3}$$

where $N_t(\cdot)$ is i.i.d. over time. We impose the following assumption on the model.

Assumption 2.1 *(i) The innovations $N_t(\cdot)$ are i.i.d. over time, and (ii) $\alpha + \beta < 1$.*

Assumption 2.1(i) allows us to embed Y_t in a Markov chain framework such that we can apply the theory of weak dependence. The second part of Assumption 2.1 states that the function $L(Y, \lambda) = \omega + \alpha Y + \beta\lambda$ is Lipschitz with Lipschitz coefficient $\alpha + \beta < 1$. This condition is standard in the literature; cf. Doukhan and Wintenberger (2008).

Finally, Assumption 2.1 implies that the PAR model has a stationary and weakly dependent solution, as proven in the next lemma.

Lemma 2.1 *Under Assumption 2.1, there exists a weakly dependent stationary and ergodic solution to eq.* (2.1) *for which $E(|Y_t|) < \infty$.*

As a result of Lemma 2.1, we can use the LLN for stationary and ergodic processes that will be needed in the next section for deriving the large sample properties of the below proposed maximum likelihood estimator.

2.2 Estimation

Next, we discuss estimation of the PAR model by means of maximum likelihood. Moreover, we outline the well understood asymptotic theory for the estimated parameters; see, among other, Fokianos, Rahbek, and Tjøstheim (2009), Agosto, Cavaliere, Kristensen, and Rahbek (2016), and Ahmad and Francq (2016).

We consider the PAR model for Y_t as given in eq. (2.1), that is, the conditional intensity is given by

$$\lambda_t(\theta) = \omega + \alpha Y_{t-1} + \beta\lambda_{t-1}(\theta). \tag{2.4}$$

Then, the conditional log-likelihood function of θ in terms of the observations, $\{Y_1, \ldots, Y_T\}$, given initial values, $\{Y_0, \lambda_0\}$, is given by

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta), \; l_t(\theta) \coloneqq Y_t \log \lambda_t(\theta) - \lambda_t(\theta), \tag{2.5}$$

where any constant terms are omitted. The maximum likelihood estimator (MLE) is then defined as

$$\hat{\theta} \coloneqq \arg\max_{\theta \in \Theta} L_T(\theta). \tag{2.6}$$

The large sample properties of $\hat{\theta}$ are well-understood in the literature. We follow Ahmad and Francq (2016) and assume that the following conditions hold throughout the rest of this paper.

Assumption 2.2 *The parameter space is given by* $\Theta = [\underline{\omega}, \overline{\omega}] \times [\underline{\alpha}, \overline{\alpha}] \times [\underline{\beta}, \overline{\beta}]$ *for some* $0 < \underline{\omega} < \overline{\omega} < \infty$, $0 < \underline{\alpha} < \overline{\alpha} < \infty$, $0 < \underline{\beta} < \overline{\beta} < 1$.

Assumption 2.3 *The conditional distribution of* Y_t *is not degenerated.*

Assumption 2.2 restricts the parameter space and ensures, for instance, that λ_t is bounded away from zero by assuming that $\underline{\omega} > 0$. Assumption 2.3 ensures identification of the parameters. Both assumptions are standard in the literature.

Under the above assumptions, together with the Assumption 2.1 used to show stationarity and existence of moments, we obtain the following asymptotic result for the MLE; see Fokianos, Rahbek, and Tjøstheim (2009), and Ahmad and Francq (2016).

Theorem 2.1 *Suppose that Assumptions 2.1, 2.2, and 2.3 hold, then*

$$\hat{\theta} \to_{a.s.} \theta_0 \text{ as } T \to \infty.$$

If, in addition, θ_0 *is an interior point of* Θ, *then*

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \to_d \mathcal{N}(0, \Sigma), \text{ as } T \to \infty,$$

where Σ *is some positive definite matrix.*

Remark 2.1 *Under the conditions of Theorem 2.1, Ahmad and Francq (2016) show that the asymptotic variance of MLE can be consistently estimated by* $\hat{\Sigma} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$ *with*

$$\hat{J} \coloneqq \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\lambda_t\left(\hat{\theta}_t\right)} \frac{\partial \lambda_t\left(\hat{\theta}\right)}{\partial \theta} \frac{\partial \lambda_t\left(\hat{\theta}\right)}{\partial \theta'}, \tag{2.7}$$

$$\hat{I} \coloneqq \frac{1}{T}\sum_{t=1}^{T} \left(\frac{Y_t}{\lambda_t\left(\hat{\theta}\right)} - 1\right)^2 \frac{\partial \lambda_t\left(\hat{\theta}\right)}{\partial \theta} \frac{\partial \lambda_t\left(\hat{\theta}\right)}{\partial \theta'}. \tag{2.8}$$

Once the model parameters have been estimated, we can use them to forecast future number of counts with the PAR model in eq. (2.1).

2.3 Forecasting

In this section, we outline how we can obtain multi-step ahead forecast of the count process, Y_t, and its distribution without relying on bootstrap based simulations. This approach is first discussed in Agosto, Cavaliere, Kristensen, and Rahbek (2016); henceforth ACKR(2016). More precisely, we apply a two step procedure: At first, we obtain a multi-step ahead forecast of the time-varying intensity parameter, λ_t. Then, this point forecast is inserted into the conditional distribution of the count to process to obtain prediction intervals, for instance.

In terms of the mean square error, the optimal one-step ahead forecast of λ_t, given the information available at time T and the model parameters, is

$$\lambda_{T+1|T}(\theta) = \omega + \alpha Y_T + \beta \lambda_T(\theta). \tag{2.9}$$

In general, we generate a multi-step ahead forecast of λ_{T+h}, for $h > 1$, by realizing that for any $k \geq 1$ the conditional intensity for λ_{T+k} can be represented as

$$\lambda_{T+k}(\theta) = \omega + (\alpha + \beta)\lambda_{T+k-1}(\theta) + \alpha\eta_{T+k-1},$$

where $\eta_t := Y_t - \lambda_t(\theta)$ such that $\mathrm{E}\left(\eta_t | \mathcal{F}_{t-1}^Y\right) = 0$. We then form a multi-step ahead forecast of $\lambda_{T+h|T}$ recursively through

$$\lambda_{T+k|T}(\theta) = \omega + (\alpha + \beta)\lambda_{T+k-1|T}(\theta), \ k = 2, \ldots, h, \tag{2.10}$$

with $\lambda_{T+1|T}(\theta)$ being defined in eq. (2.9). Finally, given a point forecast of the intensity, $\lambda_{T+h|T}$, we generate a forecast distribution of Y_{T+h} by

$$P(Y_{T+h} = y | \mathcal{F}_T) = \frac{\lambda_{T+h|T}^y(\theta) \exp\left(-\lambda_{T+h|T}(\theta)\right)}{y!}, \ y \in \{0, 1, 2, \ldots\}. \tag{2.11}$$

The before outlined approach is closely related to the concept of density forecasts; cf. Tay and Wallis (2000). The only difference is that we are concerned with a discrete valued-distribution. An easy way to illustrate the forecast distribution is by stating the $100(1-p)\,\%$ prediction interval implied by the forecast distribution for some $p \in (0, 1)$. That is, the $(1-p)$ prediction interval has following form:

$$\left[\mathrm{Q}\left(p/2 | \lambda_{T+h|T}(\theta)\right), \mathrm{Q}\left((1-p/2) | \lambda_{T+h|T}(\theta)\right)\right], \tag{2.12}$$

where $q \mapsto \mathrm{Q}(q|\lambda)$ is the quantile function of a Poisson distribution with intensity λ.

Observe that the expression in eq. (2.11) holds only for $h = 1$. The true distribution of Y_{T+h}, given the information available at time T, is in general analytically unknown for

$h > 1$. Therefore eq. (2.11) can only be regarded as an approximation of the probability mass function for more than one-step ahead. As an alternative, we can simulate the pmf by implementing our parametric bootstrap scheme as detailed in the next section.

3 Parametric Bootstrap Forecasts

As discussed before, the h-step ahead conditional distribution of the count process, $P(Y_h = y|Y_T = x, \lambda_T = \lambda)$ is analytically unknown for forecasting horizons with $h > 1$. As a solution, we can simulate multi-step ahead forecasts via recursive one-step ahead predictions of future counts using the Poisson distribution and the estimated model parameters. Then, we create a large number of these recursive multi-step ahead forecasts and use them to numerically approximate the unknown probability mass function of the count process.

We summarize the above outlined bootstrap in the following algorithm:

Algorithm A:

(i) Obtain $\hat{\theta}$ in eq. (2.1) on the original sample $\{Y_t\}_{t=1}^{T}$ using ML estimation.

(ii) Given the information available at time T, construct forecasts recursively through

$$\begin{aligned} \lambda^*_{T+h}\left(\hat{\theta}\right) &= \hat{\omega} + \hat{\alpha} Y^*_{T+h-1} + \hat{\beta}\lambda^*_{T+h-1}\left(\hat{\theta}\right), \\ Y^*_{T+h} &= N_{T+h}\left(\lambda^*_{T+h}\left(\hat{\theta}\right)\right), \; h = 1, \ldots, H, \end{aligned} \tag{3.1}$$

where $Y^*_T = x$, $\lambda^*_T = \lambda$, and $N_t(\cdot)$ is defined as before.

(iii) Obtain a set of replicates for $\lambda^{*(i)}_{T+h}\left(\hat{\theta}\right)$ and $Y^{*(i)}_{T+h}$ by repeating step (ii), say, $i = 1, \ldots, B$ times.

(iv) Calculate $\mathrm{P}^*\left(Y^*_{T+h} = y|Y^*_T = x, \lambda^*_T = \lambda\right) = \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}\left(Y^{*(i)}_{T+h} = y\right)$ for $y \in \mathbb{N}_0$.

Remark 3.1 *Step (iv) of Algorithm A generates point forecasts of the probability mass function, which can be used to construct prediction intervals for the future number of counts, Y_{T+h}.*

Remark 3.2 *In practice, we set $Y^*_T = Y_T = x$ to initialize Algorithm A in the case where $\beta = 0$.*

3.1 Asymptotic Properties of the Bootstrap

Next, we state the asymptotic validity of our parametric PAR bootstrap introduced in Algorithm A. In the following, we split our results in two parts. The first result states

the validity of Algorithm A for the PAR(1,1) while the second result demonstrates the validity of the PAR(1,0) process with $\beta = 0$.

At first, we turn to the case where $\beta > 0$ such that we look at the PAR(1,1) process. For this specification the one-step ahead intensity is

$$\lambda_{T+1}(\theta) = \omega + \alpha N_T(\lambda) + \beta\lambda, \tag{3.2}$$

where we use that $Y_T = N_T(\lambda_{t-1})$ with $\lambda_{t-1} = \lambda$, where λ is fixed. This expression leads directly to the following result for the simulated one step ahead pmf.

Theorem 3.1 *Suppose that Assumptions 2.1, 2.2, and 2.3 hold. Let* $P^*\left(Y^*_{T+1} = y | N_T(\lambda) = x\right)$ *be generated as detailed in Algorithm A. Then*

$$\left| P\left(Y_{T+1} = y | N_T(\lambda) = x\right) - P^*\left(Y^*_{T+1} = y | N_T(\lambda) = x\right) \right| \to_p 0,$$

as $T \to \infty$.

Next, we set $\beta = 0$ such that we are looking at the case where the intensity does not depend on its own past values. This specification makes λ_t a Markov Chain which is exploited to show the following result.

Theorem 3.2 *Suppose that Assumptions 2.1, 2.2, and 2.3 hold. Assume that* $\beta = 0$ *and let* $P^*\left(Y^*_{T+h} = y | Y^*_T = x\right)$ *be generated as detailed in Algorithm A. Then*

$$\left| P\left(Y_{T+h} = y | Y_T = x\right) - P^*\left(Y^*_{T+h} = y | Y^*_T = x\right) \right| \to_p 0,$$

as $T \to \infty$.

Remark 3.3 *We conjecture that the results in Theorem 3.1 can be extended to allow for* $h > 1$. *This conjecture is confirmed by simulations results in Section 4. The theoretical analysis of simulated multi-step ahead forecasts is beyond the scope of this paper and left for future research.*

4 Monte Carlo Experiments

4.1 Algorithm A

In this section, we verify the performance of our Algorithm A by means of a Monte Carlo experiment. As an alternative to our algorithm, we also implement the approach suggested by ACKR(2016); cf. eq. (2.12). For the Monte Carlo experiment we repeat the following steps:

1. Generate a time series, $\{Y_t\}_{t=1}^{T+H}$, for the PAR model given in eq. (2.1).

2. For this generated series, apply Algorithm A and the procedure of ACKR(2016) to calculate prediction intervals with nominal coverage of 90%, 95%, and 99%.

We repeat these two steps N times such that we have a set of N prediction intervals with some given nominal coverage for both methods. From this set we can empirically access the actual coverage of the prediction intervals from Algorithm A and ACKR(2016). That is, we simply compute the time series average of the number of times the true Y_{T+h} lies outside the prediction interval. Then, we take the average of this number over the N replications. In this experiment we set $N = 1000$ and use $B = 1000$ repetitions in Algorithm A to simulate the prediction intervals.

We first look at the case where $\beta = 0$. Here, we consider sample sizes, $T \in \{200, 500, 1000\}$, and we generate the time series for four vales of $\alpha \in \{0.6, 0.7, 0.8, 0.9\}$ with $\omega = 4.2$.

We present the results in Table 4.1. Our main findings are as follows: First, the method of ACKR(2016) performs worse than Algorithm A over all considered scenarios. Second, the performance of the method in ACKR(2016) decreases as the persistence of the count process increases. Third, the actual coverage of Algorithm A tends to be a slightly larger than the nominal coverage.

Looking at Figure 4.1, we can explain why the method of ACKR(2016) is not performing as good as Algorithm A. In this figure, we plot the true simulated pmf and compare it to the pmf's obtained via Algorithm A and the method of ACKR(2016). The true pmf is simulated using 1000 repetitions while $T = 1000$, $\alpha = 0.8$, and $\omega = 4.2$. We observe that forecasts based on ACKR(2016) can only approximate the true pmf for $h = 1$ reasonably well. For larger forecasting horizons, the approximation based on ACKR(2016) performs worse because it relies on the pmf of the Poisson distribution even though we do not analytically know the pmf of the count process for $h > 1$. Hence, the method of ACKR(2016) cannot model the heavier tails of the future count process in contrast to Algorithm A. As we see from Figure 4.1, Algorithm A is quite accurate and can capture the tails of the count process.

Next, we look at the situation where the DGP for the count series follows a negative binomial distribution. For the estimation step and forecasting exercise we maintain the assumption that the conditional distribution is Poisson. Hence, we are in a situation where our model is misspecified. This type of misspecification plays no role for the estimation step since our estimates are consistent even though we use the Poisson distribution; cf. Ahmad and Francq (2016). Yet, we use the wrong distribution when we construct prediction intervals. Specifically, we assume that $Y_t|\mathcal{F}_{t-1} \sim NB\left(r, r/(\lambda_t + r)\right)$ with $r > 0$ and $\lambda_t = \omega + \alpha Y_{t-1}$. Here, we set $\alpha \in \{0.6, 0.7, 0.8\}$ while $\omega = 4.2$ with $r = 10$ for $T \in \{200, 500, 1000\}$. We refer to this specification as NBAR(1). The results for this setting are available in Table 4.2 and Appendix B. From these tables we note that neither Algorithm A nor the procedure of ACKR(2016) perform well. That is, the actual coverage is well below the nominal coverage for all specifications.

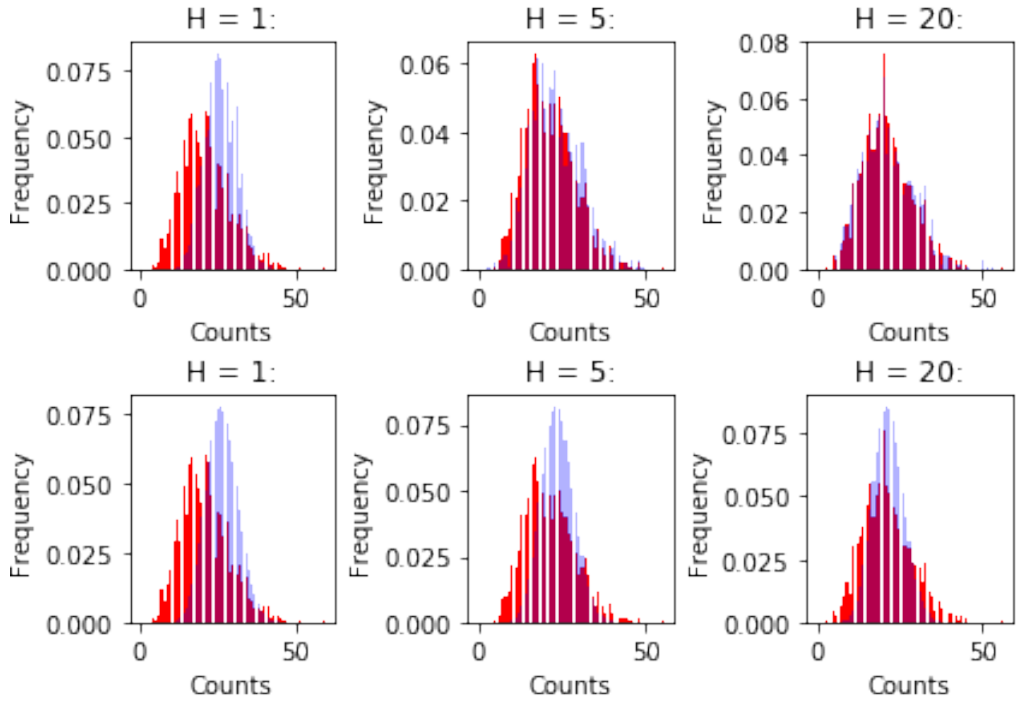

Figure 4.1: PAR(1): The true simulated pmf (red) is compared to pmf's obtained via Algorithm A (top row); respectively ACKR(2016) (bottom row) for $h \in \{1, 5, 20\}$. The sample size is $T = 1000$ and $\alpha = 0.8$ while $\omega = 4.2$.

We also consider the situation where the DGP is a PAR(1,1) model with $\alpha = 0.2$ and $\beta = \{0.6, 0.7\}$ while $\omega = 4.2$ and $T \in \{200, 500, 1000\}$. The values for α and β are chosen such that they reflect results often found in the literature where the sum of the two coefficients is close to 1; see Fokianos, Rahbek, and Tjøstheim (2009). Then we follow the outline from before and obtain prediction intervals with 90%, 95%, and 99% coverage for both forecasting methods. The results are available in Appendix B and allow for the following conclusions: First, Algorithm A performs better than ACKR(2016) in all settings. Especially for longer forecasting horizons the performance of Algorithm A is better than that of ACKR(2016). Second, compared to the PAR(1) model as DGP, the method of ACKR(2016) shows an increased performance and often provides an actual coverage close to nominal.

Finally, we conduct the same robustness check as before. That is, we study the case where the DGP of the count process follows a negative binomial distributions. We set $\alpha = 0.2$, $\beta = 0.7$ while $\omega = 4.2$, $T \in \{200, 500, 1000\}$, and $r = 10$. The results of this robustness check are relegated to Appendix B since its findings are qualitatively the same as before.

Table 4.1: PAR(1) model: Prediction intervals with 95% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

α	T	Method	h			
			1	5	10	20
0.6	200	Algorithm A	0.972 (0.165)	0.974 (0.089)	0.973 (0.065)	0.969 (0.051)
		ACKR(2016)	0.968 (0.176)	0.923 (0.142)	0.915 (0.112)	0.911 (0.080)
	500	Algorithm A	0.979 (0.143)	0.971 (0.089)	0.970 (0.067)	0.970 (0.050)
		ACKR(2016)	0.976 (0.153)	0.928 (0.140)	0.914 (0.115)	0.910 (0.086)
	1000	Algorithm A	0.983 (0.129)	0.971 (0.096)	0.969 (0.071)	0.969 (0.050)
		ACKR(2016)	0.973 (0.162)	0.929 (0.146)	0.913 (0.113)	0.907 (0.081)
0.7	200	Algorithm A	0.973 (0.162)	0.968 (0.098)	0.967 (0.080)	0.965 (0.064)
		ACKR(2016)	0.964 (0.186)	0.910 (0.164)	0.891 (0.137)	0.877 (0.110)
	500	Algorithm A	0.964 (0.186)	0.963 (0.108)	0.967 (0.075)	0.968 (0.056)
		ACKR(2016)	0.957 (0.203)	0.896 (0.178)	0.881 (0.141)	0.875 (0.109)
	1000	Algorithm A	0.967 (0.179)	0.969 (0.097)	0.969 (0.072)	0.966 (0.057)
		ACKR(2016)	0.955 (0.207)	0.905 (0.170)	0.888 (0.140)	0.877 (0.108)
0.8	200	Algorithm A	0.966 (0.181)	0.961 (0.110)	0.957 (0.094)	0.958 (0.077)
		ACKR(2016)	0.959 (0.198)	0.863 (0.218)	0.832 (0.189)	0.810 (0.154)
	500	Algorithm A	0.971 (0.168)	0.964 (0.110)	0.963 (0.088)	0.963 (0.070)
		ACKR(2016)	0.965 (0.184)	0.865 (0.208)	0.835 (0.185)	0.812 (0.149)
	1000	Algorithm A	0.962 (0.191)	0.962 (0.110)	0.959 (0.098)	0.958 (0.080)
		ACKR(2016)	0.956 (0.205)	0.874 (0.198)	0.833 (0.185)	0.810 (0.149)
0.9	200	Algorithm A	0.960 (0.196)	0.952 (0.135)	0.949 (0.119)	0.938 (0.122)
		ACKR(2016)	0.957 (0.203)	0.821 (0.246)	0.746 (0.247)	0.684 (0.222)
	500	Algorithm A	0.963 (0.189)	0.956 (0.130)	0.959 (0.110)	0.954 (0.096)
		ACKR(2016)	0.956 (0.205)	0.832 (0.238)	0.763 (0.232)	0.699 (0.209)
	1000	Algorithm A	0.965 (0.184)	0.961 (0.123)	0.950 (0.121)	0.950 (0.105)
		ACKR(2016)	0.964 (0.186)	0.840 (0.229)	0.759 (0.240)	0.698 (0.218)

Table 4.2: NBAR(1) model: Prediction intervals with 95% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

α	T	Method	h			
			1	5	10	20
0.6	200	Algorithm A	0.870 (0.336)	0.855 (0.195)	0.856 (0.145)	0.856 (0.103)
		ACKR(2016)	0.864 (0.343)	0.793 (0.222)	0.785 (0.168)	0.779 (0.123)
	500	Algorithm A	0.855 (0.352)	0.852 (0.193)	0.860 (0.139)	0.861 (0.103)
		ACKR(2016)	0.851 (0.356)	0.793 (0.215)	0.790 (0.163)	0.785 (0.120)
	1000	Algorithm A	0.882 (0.323)	0.873 (0.179)	0.872 (0.130)	0.865 (0.099)
		ACKR(2016)	0.883 (0.321)	0.814 (0.209)	0.802 (0.155)	0.786 (0.117)
0.7	200	Algorithm A	0.833 (0.373)	0.827 (0.225)	0.828 (0.170)	0.826 (0.126)
		ACKR(2016)	0.828 (0.377)	0.743 (0.257)	0.716 (0.199)	0.695 (0.152)
	500	Algorithm A	0.825 (0.380)	0.826 (0.221)	0.829 (0.166)	0.831 (0.124)
		ACKR(2016)	0.829 (0.377)	0.738 (0.256)	0.717 (0.192)	0.706 (0.143)
	1000	Algorithm A	0.819 (0.385)	0.831 (0.212)	0.833 (0.165)	0.829 (0.119)
		ACKR(2016)	0.820 (0.384)	0.736 (0.245)	0.713 (0.195)	0.698 (0.139)
0.8	200	Algorithm A	0.763 (0.425)	0.738 (0.297)	0.733 (0.249)	0.735 (0.208)
		ACKR(2016)	0.799 (0.401)	0.645 (0.297)	0.593 (0.240)	0.568 (0.191)
	500	Algorithm A	0.714 (0.452)	0.716 (0.308)	0.721 (0.266)	0.723 (0.231)
		ACKR(2016)	0.761 (0.426)	0.636 (0.291)	0.600 (0.245)	0.581 (0.202)
	1000	Algorithm A	0.689 (0.463)	0.678 (0.347)	0.681 (0.309)	0.685 (0.284)
		ACKR(2016)	0.794 (0.404)	0.667 (0.304)	0.629 (0.259)	0.609 (0.218)

5 Empirical applications

To further illustrate the performance of Algorithm A, we look at two empirical applications. First, we use the data set from Fokianos, Rahbek, and Tjøstheim (2009) which contains the number of transactions per minute for the stock Ericsson B during July 2, 2002. Second, we use the data from ACKR(2016) which cover monthly US corporate default counts from 1982 through 2011.

5.1 Ericsson B Data

The first data set contains the number of transactions per minute for the stock Ericsson B during July 2, 2002. The data set has 460 observations which cover eight hours of trading. Note that the first and last minutes of transactions are not taken into account. The data are also used in Fokianos, Rahbek, and Tjøstheim (2009) to illustrate the empirical performance of PAR(1,1) model. Here, we estimate at first the model parameters. Then, we use the method of ACKR(2016) and Algorithm A to construct out-of-sample forecasts for the data.

When looking at Figure 5.1, we observe that the number of trades are highly autocorrelated. Moreover, the data displays overdispersion since its mean is 9.9 while its sample variance equals 32.8. In a next step, we fit a PAR(1,1) model as in Fokianos, Rahbek, and Tjøstheim (2009) to the data at hand. The parameter estimates are given in Table and are comparable with Fokianos, Rahbek, and Tjøstheim (2009). We note that $\hat{\alpha} + \hat{\beta}$ are close to unity which is comparable to the unit-root phenomenon in autoregressive time series modeling. For instance, this behavior is similar to the so-called IGARCH case where there is high-persistence in the conditional variance.

Finally, Figure 5.2 plots diagnostic tools to help us judge the fit of the PAR(1,1) model. First, the upper left part of this figure illustrates that the predicted values defined as $\hat{Y}_t = \hat{\lambda}_t(\hat{\theta})$ are a good approximation of the observed transactions per minute. Second, the upper right part of Figure 5.2 plots the sample autocorrelation function of the Pearson residuals which are defined as $\hat{\epsilon}_t = \left(Y_t - \lambda_t(\hat{\theta})\right) / \sqrt{\lambda_t(\hat{\theta})}$. Under a correct model, the Pearson residuals should be a white noise sequence with constant variance as seen in the figure. Third, in the bottom left part of Figure 5.2, we see the probability integral transform (PIT) which is not uniformly distributed on $[0, 1]$. A deviation from a uniform distribution over the unity interval indicates that the assumption of a Poisson distribution might be violated; cf. Liboschik, Fokianos, and Fried (2017). Finally, we plot the marginal calibration in the bottom right of Figure 5.2. For a well specified model, we expect that the marginal distribution of the fitted values resembles the marginal distribution of the observations. Major deviations from zero indicate a misspecified model; cf. Christou and Fokianos (2015b). Hence, we conclude that the PAR(1,1) model successfully removes the

temporal dependence, but the Poisson assumption is likely to be violated.

Next, we perform a pseudo-out-of-sample forecasting exercise for the PAR(1,1) model. That is, we split the data set in two parts where the first part of size $T_0 = 300$ is used for initial estimation of the PAR(1,1) model. The remaining data Y_t with $t = T_0 + 1, \ldots, T$ is retained for the forecasting exercise outlined next.

First, we define

$$\hat{\theta}_t = \arg\max_{\theta} L_t(\theta)$$

to be the MLE using data until $t \geq T_0$, where $L_t(\theta) = \sum_{s=1}^{t} l_s(\theta)$ with $l_s(\theta) = y_s \log \lambda_s(\theta) - \lambda_s(\theta)$.

Then, we use $\hat{\theta}$ and compute the corresponding one-step ahead prediction interval based on Algorithm A. In addition, we forecast using on the approach of ACKR(2016); cf. eq. (2.12). Next, we repeat this exercise for $t = T_0 + 1, \ldots, T$ such that we create a series of prediction intervals. That is, we obtain $\{L_{t+1|t}(p), U_{t+1|t}(p)\}$ for $t = T_0 + 1, \ldots, T$, where $L_{t+1|t}(p)$, and $U_{t+1|t}(p)$ are the lower and upper limits of the prediction intervals for the coverage probability, p. This rolling approach is comparable to the setup in ACKR(2016) and mimics a situation where a forecaster starts at time T_0. Then, as new observations arrive, the forecaster updates his estimates and obtains new forecasts. The two series are plotted in Figure 5.3 for a coverage probability of 95%. Upon visual inspection, both procedures for constructing prediction intervals appear to perform equally well.

Next, we use the $\{L_{t+1|t}(p), U_{t+1|t}(p)\}$, $t = T_0+1, \ldots, T$, to calculate the actual coverage and evaluate the intervals via standard backtesting methods. The first test we apply is the unconditional coverage (UC) test by Kupiec (1995). The second method is the conditional coverage test (CC) by Christoffersen (1998). The results are available in Table 5.2 and allow for the following conclusions: First, we conclude that the prediction intervals based on Algorithm A always have a coverage closer to nominal than the ones based on ACKR(2016). Second, none of the two methods passes any of the formal statistical tests. This result can be explained by the fact that both procedures rely on the Poisson distribution to create the prediction intervals. However, as evident from the PIT and marginal calibration plot in Figure 5.2, the Poisson distribution does not provide a good fit for the data. And as a result, forecasts based on this assumption will also perform poorly. This observations is also in line with our results from the Monte Carlo experiment.

5.2 US corporate default count

Our second data example looks at the data set used in Agosto, Cavaliere, Kristensen, and Rahbek (2016) which contains monthly observations of US cooperate default counts. More precisely, the data contains the monthly number of bankruptcies among Moody's rated industrial firms in the United States from 1982 until 2011 which gives us 360 observations.

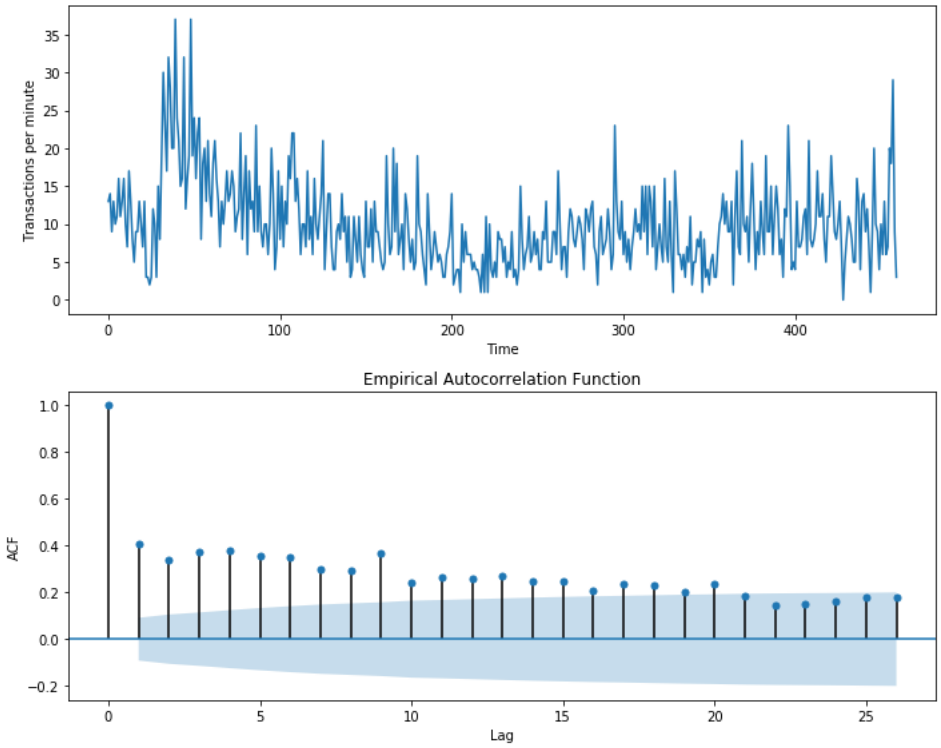

Figure 5.1: Top: The number of transactions per minute for the stock Ericsson B during July 2, 2002. Bottom: Empirical autocorrelation function of the number of transactions.

Table 5.1: PAR(1,1) estimates for the Ericsson B data obtained via MLE.

ω	α	β	$\alpha+\beta$
0.581	0.198	0.744	0.943
(0.277)	(0.034)	(0.052)	

Robust standard errors are given in parentheses.

Table 5.2: Coverage of the one-step ahead prediction intervals and p-values for Unconditional Coverage (UC) and Conditional Coverage (CC) Test for the Ericsson B data.

		Coverage probability, p		
Method		90%	95%	99%
ACKR(2016)	Coverage	0.812	0.843	0.906
	UC	0.000	0.000	0.000
	CC	0.001	0.000	0.000
Algorithm A		0.843	0.868	0.918
		0.014	0.000	0.000
		0.034	0.000	0.000

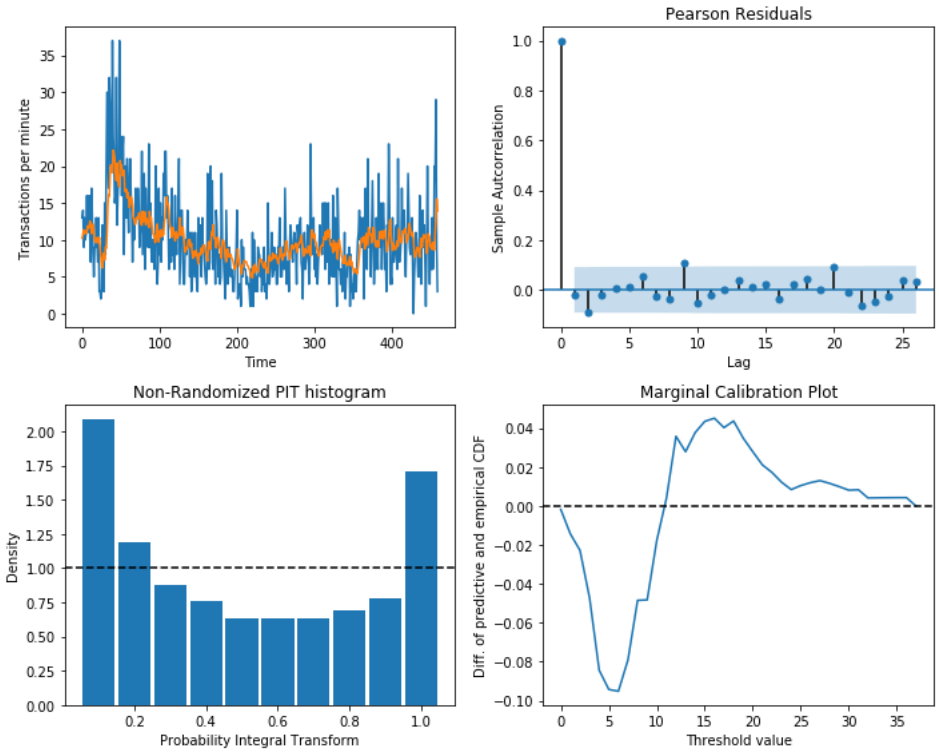

Figure 5.2: Top left: Fitted values and observed process. Top right: Sample autocorrelation of the Pearson Residuals. Bottom left: Probability Integral Transform. Bottom Right: Marginal Calibration Plot.

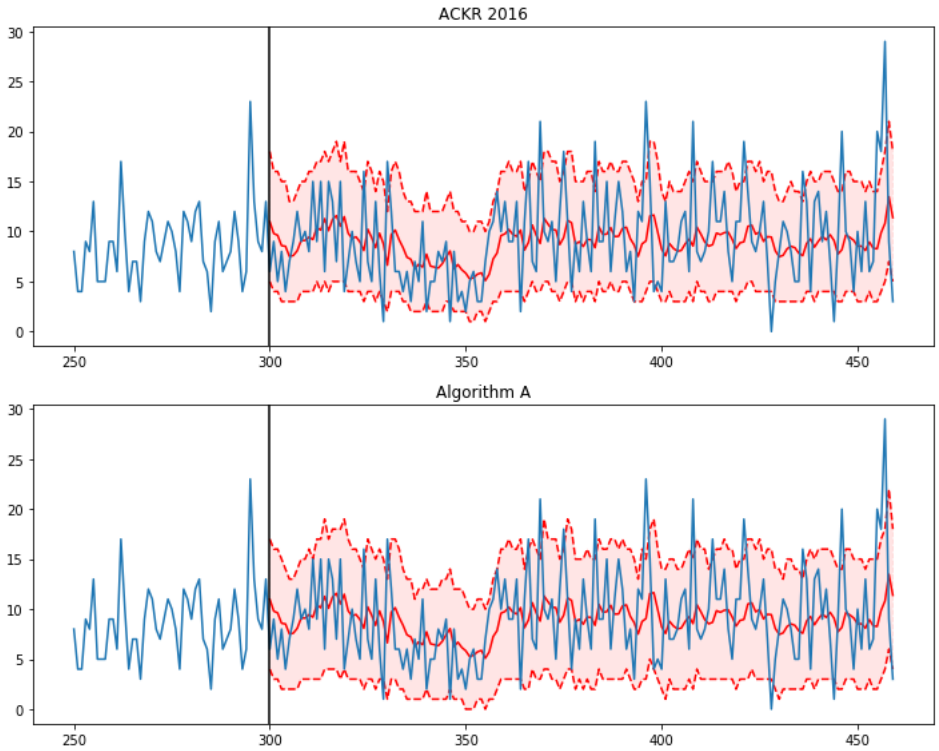

Figure 5.3: Top: One-step ahead prediction intervals based on ACKR(2016) with nominal coverage of 95%. Bottom: One-step ahead prediction intervals based on Algorithm A with nominal coverage of 95%.

The data is collected from Moody's Credit Risk Calculator. As with the transaction data, we see from Figure 5.4 that the default counts are highly correlated and cluster over time. Moreover, the distribution of the default counts is over-dispersed with a mean of 3.51 while the empirical variance is 15.57.

In ACKR(2016), the authors also include exogenous regressors in the intensity specification in their empirical analysis. We denote this specification by PARX in the following. Adding exogenous regressors allows the authors to study to what extent autocorrelation and clustering of the default counts depend on common risk factors. Among the covariates are financial, credit market, and macroeconomic variables such the Leading Index released by the Federal Reserve denoted by LI. Moreover, ACKR(2016) also consider realized volatility (RV) on the S&P 500 to take into the impact of uncertainty in financial markets on default counts. For more details see ACKR(2016).

The analysis of ACKR(2016) arrives at two preferred models which are a PAR(2,1) model and a PARX(2,1) specification which includes RV and LI as exogenous covariates. The parameter estimates are obtained via MLE and reproduced in Table 5.3. From this table we note that $\hat{\alpha}_1+\hat{\alpha}_2$ decreases for the specification with covariates. This observation shows that covariates are important in explaining the dependence of default counts over time. However, a strong link between the conditional intensity and past default counts remains since $\hat{\alpha}_1 + \hat{\alpha}_2$ is not close to zero which would imply conditional independence of default counts over time. In terms of misspecification, we turn to Figure 5.5 and Figure 5.6 which plot the fitted values, the sample autocorrelation function of the Pearson Residuals, the PIT and the marginal calibration plot for the corresponding model. Both specifications successfully remove the temporal dependence according to the autocorrelation function of the residuals. Moreover, in this example it appears that the Poisson distribution is a reasonable assumption when looking at the PIT plot and marginal calibration plot for both models.

As in the previous section, we continue with a pseudo-out-of-sample forecasting exercise for the PAR(2,1) and PARX(2,1) model. We split the data into two parts where the first part is of size $T_0 = 200$ and is used for initial estimation. The remaining data Y_t for $t = T_0 + 1, \ldots, T$ is reserved for the forecasting exercise. For this exercise, we stick to the same approach as in the section before. That is, we estimate $\hat{\theta}$ based on a rolling window and then obtain one-step ahead prediction intervals for both model specifications using the method of ACKR(2016) and Algorithm A. We repeat this exercise for $t = T_0+1, \ldots, T$ such that we get a series of prediction intervals and intensity forecasts. These series are plotted in Figure 5.7 and 5.8, respectively. Upon visual inspection, both models and both forecasting procedures perform equally well.

Finally, we apply the tests for UC and CC on the series of prediction intervals. The results of the tests are available in Table 5.4 and allow for the following observations: First, we conclude that the intervals based on Algorithm A always have coverage closer to

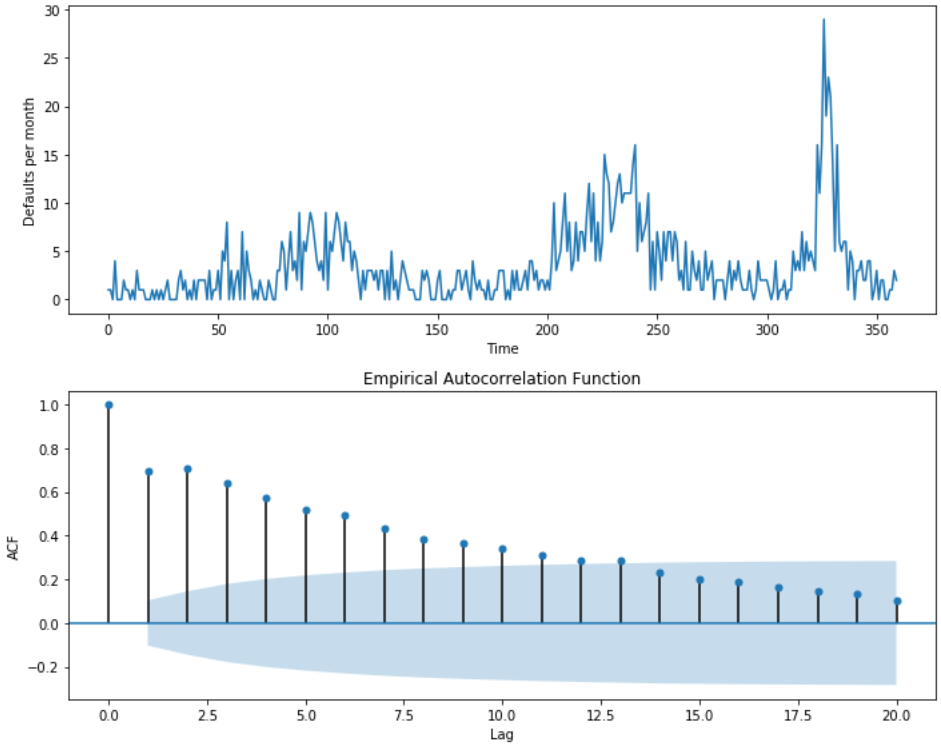

Figure 5.4: Top: Number of defaults per month among Moody's rated US industrial firms from 1982 until 2012. Bottom: Empirical autocorrelation function of the default counts.

Table 5.3: Estimations results for PAR and PARX model specification via MLE.

	ω	α_1	α_1	β	RV	LI^-	$\hat{\alpha}_1 + \hat{\alpha}_2$	AIC	BIC
PAR(2,1)	0.313	0.239	0.224	0.444	-	-	0.464	-1515.33	-1530.66
PARX(2,1)	0.255	0.189	0.204	0.489	27.99	0.772	0.393	-1491.64	-1514.95

Robust standard errors are given in parentheses. Moreover, note that L^- denotes the negative part of LI, that is, $LI^- = \mathbb{I}_{\{LI<0\}}|LI|$. This step is requited to ensure the non-negativity of the chosen intensity specification.

nominal, except for the interval with 99% coverage. Second, Algorithm A passes the tests for UC and CC for all confidence levels except for the 99% level. Third, the approach of ACKR(2016) only passes both tests for the PARX model for prediction intervals with 90% coverage.

When we compare these results with the test results for the transaction counts per minute, we see that our Algorithm A delivers prediction intervals with the correct coverage. However, as a prerequisite for this good performance, we need that the assumption of a Poisson distribution is reasonable for the studied count data. In addition, our results are line with ACKR(2016) and confirm that exogenous covariates seem to improve the forecasting performance of the model at hand. Finally, we note how difficult it is to model extreme quantiles, e.g. the 99% quantile, of the distribution correctly. This observation is in line with Tay and Wallis (2000).

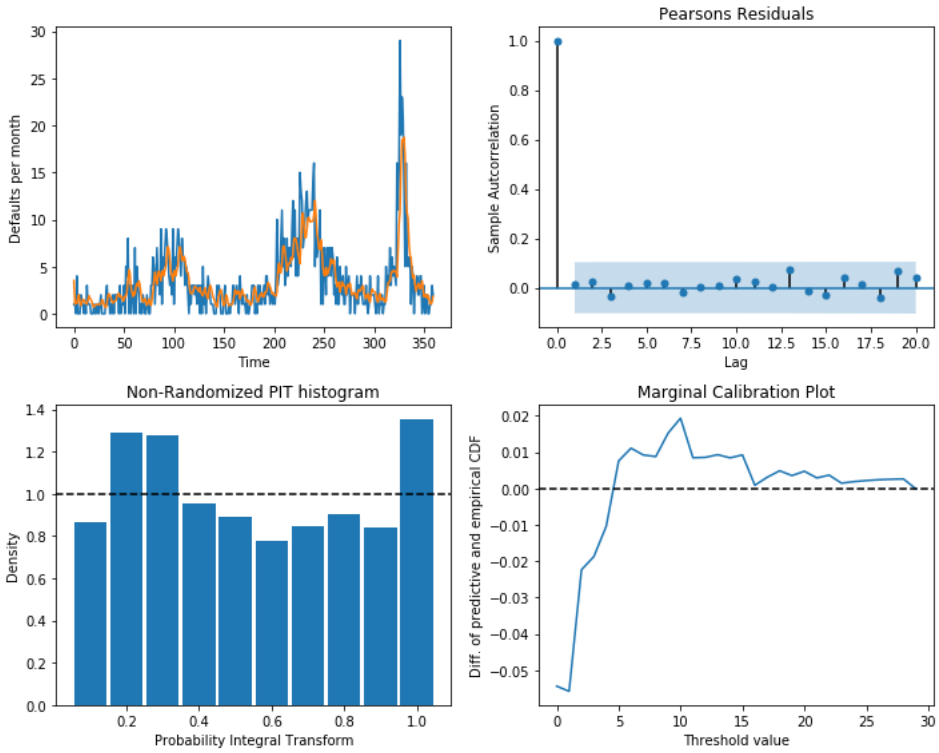

Figure 5.5: PAR(2,1) model. Top left: Fitted values and observed process. Top right: Sample autocorrelation of the Pearson Residuals. Bottom left: Probability Integral Transform. Bottom Right: Marginal Calibration Plot.

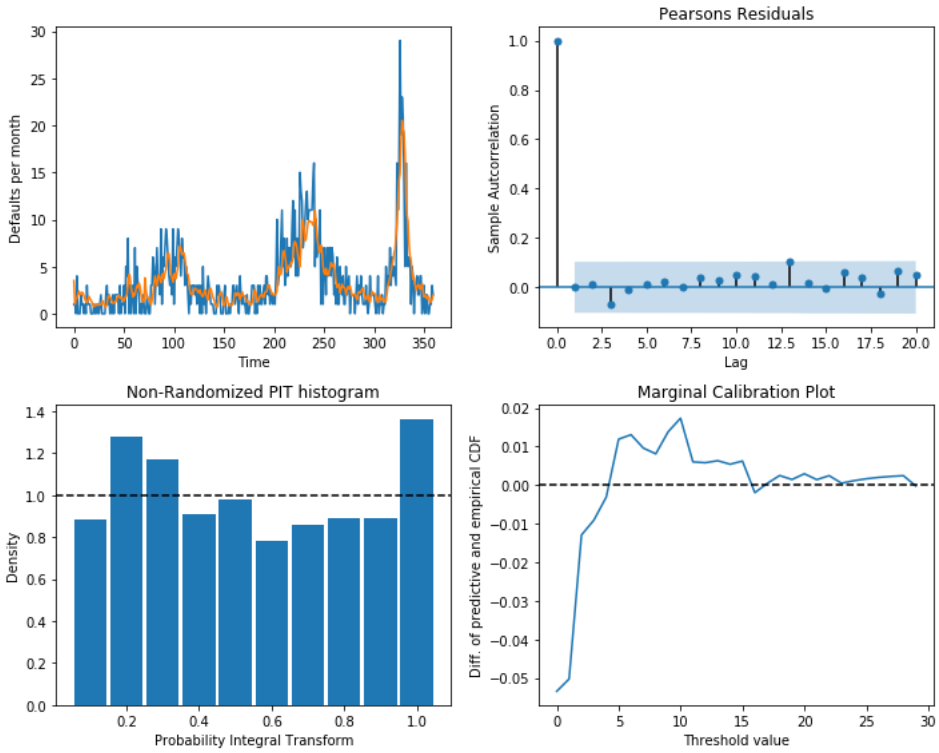

Figure 5.6: PARX(2,1) model. Top left: Fitted values of and observed process. Top right: Sample autocorrelation of the Pearson Residuals. Bottom left: Probability Integral Transform. Bottom Right: Marginal Calibration Plot.

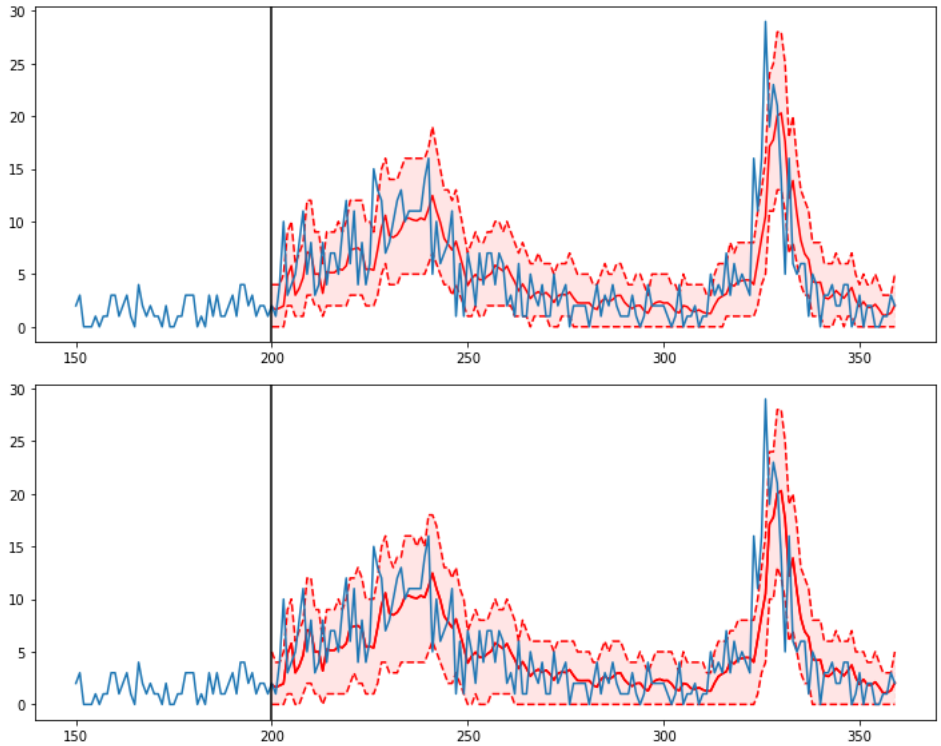

Figure 5.7: PAR(2,1) model. Top: One-step ahead prediction intervals based on ACKR(2016) with nominal coverage of 95%. Bottom: One-step ahead prediction intervals based on Algorithm A with nominal coverage of 95%.

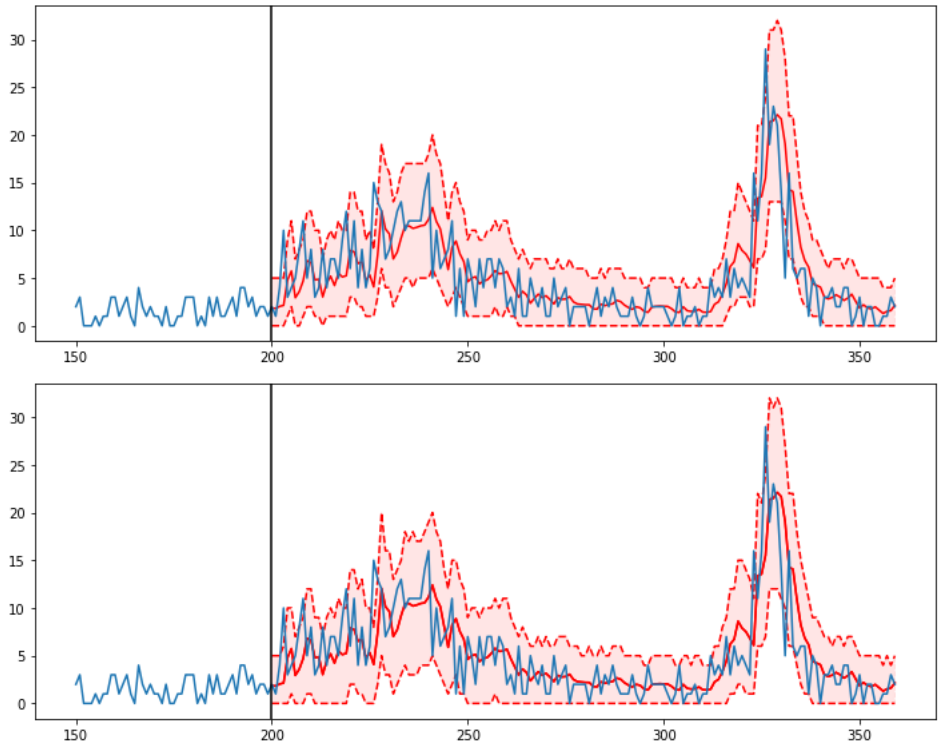

Figure 5.8: PARX(2,1) model. Top: One-step ahead prediction intervals based on ACKR(2016) with nominal coverage of 95%. Bottom: One-step ahead prediction intervals based on Algorithm A with nominal coverage of 95%.

Table 5.4: Coverage of the one-step ahead prediction intervals and p-values for Unconditional Coverage (UC) and Conditional Coverage (CC) Test for the US default count data.

		PAR			PARX		
		Coverage probability, p			Coverage probability, p		
Method		0.1	0.05	0.01	0.1	0.05	0.01
ACKR(2016)	Coverage	0.850	0.900	0.962	0.875	0.906	0.962
	UC	0.047	0.010	0.007	0.308	0.022	0.007
	CC	0.122	0.027	0.020	0.531	0.014	0.020
Algorithm A		0.875	0.931	0.962	0.906	0.937	0.962
		0.308	0.301	0.007	0.790	0.484	0.007
		0.149	0.542	0.020	0.858	0.387	0.020

6 Conclusion

In this article, we introduce a fully parametric bootstrap scheme to simulate the analytically unknown h-step probability mass function of the PAR model. In addition, we study the asymptotic properties of our bootstrap approach. Especially, we show that our bootstrap forecasts are asymptotically valid.

We also look at the finite-sample properties of our bootstrap approach in a controlled Monte Carlo setup. We showed that prediction intervals based on our Algorithm A are more accurate in terms of coverage compared to the procedure proposed in ACKR(2016). When the DGP is governed by a negative binomial distribution, then the performance of both methods decreases. However, Algorithm A still delivers the best results. Finally, in the empirical applications, we illustrate how Algorithm A produces prediction intervals with correct coverage for the US default data set when using a well specified model.

Further issues are left for future research. First, the performance of our Algorithm A crucially depends on the assumption that the conditional distribution is Poisson. Hence, when this assumption is violated, Algorithm A performs poorly and some form of robustness against model mis-misspecification would be desirable. Second, Algorithm A does not take estimation uncertainty into account which is also an interesting extension; cf. Hansen (2006).

Appendix

A Proofs

Proof of Theorem 2.1.

The results follow from using the same arguments as given in Section 3.1 in Ahmad and Francq (2016). □

Proof of Lemma 2.1.

To show the result, we follow Agosto, Cavaliere, Kristensen, and Rahbek (2016) and verify that the conditions of Theorem 3.1 in Doukhan and Wintenberger (2008) hold for the process $\{Y_t\}_{t=1}^T$. First, we write $Y_t = N_t(\lambda_t) = N_t(\omega + \alpha Y_{t-1} + \beta\lambda_{t-1}) =: F(X_{t-1}; N_t)$, where $X_t = (Y_t, \lambda_t)$. Then, we choose an Orlicz function, $\Phi(x)$, such that $\Phi(x) = x, x \in \mathbb{R}_+$, and define the norm $||X||_\Phi$ as

$$||X||_\Phi = \inf\left\{u > 0 \text{ with } \mathrm{E}\left[\Phi\left(\frac{||X||}{u}\right)\right] \leq 1\right\},$$

where $||\cdot||$ denotes the norm of a Banach space. In our case, we note that $||X||_\Phi = \mathrm{E}(|X|)$. Thus, by definition, $||F(0; N_t)||_\Phi = \mathrm{E}|N_t(\omega)| < \infty$ and we have that condition (3.3) of Doukhan and Wintenberger (2008) is satisfied. Next, for any two deterministic sequences $x = (y, \lambda) \in \mathbb{N}_0 \times (0, \infty)$ and $x' = (y', \lambda') \in \mathbb{N}_0 \times (0, \infty)$, we find that

$$\begin{aligned} ||F(x; N_t) - F(x'; N_t)||_\Phi &= \mathrm{E}(|N_t(\omega + \alpha y + \beta\lambda) - N_t(\omega + \alpha y' + \beta\lambda')|) \\ &= |\omega + \alpha y + \beta\lambda - \omega - \alpha y' + \beta\lambda'| \\ &= \alpha|y - y'| + \beta|\lambda - \lambda'|, \end{aligned} \tag{A.1}$$

where the second equality follows from the properties of the Poisson process $N_t(\cdot)$. Consequently, condition (3.1) of Doukhan and Wintenberger (2008) holds if $\alpha + \beta < 1$. Since $\alpha + \beta < 1$ by assumption, Theorem 3.1 in Doukhan and Wintenberger (2008) applies and there exists a weakly dependent stationary and ergodic solution to eq. (2.1). Moreover, all moments of the joint process (Y_t, λ_t) exist and are finite. □

The main tool for showing Theorem 3.2 is the following:

Proposition 1 *(Billingsley 1999, Theorem 3.2) Suppose that $(X_{M,T}, X_T)$ are some random variables, that X_M and X are non-stochastic, and that*

$$\begin{aligned} &(i) \quad |X_{M,T} - X_M| \xrightarrow{p} 0, \text{ as } T \to \infty, \\ &(ii) \quad |X_M - X| \to 0, \text{ as } M \to \infty, \\ &(iii) \quad \lim_{M\to\infty} \limsup_{T\to\infty} P(|X_{M,T} - X_T| \geq \varepsilon) = 0, \text{ for each } \varepsilon > 0. \end{aligned}$$

Then $X_T \to_p X$ *as* $T \to \infty$.

Proof of Theorem 3.1.

It follows that

$$\begin{aligned} & \left|P\left(Y_{T+1}=y|N_T(\lambda)=x\right)-P^*\left(Y^*_{T+1}=y|N_T(\lambda)=x\right)\right| \\ = & \left|\frac{\exp\left(\omega+\alpha x+\beta\lambda\right)\left(\omega+\alpha x+\beta\lambda\right)^y}{y!}\right. \\ & \left.-\frac{\exp\left(\hat{\omega}+\hat{\alpha}x+\hat{\beta}\lambda\right)\left(\hat{\omega}+\hat{\alpha}x+\hat{\beta}\lambda\right)^y}{y!}\right| \to_p 0, \end{aligned} \tag{A.2}$$

as $T \to \infty$, by Theorem 2.1 and Proposition 2.27 in White (2001). □

Proof of Theorem 3.2.

For $h = 1$ it follows that

$$\begin{aligned} & \left|P\left(Y_{T+1}=y|Y_T=x\right)-P^*\left(Y^*_{T+1}=y|Y^*_T=x\right)\right| \\ = & \left|\frac{\exp\left(\lambda_{T+1}(\theta)\right)\lambda^y_{T+1}(\theta)}{y!}-\frac{exp\left(\lambda^*_{T+1}(\hat{\theta})\right)\lambda^{*y}_{T+1}(\hat{\theta})}{y!}\right| \\ = & \left|\frac{\exp\left(\omega+\alpha x\right)\left(\omega+\alpha x\right)^y}{y!}-\frac{exp\left(\hat{\omega}+\hat{\alpha}x\right)\left(\hat{\omega}+\hat{\alpha}x\right)^y}{y!}\right| \to_p 0, \end{aligned} \tag{A.3}$$

as $T \to \infty$, by Theorem 2.1 and Proposition 2.27 in White (2001).

Let $h = 2$. Then, since Y_{T+h} is Markovian,

$$P\left(Y_{T+2}=y|Y_T=x\right)=\sum_{i=0}^{\infty}P\left(Y_{T+2}=y|Y_{T+1}=i\right)P\left(Y_{T+1}=i|Y_T=x\right)=:\sum_{i=0}^{\infty}f_i\leq 1,$$

with f_i non-random. Likewise,

$$P^*\left(Y^*_{T+2}=y|Y^*_T=x\right)=\sum_{i=0}^{\infty}P^*\left(Y^*_{T+2}=y|Y^*_{T+1}=i\right)P^*\left(Y^*_{T+1}=i|Y^*_T=x\right)=:\sum_{i=0}^{\infty}f_{i,T}\leq 1,$$

where the index T signifies the dependence on $\hat{\theta}=(\hat{\omega},\hat{\alpha})'$. We have that

$$\begin{aligned} & P\left(Y_{T+2}=y|Y_T=x\right)-P^*\left(Y^*_{T+2}=y|Y^*_T=x\right) \\ = & \sum_{i=0}^{\infty}\left(f_i-f_{i,T}\right). \end{aligned}$$

For any positive integer M, define

$$S_M := \sum_{i=0}^{M} f_i \quad \text{and } S_{M,T} := \sum_{i=0}^{M} f_{i,T}.$$

Next, we check that Proposition 1 applies. First, we obtain that

$$|S_M - S_{M,T}| \;=\; \left|\sum_{i=0}^{M} (f_i - f_{i,T})\right| \to_p 0, \text{ as } T \to \infty,$$

by the same arguments from eq. (A.3). Second, by definition,

$$|S_M - S| \to 0, \text{ as } M \to \infty,$$

where $S := \sum_{i=0}^{\infty} f_i = P\left(Y_{T+2} = y | Y_T = x\right)$. Finally, it remains to show that

$$\lim_{M\to\infty} \limsup_{T\to\infty} P\left(|S_{M,T} - S_T| \geq \varepsilon\right) = 0,$$

where $S_T := \sum_{i=0}^{\infty} f_{i,T}$. Note that

$$\begin{aligned} |S_{M,T} - S_T| &= \left|\sum_{i=0}^{M} f_{i,T} - \sum_{i=0}^{\infty} f_{i,T}\right| \\ &= \left|\sum_{i=M}^{\infty} f_{i,T}\right|. \end{aligned}$$

Further, $f_{i,T} \to_p f_i$, as $T \to \infty$, by the same arguments from eq. (A.3). In addition, by Theorem 6.6.2 in Resnick (2001), $E\left[|f_{i,T} - f_i|\right] \to 0$, as $T \to \infty$, since $f_{i,T} \leq 1$. Then,

$$\begin{aligned} P\left(\left|\sum_{i=M}^{\infty} f_{i,T} - \sum_{i=M}^{\infty} f_i\right| > \epsilon\right) &\leq \frac{E\left|\sum_{i=M}^{\infty} f_{i,T} - \sum_{i=M}^{\infty} f_i\right|}{\epsilon} \\ &= \frac{E\left|\sum_{i=M}^{\infty} (f_{i,T} - f_i)\right|}{\epsilon} \\ &\leq \frac{E\sum_{i=M}^{\infty} |f_{i,T} - f_i|}{\epsilon} \\ &= \frac{\sum_{i=M}^{\infty} E\,|f_{i,T} - f_i|}{\epsilon} \to 0, \text{ as } T \to \infty, \end{aligned}$$

where the last equality follows because $|f_{i,T} - f_i|$ is positive for all i. Moreover, as $(f_i : i = 0, ...)$ is summable, $\sum_{i=M}^{\infty} f_i \to 0$, as $M \to \infty$, due to Theorem 2.25 in Davidson (1994). Hence $\lim_{M\to\infty} \limsup_{T\to\infty} P\left(|S_{M,T} - S_T| \geq \varepsilon\right) = 0$. We conclude that

$$|S - S_T| \to_p 0, \text{ as } T \to \infty,$$

i.e. that

$$\left|P\left(Y_{T+2}=y|Y_T=x\right)-P^*\left(Y_{T+2}^*=y|Y_T^*=x\right)\right|\rightarrow_p 0,$$

as $T\rightarrow\infty$.

Next, let $h=3$. Then, since Y_{T+h} is Markovian,

$$\begin{aligned}
P\left(Y_{T+3}=y|Y_T=x\right) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P\left(Y_{T+3}=y|Y_{T+2}=i\right)P\left(Y_{T+2}=i|Y_{T+1}=j\right)\\
&\quad\times P\left(Y_{T+1}=j|Y_T=x\right)\\
&= \sum_{i=0}^{\infty}P\left(Y_{T+3}=y|Y_{T+2}=i\right)P\left(Y_{T+2}=i|Y_T=x\right)\\
&:= \sum_{i=0}^{\infty}\tilde{f}_i
\end{aligned}$$

with $\tilde{f}_i$ non-random. Likewise,

$$\begin{aligned}
P^*\left(Y_{T+3}^*=y|Y_T^*=x\right) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P^*\left(Y_{T+3}^*=y|Y_{T+2}^*=i\right)P^*\left(Y_{T+2}^*=i|Y_{T+1}^*=j\right)\\
&\quad\times P^*\left(Y_{T+1}^*=j|Y_T^*=x\right)\\
&= \sum_{i=0}^{\infty}P^*\left(Y_{T+3}^*=y|Y_{T+2}^*=i\right)P^*\left(Y_{T+2}^*=i|Y_{T+x}^*=x\right)\\
&:= \sum_{i=0}^{\infty}\tilde{f}_{i,T},
\end{aligned}$$

where T index signifies the dependence on $\hat{\theta}$. Hence we can use the same arguments made for $h=2$ to show the result for $h=3$.

Moreover, for $h\geq 4$ similar arguments as before hold and the result follows. □

B Additional simulations results

Table B.1: PAR(1) model: Prediction intervals with 90% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

α	T	Method	h			
			1	5	10	20
0.6	200	Algorithm A	0.944 (0.230)	0.941 (0.133)	0.943 (0.098)	0.939 (0.072)
		ACKR(2016)	0.927 (0.260)	0.871 (0.181)	0.860 (0.140)	0.854 (0.100)
	500	Algorithm A	0.959 (0.198)	0.943 (0.126)	0.940 (0.095)	0.939 (0.073)
		ACKR(2016)	0.941 (0.236)	0.878 (0.175)	0.862 (0.143)	0.855 (0.106)
	1000	Algorithm A	0.953 (0.212)	0.943 (0.135)	0.939 (0.101)	0.938 (0.070)
		ACKR(2016)	0.934 (0.248)	0.875 (0.185)	0.856 (0.144)	0.850 (0.102)
0.7	200	Algorithm A	0.946 (0.226)	0.937 (0.137)	0.934 (0.111)	0.931 (0.089)
		ACKR(2016)	0.924 (0.265)	0.849 (0.201)	0.824 (0.166)	0.808 (0.129)
	500	Algorithm A	0.936 (0.245)	0.929 (0.147)	0.933 (0.109)	0.932 (0.083)
		ACKR(2016)	0.916 (0.277)	0.831 (0.215)	0.816 (0.170)	0.806 (0.130)
	1000	Algorithm A	0.940 (0.237)	0.937 (0.141)	0.936 (0.106)	0.934 (0.083)
		ACKR(2016)	0.916 (0.277)	0.842 (0.210)	0.823 (0.167)	0.809 (0.127)
0.8	200	Algorithm A	0.937 (0.243)	0.922 (0.167)	0.916 (0.138)	0.919 (0.107)
		ACKR(2016)	0.919 (0.273)	0.796 (0.246)	0.759 (0.213)	0.733 (0.170)
	500	Algorithm A	0.942 (0.234)	0.925 (0.164)	0.927 (0.129)	0.926 (0.098)
		ACKR(2016)	0.920 (0.271)	0.795 (0.244)	0.761 (0.216)	0.738 (0.167)
	1000	Algorithm A	0.942 (0.234)	0.928 (0.157)	0.923 (0.135)	0.922 (0.108)
		ACKR(2016)	0.921 (0.270)	0.811 (0.232)	0.762 (0.210)	0.735 (0.165)
0.9	200	Algorithm A	0.927 (0.260)	0.911 (0.186)	0.902 (0.171)	0.890 (0.163)
		ACKR(2016)	0.921 (0.270)	0.753 (0.272)	0.672 (0.256)	0.606 (0.223)
	500	Algorithm A	0.924 (0.265)	0.920 (0.175)	0.921 (0.148)	0.913 (0.132)
		ACKR(2016)	0.915 (0.279)	0.760 (0.272)	0.684 (0.249)	0.619 (0.213)
	1000	Algorithm A	0.939 (0.239)	0.924 (0.175)	0.909 (0.165)	0.909 (0.142)
		ACKR(2016)	0.932 (0.252)	0.768 (0.265)	0.680 (0.255)	0.619 (0.223)

Table B.2: PAR(1) model: Prediction intervals with 99% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

α	T	Method	h			
			1	5	10	20
0.6	200	Algorithm A	0.995 (0.071)	0.994 (0.039)	0.994 (0.028)	0.992 (0.023)
		ACKR(2016)	0.994 (0.077)	0.981 (0.075)	0.977 (0.059)	0.974 (0.045)
	500	Algorithm A	0.993 (0.083)	0.992 (0.041)	0.992 (0.029)	0.991 (0.023)
		ACKR(2016)	0.996 (0.063)	0.979 (0.076)	0.975 (0.062)	0.973 (0.047)
	1000	Algorithm A	0.997 (0.055)	0.992 (0.044)	0.991 (0.033)	0.991 (0.024)
		ACKR(2016)	0.997 (0.055)	0.977 (0.081)	0.974 (0.064)	0.972 (0.047)
0.7	200	Algorithm A	0.991 (0.094)	0.992 (0.045)	0.992 (0.035)	0.991 (0.025)
		ACKR(2016)	0.994 (0.077)	0.970 (0.092)	0.959 (0.082)	0.952 (0.070)
	500	Algorithm A	0.994 (0.077)	0.991 (0.052)	0.992 (0.032)	0.992 (0.023)
		ACKR(2016)	0.995 (0.071)	0.964 (0.105)	0.959 (0.083)	0.954 (0.067)
	1000	Algorithm A	0.990 (0.099)	0.992 (0.047)	0.992 (0.034)	0.991 (0.027)
		ACKR(2016)	0.992 (0.089)	0.969 (0.096)	0.960 (0.082)	0.954 (0.065)
0.8	200	Algorithm A	0.996 (0.063)	0.993 (0.041)	0.990 (0.043)	0.989 (0.038)
		ACKR(2016)	0.994 (0.077)	0.944 (0.137)	0.918 (0.134)	0.906 (0.113)
	500	Algorithm A	0.995 (0.071)	0.992 (0.050)	0.992 (0.036)	0.991 (0.032)
		ACKR(2016)	0.995 (0.071)	0.944 (0.135)	0.928 (0.126)	0.912 (0.107)
	1000	Algorithm A	0.987 (0.113)	0.991 (0.046)	0.990 (0.043)	0.989 (0.034)
		ACKR(2016)	0.986 (0.117)	0.947 (0.134)	0.922 (0.135)	0.908 (0.114)
0.9	200	Algorithm A	0.991 (0.094)	0.990 (0.052)	0.989 (0.048)	0.983 (0.057)
		ACKR(2016)	0.994 (0.077)	0.917 (0.169)	0.858 (0.199)	0.802 (0.200)
	500	Algorithm A	0.992 (0.089)	0.992 (0.045)	0.992 (0.042)	0.991 (0.035)
		ACKR(2016)	0.996 (0.063)	0.922 (0.169)	0.875 (0.182)	0.817 (0.184)
	1000	Algorithm A	0.990 (0.099)	0.990 (0.058)	0.986 (0.058)	0.986 (0.051)
		ACKR(2016)	0.989 (0.104)	0.919 (0.171)	0.854 (0.205)	0.807 (0.196)

Table B.3: NBAR(1) model: Prediction intervals with 90% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

α	T	Method	h			
			1	5	10	20
0.6	200	Algorithm A	0.791 (0.407)	0.781 (0.225)	0.785 (0.167)	0.785 (0.120)
		ACKR(2016)	0.791 (0.407)	0.714 (0.245)	0.708 (0.185)	0.702 (0.133)
	500	Algorithm A	0.792 (0.406)	0.787 (0.217)	0.793 (0.163)	0.792 (0.118)
		ACKR(2016)	0.794 (0.404)	0.727 (0.235)	0.719 (0.173)	0.709 (0.128)
	1000	Algorithm A	0.803 (0.398)	0.803 (0.214)	0.804 (0.152)	0.796 (0.114)
		ACKR(2016)	0.801 (0.399)	0.736 (0.230)	0.723 (0.169)	0.708 (0.126)
0.7	200	Algorithm A	0.754 (0.431)	0.753 (0.254)	0.752 (0.191)	0.748 (0.143)
		ACKR(2016)	0.752 (0.432)	0.660 (0.269)	0.631 (0.206)	0.610 (0.156)
	500	Algorithm A	0.757 (0.429)	0.749 (0.256)	0.755 (0.187)	0.756 (0.137)
		ACKR(2016)	0.761 (0.426)	0.663 (0.273)	0.638 (0.201)	0.621 (0.149)
	1000	Algorithm A	0.752 (0.432)	0.757 (0.240)	0.756 (0.186)	0.752 (0.132)
		ACKR(2016)	0.756 (0.429)	0.662 (0.264)	0.638 (0.203)	0.617 (0.145)
0.8	200	Algorithm A	0.710 (0.454)	0.664 (0.307)	0.656 (0.254)	0.654 (0.210)
		ACKR(2016)	0.736 (0.441)	0.570 (0.297)	0.521 (0.237)	0.495 (0.189)
	500	Algorithm A	0.640 (0.480)	0.633 (0.311)	0.639 (0.265)	0.643 (0.228)
		ACKR(2016)	0.694 (0.461)	0.566 (0.299)	0.529 (0.247)	0.511 (0.207)
	1000	Algorithm A	0.624 (0.484)	0.608 (0.347)	0.608 (0.302)	0.610 (0.269)
		ACKR(2016)	0.736 (0.441)	0.603 (0.313)	0.566 (0.269)	0.544 (0.229)

Table B.4: NBAR(1) model: Prediction intervals with 99% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

α	T	Method	h			
			1	5	10	20
0.6	200	Algorithm A	0.932 (0.252)	0.938 (0.137)	0.942 (0.098)	0.943 (0.069)
		ACKR(2016)	0.933 (0.250)	0.895 (0.173)	0.889 (0.131)	0.886 (0.095)
	500	Algorithm A	0.935 (0.247)	0.940 (0.125)	0.946 (0.091)	0.947 (0.068)
		ACKR(2016)	0.933 (0.250)	0.893 (0.168)	0.894 (0.125)	0.893 (0.095)
	1000	Algorithm A	0.961 (0.194)	0.953 (0.117)	0.953 (0.084)	0.949 (0.066)
		ACKR(2016)	0.959 (0.198)	0.914 (0.148)	0.907 (0.115)	0.898 (0.089)
0.7	200	Algorithm A	0.916 (0.277)	0.922 (0.162)	0.922 (0.123)	0.921 (0.093)
		ACKR(2016)	0.913 (0.282)	0.853 (0.217)	0.837 (0.171)	0.823 (0.131)
	500	Algorithm A	0.932 (0.252)	0.930 (0.146)	0.928 (0.115)	0.928 (0.092)
		ACKR(2016)	0.930 (0.255)	0.854 (0.206)	0.838 (0.165)	0.831 (0.125)
	1000	Algorithm A	0.924 (0.265)	0.925 (0.154)	0.924 (0.124)	0.923 (0.090)
		ACKR(2016)	0.919 (0.273)	0.858 (0.196)	0.839 (0.166)	0.825 (0.121)
0.8	200	Algorithm A	0.856 (0.351)	0.844 (0.260)	0.846 (0.225)	0.848 (0.199)
		ACKR(2016)	0.882 (0.323)	0.764 (0.271)	0.723 (0.230)	0.697 (0.185)
	500	Algorithm A	0.823 (0.382)	0.826 (0.282)	0.830 (0.253)	0.831 (0.230)
		ACKR(2016)	0.873 (0.333)	0.763 (0.267)	0.728 (0.230)	0.707 (0.186)
	1000	Algorithm A	0.772 (0.420)	0.782 (0.336)	0.786 (0.312)	0.785 (0.299)
		ACKR(2016)	0.877 (0.328)	0.779 (0.266)	0.745 (0.230)	0.726 (0.193)

Table B.5: PAR(1,1) model: Prediction intervals with 90% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

$\alpha + \beta$	T	Method	h			
			1	5	10	20
$0.2 + 0.6$	200	Algorithm A	0.873 (0.333)	0.900 (0.145)	0.907 (0.102)	0.912 (0.073)
		ACKR(2016)	0.884 (0.320)	0.897 (0.148)	0.901 (0.106)	0.902 (0.077)
	500	Algorithm A	0.917 (0.276)	0.914 (0.128)	0.917 (0.091)	0.915 (0.069)
		ACKR(2016)	0.921 (0.270)	0.908 (0.133)	0.909 (0.094)	0.904 (0.074)
	1000	Algorithm A	0.905 (0.293)	0.915 (0.127)	0.916 (0.094)	0.917 (0.073)
		ACKR(2016)	0.902 (0.297)	0.907 (0.134)	0.903 (0.101)	0.903 (0.078)
$0.2 + 0.7$	200	Algorithm A	0.860 (0.347)	0.884 (0.158)	0.892 (0.117)	0.896 (0.085)
		ACKR(2016)	0.865 (0.342)	0.876 (0.162)	0.879 (0.122)	0.878 (0.092)
	500	Algorithm A	0.883 (0.321)	0.901 (0.152)	0.897 (0.112)	0.903 (0.081)
		ACKR(2016)	0.889 (0.314)	0.896 (0.153)	0.883 (0.118)	0.883 (0.088)
	1000	Algorithm A	0.896 (0.305)	0.892 (0.150)	0.898 (0.112)	0.905 (0.079)
		ACKR(2016)	0.895 (0.307)	0.884 (0.155)	0.881 (0.119)	0.880 (0.089)

Table B.6: PAR(1,1) model: Prediction intervals with 95% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

$\alpha + \beta$	T	Method	h			
			1	5	10	20
$0.2 + 0.6$	200	Algorithm A	0.948 (0.222)	0.954 (0.100)	0.956 (0.071)	0.957 (0.052)
		ACKR(2016)	0.949 (0.220)	0.951 (0.101)	0.949 (0.075)	0.949 (0.056)
	500	Algorithm A	0.959 (0.198)	0.957 (0.097)	0.958 (0.069)	0.957 (0.051)
		ACKR(2016)	0.954 (0.209)	0.950 (0.100)	0.951 (0.073)	0.948 (0.057)
	1000	Algorithm A	0.956 (0.205)	0.960 (0.089)	0.960 (0.064)	0.961 (0.048)
		ACKR(2016)	0.955 (0.207)	0.956 (0.094)	0.956 (0.068)	0.952 (0.053)
$0.2 + 0.7$	200	Algorithm A	0.925 (0.263)	0.942 (0.117)	0.945 (0.084)	0.947 (0.062)
		ACKR(2016)	0.923 (0.267)	0.937 (0.121)	0.935 (0.093)	0.932 (0.071)
	500	Algorithm A	0.939 (0.239)	0.951 (0.101)	0.948 (0.076)	0.952 (0.055)
		ACKR(2016)	0.940 (0.237)	0.944 (0.112)	0.936 (0.087)	0.936 (0.066)
	1000	Algorithm A	0.940 (0.237)	0.943 (0.109)	0.949 (0.078)	0.953 (0.054)
		ACKR(2016)	0.941 (0.236)	0.939 (0.114)	0.938 (0.087)	0.937 (0.064)

Table B.7: PAR(1,1) model: Prediction intervals with 99% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

$\alpha + \beta$	T	Method	h			
			1	5	10	20
$0.2 + 0.6$	200	Algorithm A	0.985 (0.122)	0.989 (0.047)	0.991 (0.030)	0.992 (0.022)
		ACKR(2016)	0.982 (0.133)	0.986 (0.056)	0.988 (0.037)	0.989 (0.026)
	500	Algorithm A	0.989 (0.104)	0.991 (0.042)	0.992 (0.028)	0.991 (0.022)
		ACKR(2016)	0.991 (0.094)	0.989 (0.046)	0.990 (0.032)	0.989 (0.025)
	1000	Algorithm A	0.991 (0.094)	0.992 (0.042)	0.993 (0.028)	0.992 (0.020)
		ACKR(2016)	0.989 (0.104)	0.989 (0.049)	0.990 (0.033)	0.990 (0.023)
$0.2 + 0.7$	200	Algorithm A	0.983 (0.129)	0.988 (0.054)	0.988 (0.037)	0.989 (0.029)
		ACKR(2016)	0.989 (0.104)	0.986 (0.059)	0.984 (0.046)	0.983 (0.036)
	500	Algorithm A	0.981 (0.137)	0.987 (0.051)	0.988 (0.036)	0.990 (0.024)
		ACKR(2016)	0.985 (0.122)	0.986 (0.055)	0.985 (0.043)	0.985 (0.030)
	1000	Algorithm A	0.990 (0.099)	0.988 (0.048)	0.989 (0.035)	0.991 (0.023)
		ACKR(2016)	0.992 (0.089)	0.984 (0.056)	0.985 (0.042)	0.985 (0.030)

Table B.8: NBAR(1,1) model: Prediction intervals with 90% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

$\alpha + \beta$	T	Method	h			
			1	5	10	20
$0.2 + 0.6$	200	Algorithm A	0.685 (0.465)	0.670 (0.222)	0.672 (0.160)	0.677 (0.117)
		ACKR(2016)	0.685 (0.465)	0.663 (0.222)	0.658 (0.160)	0.659 (0.117)
	500	Algorithm A	0.646 (0.478)	0.663 (0.218)	0.675 (0.159)	0.672 (0.115)
		ACKR(2016)	0.653 (0.476)	0.652 (0.223)	0.660 (0.163)	0.654 (0.118)
	1000	Algorithm A	0.690 (0.462)	0.672 (0.219)	0.681 (0.157)	0.682 (0.113)
		ACKR(2016)	0.694 (0.461)	0.663 (0.218)	0.665 (0.159)	0.662 (0.115)

Table B.9: NBAR(1,1) model: Prediction intervals with 95% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

$\alpha + \beta$	T	Method	h			
			1	5	10	20
$0.2 + 0.6$	200	Algorithm A	0.773 (0.419)	0.756 (0.204)	0.758 (0.148)	0.760 (0.109)
		ACKR(2016)	0.776 (0.417)	0.745 (0.206)	0.742 (0.151)	0.743 (0.111)
	500	Algorithm A	0.737 (0.440)	0.750 (0.203)	0.759 (0.149)	0.758 (0.107)
		ACKR(2016)	0.732 (0.443)	0.737 (0.207)	0.743 (0.153)	0.738 (0.110)
	1000	Algorithm A	0.762 (0.426)	0.747 (0.206)	0.759 (0.144)	0.763 (0.105)
		ACKR(2016)	0.766 (0.423)	0.735 (0.209)	0.742 (0.148)	0.743 (0.108)

Table B.10: NBAR(1,1) model: Prediction intervals with 99% coverage based on Algorithm A and ACKR(2016). The results for Algorithm A are based on $B = 1000$ simulations. Standard errors are given in parenthesis and based on $N = 1000$.

$\alpha + \beta$	T	Method	h			
			1	5	10	20
$0.2 + 0.6$	200	Algorithm A	0.891 (0.312)	0.880 (0.153)	0.878 (0.116)	0.878 (0.085)
		ACKR(2016)	0.886 (0.318)	0.864 (0.164)	0.861 (0.122)	0.860 (0.091)
	500	Algorithm A	0.870 (0.336)	0.875 (0.154)	0.878 (0.114)	0.878 (0.082)
		ACKR(2016)	0.871 (0.335)	0.860 (0.163)	0.862 (0.119)	0.859 (0.086)
	1000	Algorithm A	0.874 (0.332)	0.867 (0.167)	0.878 (0.115)	0.881 (0.081)
		ACKR(2016)	0.871 (0.335)	0.855 (0.173)	0.862 (0.121)	0.862 (0.085)

Part III

Estimation Uncertainty in GARCH Option Prices

Estimation Uncertainty in GARCH Option Prices*

Philipp Christian Kless,
Department of Economics,
University of Copenhagen, Denmark.

Abstract

We investigate the impact of estimation uncertainty on GARCH option prices via two novel bootstrap algorithms. First, we design a bootstrap algorithm that allows us to assess the impact of parameter uncertainty on GARCH option prices. For this bootstrap, we assume that the conditional distribution of the innovations is normal under the equivalent martingale measure, $\mathbb{Q}$. Next, we introduce estimation uncertainty by varying the GARCH model parameters in each bootstrap repetition while we keep the innovations fixed. In each bootstrap repetition, we vary the model parameters by drawing them from the asymptotic distribution of the proposed estimator. This design allows us to make sure that the only variation in the bootstrap stems from parameter uncertainty. By means of simulations, we find that the impact of estimation uncertainty is higher for options at the money. In addition, we introduce a second bootstrap which is more general. The second bootstrap allows for conditional distributions under $\mathbb{Q}$ with more skewness and heavier tails than the normal distribution. In an empirical application of the second bootstrap, we show that the uncertainty contained in option prices leads to variation in the metric used to evaluate the pricing performance of competing models. As a result, we find that no single GARCH specification dominates in terms of average relative pricing error when taking parameter uncertainty into account.

Keywords: GARCH; Option pricing; Bootstrap

*I thank Anders Rahbek and Rasmus Søndergaard Pedersen for helpful comments and suggestions. Moreover, I thank Ke Zhu for providing his MATLAB code.

1 Introduction

We consider option pricing under a general GARCH framework. In this setup, option prices are obtained via Monte Carlo simulations since the option prices have no closed form solution in general. Moreover, option prices depend on the unknown GARCH model parameters which are typically recovered from historical time series data using quasi maximum likelihood estimation (QMLE). As a result of relying on estimated parameters, instead of the true unknown model parameters, estimated option prices are subject to estimation uncertainty.

In this paper, we contribute to the literature by making this uncertainty explicit with the help of two bootstrap methods. At first, we design a bootstrap algorithm where we simulate a large number of option prices where the only source of variation stems from variation in the used model parameters. We achieve this by fixing the innovation path in each bootstrap repetition. More precisely, we first construct the risk-neutral Esscher measure under $\mathbb{Q}$ such that the discounted asset price is a martingale. Then, we use Monte Carlo methods to simulate one option price. In a subsequent bootstrap repetition, we simulate another option price via Monte Carlo methods where we use the same innovations from before and only change the parameter estimates. For our bootstrap, we generate new model parameters by drawing them from their asymptotic distribution with an estimated covariance matrix. That is, we assume that parameter uncertainty is reflected by the covariance matrix of the parameter estimates and their asymptotic distribution. This idea has been previously used in Bams, Lehnert, and Wolff (2005) to construct prediction intervals for Value-at-Risk models; see also Blasques, Koopman, Lasak, and Lucas (2015).

We implement our first bootstrap method and quantify the amount of estimation uncertainty in option prices. For this experiment, we assume that the DGP is a GARCH-in-mean (GIM) or AR(1)-GARCH model with normal innovations under $\mathbb{Q}$. The model parameters and their covariance matrix are estimated on real S&P500 data using QMLE. For both models, we find that estimation uncertainty is larger for option prices at the money.

For our second bootstrap method, we generalize our first method. That is, we let the innovations paths differ in each bootstrap repetition. Making this change allows us to extend our bootstrap also to GARCH models where the conditional distribution of the innovations under $\mathbb{Q}$ is more skewed and has heavier tails than the normal distribution. For instance, we later assume that the conditional distribution of the innovations is shift negative gamma. This type of distribution is attractive because it is skewed and has heavier tails than a normal distribution which leads to an improved pricing performance; cf. Siu, Howell, and Yang (2004), and Zhu and Ling (2015).

In an empirical illustration, we implement our second bootstrap to option prices written

on the S&P500. We implement different GARCH models with various specification for the conditional distribution of the innovations. We demonstrate that option prices are subject to parameter uncertainty and show therefore high variation. In addition, we show that this variation also makes the metric used to evaluate pricing performance subject to estimation uncertainty. As a consequence, no single GARCH model dominates all other models when we take parameter uncertainty into account.

In terms of related literature, we note that discrete-time GARCH option pricing models have become popular in recent years. One reason for this increased popularity is that GARCH models are easily estimated compared to continuous-time models. In most of the studies using GARCH models for option pricing, the model parameters are recovered from historical time series data using QMLE; see among others Duan and Zhang (2001), Badescu and Kulperger (2008), and, Zhu and Ling (2015). However, under the GARCH framework, the markets are incomplete and an infinite number of risk neutral measures exists. In the following, we use the Esscher transform to find a risk-neutral measure which allows us to price options. This concept was first used by Siu, Howell, and Yang (2004) in the context of option pricing. One advantage of this measure is that it allows for option pricing under GARCH models with non-normal innovations. In addition, this transform has shown a good performance in empirical applications; cf. Badescu and Kulperger (2008), and Chorro, Guégan, and Ielpo (2015), and Zhu and Ling (2015). In terms of the impact of parameter estimation on option prices, the literature is sparser. For instance, Phillips and Yu (2005) show that estimation bias is transmitted into bond option prices. As a bias correction method, they propose a jackknife estimator which mitigated the bias due to parameter estimation; see also Dotsis and Markellos (2007). Similar studies exist for Value-at-Risk (VaR) models where Bams, Lehnert, and Wolff (2005) show how to adjust VaR calculations to account for parameter uncertainty; see also Pascual, Romo, and Ruiz (2006). Moreover, Blasques, Koopman, Lasak, and Lucas (2015) follow a comparable approach and demonstrate how to adapt in- and out-sample confidence bands for observation driven models to parameter uncertainty. To the best of our knowledge, estimation uncertainty has not been studied for GARCH option prices.

This article is organized as follows: In Section 2, we introduce a risk-neutral Esscher measure under $\mathbb{Q}$. Section 3 provides details about the used GARCH processes under $\mathbb{Q}$. Section 4 shows how to obtain option prices via Monte Carlo methods. Section 5 discusses parameter estimation and its related uncertainty. Section 6 introduces our first bootstrap to quantify estimation uncertainty while Section 7 quantifies numerically estimation uncertainty. Section 8 gives details about our generalized bootstrap while Section 9 contains an empirical application. Section 10 concludes the paper. Appendix A contains additional derivations. Finally, Appendix B provides additional simulation along with additional empirical results.

2 Conditional Esscher Transforms for GARCH models

Similar to Badescu and Kulperger (2008), we consider a discrete-time economy with the time index $\mathcal{T} = \{t|t = 0, \ldots, \tau\}$ of trading dates consisting of one risk-free asset and one risky stock. Let $(\Sigma, \mathbb{P}, \mathcal{F})$ be a complete filtered probability space where $\mathbb{P}$ is the historical probability measure. Denote by $\mathcal{F}_t$ a sequence of increasing σ-fields of $\mathcal{F}$ representing all market information up to time t; we assume that $\mathcal{F}_0 = \{\emptyset, \Sigma\}$ and $\mathcal{F}_\tau = \mathcal{F}$.

Let $(S^0, S) = (S_t^0, S_t)_{t\geq 0}$ be the price process of the risk-free asset and the risky stock; respectively. In this discrete-time setting, we assume that the stock price process S_t is adapted to the filtration $\mathcal{F}$. The risk-free asset price process S_t^0 is deterministic with $S_t^0 = e^{-rt}$, where r is a constant continuously compounded risk-free interest rate. Moreover, we write $y_t = \log \frac{S_t}{S_{t-1}}$ for the continuously compounded return process following a GARCH process under the historical probability measure $\mathbb{P}$. That is,

$$y_t \quad = \quad m_t + z_t\sqrt{h_t}, \tag{2.1}$$

where $z_t|\mathcal{F}_{t-1}$ follows an arbitrary distribution $D(0,1)$ with mean zero and variance one. For the conditional variance, we assume that $h_t \in \mathcal{F}_{t-1}$. Moreover, $m_t \in \mathcal{F}_{t-1}$ is the conditional mean which may depend on h_t and some other unknown parameters such as the constant unit risk premium ν, for instance. The model in eq. (2.1) includes a large class of GARCH models; cf. Section 3.

We impose a no-arbitrage condition by assuming that the stock price admits an equivalent martingale measure (EMM). A probability measure $\mathbb{Q}$ is an EMM wrt $\mathbb{P}$ if

$$E^{\mathbb{Q}}[e^{-rt}S_t|\mathcal{F}_{t-1}] = e^{-r(t-1)}S_{t-1}.$$

Remark 2.1 *In the option pricing literature, it is well known that the market is incomplete in our discrete time setting; see Jacod and Shiryaev (1998). Therefore there may be an infinite number of risk neutral measures to price contingent claims.*

In the following, we use the approach of Gerber and Shiu (1994) to get a risk-neutral measure via the conditional Esscher transform; see also Zhu and Ling (2015), and Chapter 3 in Chorro, Guégan, and Ielpo (2015). For this approach, we assume for the conditional moment generating function of y_t, given in eq. (2.1), wrt $\mathcal{F}_{t-1}$ under $\mathbb{P}$ that

$$M^{\mathbb{P}}_{y_t|\mathcal{F}_{t-1}}(z) = E^{\mathbb{P}}[e^{zy_t}|\mathcal{F}_{t-1}] = \int_{-\infty}^{\infty} e^{zx} dF\left(\frac{x - m_t}{\sqrt{h_t}}\right) < \infty \text{ almost surely}, \tag{2.2}$$

where $z \in \mathbb{R}$, for all $t \geq 0$. Moreover, let $M^{\mathbb{Q}}_{y_t|\mathcal{F}_{t-1}}(z)$ be the conditional moment generating

function of y_t wrt $\mathcal{F}_{t-1}$ under the EMM, $\mathbb{Q}$. Then, Zhu and Ling (2015) show that

$$M^{\mathbb{Q}}_{y_t|\mathcal{F}_{t-1}}(z) = \frac{M^{\mathbb{P}}_{y_t|\mathcal{F}_{t-1}}(z+\delta_t)}{M^{\mathbb{P}}_{y_t|\mathcal{F}_{t-1}}(\delta_t)}, \tag{2.3}$$

where $\delta_t \in \mathcal{F}_{t-1}$. Moreover, Zhu and Ling (2015) show that from eq. (2.3) it follows that

$$E^{\mathbb{Q}}[e^{-rt}S_t|\mathcal{F}_{t-1}] = e^{-r(t-1)}S_{t-1}M^{\mathbb{Q}}_{y_t|\mathcal{F}_{t-1}}(1).$$

Hence, $\{e^{-rt}S_t : t = 0, \ldots\}$ is a martingale under $\mathbb{Q}$ iff $e^{-r}M^{\mathbb{Q}}_{y_t|\mathcal{F}_{t-1}}(1) = 1$ such that

$$\frac{M^{\mathbb{P}}_{y_t|\mathcal{F}_{t-1}}(1+\delta_t)}{M^{\mathbb{P}}_{y_t|\mathcal{F}_{t-1}}(\delta_t)} = e^r \text{ for all } t. \tag{2.4}$$

If there exists a unique solution to eq. (2.4) δ_t, the martingale measure $\mathbb{Q}$, associated with this δ_t, is the EMM; cf. Zhu and Ling (2015).

Now, a fair price of a European call option with strike, K, and maturity, τ, at current time t, denoted V_t, is

$$V_t = E^{\mathbb{Q}}[e^{-r(\tau-t)}\max(S_\tau - K, 0)|\mathcal{F}_{t-1}], \tag{2.5}$$

with $S_\tau = S_t \exp\left(\sum_{i=t+1}^{\tau} y_i\right)$. Note that V_t has no closed form in general. We rely on Monte Carlo simulation to approximate V_t where we simulate the dynamics of $\{y_i\}_{i=t+1}^{\tau}$ under $\mathbb{Q}$; see also Section 4. Hence, we need to use the conditional moment generating function under $\mathbb{Q}$ to obtain S_τ.

The are several economic justifications for the choice of the Esscher transform for derivative pricing. For instance, Gerber and Shiu (1996) show that this transform can be motivated if the representative agent is an expected utility maximizer with power utility functions. Moreover, Badescu and Kulperger (2008) demonstrate that the Esscher transform is justified by a log-linear pricing kernel as a function of y_t. Another advantage is that the Esscher transform can be used as long as the moment generating function of the innovations exists. Finally, Badescu, Elliott, Kulperger, Miettinen, and Siu (2011) demonstrate that the Esscher transform has a good empirical performance when compared to other risk-neutral measures.

3 GARCH processes under $\mathbb{Q}$

In this section, we discuss different specifications of m_t, h_t, and $z_t|\mathcal{F}_{t-1}$ in eq. (2.1) and derive their properties under $\mathbb{Q}$. In general, we study two choices for m_t. First, we consider the case where $m_t = r + \nu\sqrt{h_t}$ which is the usual GARCH-in-mean (GIM) specification; see Badescu and Kulperger (2008) for instance. Second, we look at the case where m_t

follows an AR(p) model; cf. Duan and Zhang (2001), and Zhu and Ling (2015). For both cases we specify the conditional variance, h_t, as a standard linear GARCH model as introduced in Bollerslev (1986). We also consider the NGARCH model of Engle and Ng (1993) which captures the leverage effect in volatility. In addition, the practical usefulness of the NGARCH model has been demonstrated in the literature; cf. Zhu and Ling (2015). Finally, we consider three choices for the conditional distribution of the innovations under $\mathbb{Q}$: z_t is conditionally (i) normal, (ii) shift negative gamma (SNG), and (iii) shift negative inverse Gaussian (SNIG). The SNG and SNIG distribution are popular choices in the literature because they allow for more kurtosis and skewness than the normal distribution and have shown good performance in practice; cf. Siu, Howell, and Yang (2004) and Christoffersen, Heston, and Jacobs (2006).

The following results can be derived using the approach outlined in the previous section; see also Zhu and Ling (2015). For more details see also Appendix A.

3.1 GIM models

First, let $m_t = r + \nu\sqrt{h_{t-1}}$ with h_t being a standard GARCH model where $z_t|\mathcal{F}_{t-1} \sim N(0,1)$. This model choice serves as a benchmark for the rest of the paper. Next we state the dynamics under $\mathbb{Q}$ for the NGARCH model with $z_t|\mathcal{F}_{t-1} \sim N(0,1)$, $SNG(a)$ and $SNIG(d)$, where a and d parameterize the conditional skewness and kurtosis of the innovations; see also Zhu and Ling (2015).

GARCH

Under $\mathbb{P}$, assume for the model in eq. (2.1) that

$$\begin{aligned} y_t &= r + \nu\sqrt{h_t} + \epsilon_t,\ \epsilon_t = z_t\sqrt{h_t}, \\ h_t &= \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}, \end{aligned} \tag{3.1}$$

where $z_t|\mathcal{F}_{t-1} \sim N(0,1)$, r is the risk-free rate, ν is the unit risk premium, and $\omega > 0$, and $\alpha, \beta \geq 0$.

Using the results from Section 2, we find that under $\mathbb{Q}$,

$$\begin{aligned} y_t &= r - h_t/2 + \epsilon_t^*,\ \epsilon_t^*|\mathcal{F}_{t-1} \sim N(0, h_t), \\ h_t &= \omega + \alpha\left(\epsilon_{t-1}^* - \frac{h_{t-1}}{2} - \nu\sqrt{h}_{t-1}\right)^2 + \beta h_{t-1}. \end{aligned} \tag{3.2}$$

These dynamics were also obtained in the seminal paper of Duan (1995).

NGARCH

Under $\mathbb{P}$, we have for the model in eq. (2.1) that

$$\begin{aligned} y_t &= r + \nu\sqrt{h_t} + \epsilon_t,\ \epsilon_t = z_t\sqrt{h_t}, \\ h_t &= \omega + \alpha\left(\epsilon_{t-1} - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}, \end{aligned} \tag{3.3}$$

where $\gamma \geq 0$, and $z_t|\mathcal{F}_{t-1} \sim N(0,1)$, $SNG(a)$, or $SNIG(d)$.

With the approach from Section 2, we get under $\mathbb{Q}$ if $z_t|\mathcal{F}_{t-1} \sim N(0,1)$ that

$$\begin{aligned} y_t &= r + \nu\sqrt{h_t} + \epsilon_t^*,\ \epsilon_t^*|\mathcal{F}_{t-1} \sim N(0, h_t), \\ h_t &= \omega + \alpha\left(\epsilon_{t-1}^* - \frac{h_{t-1}}{2} - (\gamma+\nu)\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}. \end{aligned} \tag{3.4}$$

Using the results from Section 2, we obtain under $\mathbb{Q}$ if $z_t|\mathcal{F}_{t-1} \sim SNG(a)$ that

$$\begin{aligned} y_t &= r + \nu\sqrt{h_t} + \epsilon_t^*, \\ \epsilon_t^* &= \sqrt{ah_t} + \xi_t^*,\ \xi_t^* \sim -G\left(a, b_t\right), \\ h_t &= \omega + \alpha\left(\epsilon_{t-1}^* - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}, \end{aligned} \tag{3.5}$$

where $G(a,b)$ denotes a Gamma distribution with shape parameter a and scale parameter b. Moreover,

$$b_t = \left[\exp\left(\frac{\nu\sqrt{h_t} + \sqrt{ah_t}}{a}\right) - 1\right]^{-1}.$$

Finally, if $z_t|\mathcal{F}_{t-1} \sim SNIG(d)$ we get under $\mathbb{Q}$ that

$$\begin{aligned} y_t &= r + \nu\sqrt{h_t} + \epsilon_t^*, \\ \epsilon_t^* &= \sqrt{dh_t} + c_t^{-1}\sqrt{\frac{h_t}{d}}\xi_t^*,\ \xi_t^* \sim -IG(d\sqrt{c_t}), \\ h_t &= \omega + \alpha\left(\epsilon_{t-1}^* - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}, \end{aligned} \tag{3.6}$$

where $IG\left(d\sqrt{c}\right)$ stands for a inverse Gaussian distribution with mean parameter $d\sqrt{c}$. In addition,

$$c_t = \frac{1}{4}\left(\frac{-\nu\sqrt{h_t} - \sqrt{dh_t}}{2d} - \frac{2\sqrt{dh_t}}{-\nu\sqrt{h_t} - \sqrt{dh_t}}\right)^2.$$

3.2 AR(p) models

As an alternative, we consider a class of models where we assume that $m_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i}$ for $p > 1$. We choose h_t as a standard linear GARCH model with $z_t|\mathcal{F}_{t-1} \sim N(0,1)$. Moreover, we specify h_t as NGARCH model and take $z_t|\mathcal{F}_{t-1} \sim N(0,1)$, $SNG(a)$, or $SNIG(d)$; see also Zhu and Ling (2015).

GARCH

Under $\mathbb{P}$, assume for the model in eq. (2.1) that

$$\begin{aligned} y_t &= \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t,\ \epsilon_t = z_t\sqrt{h_t}, \\ h_t &= \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}, \end{aligned} \tag{3.7}$$

where $z_t|\mathcal{F}_{t-1} \sim N(0,1)$.

Using the results from Section 2, we find that under $\mathbb{Q}$,

$$\begin{aligned} y_t &= r - h_t/2 + \epsilon_t^*,\ \epsilon_t^*|\mathcal{F}_{t-1} \sim N(0,h_t), \\ h_t &= \omega + \alpha z_{t-1}^{*2} + \beta h_{t-1}, \end{aligned} \tag{3.8}$$

where $z_t^* = \epsilon_t^* + r - \frac{h_t}{2} - \phi_0 - \sum_{i=1}^{p} \phi_i y_{t-i}$.

NGARCH

Under $\mathbb{P}$, assume for the model in eq. (2.1) that

$$\begin{aligned} y_t &= \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t,\ \epsilon_t = z_t\sqrt{h_t}, \\ h_t &= \omega + \alpha\left(\epsilon_{t-1} - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}, \end{aligned} \tag{3.9}$$

where $z_t|\mathcal{F}_{t-1} \sim N(0,1)$, $SNG(a)$, or $SNIG(d)$.

If $z_t|\mathcal{F}_{t-1} \sim N(0,1)$, we obtain under $\mathbb{Q}$ that

$$\begin{aligned} y_t &= r - \frac{h_t}{2} + \epsilon_t^*,\ \epsilon_t^*|\mathcal{F}_{t-1} \sim N(0,h_t), \\ h_t &= \omega + \alpha\left(z_{t-1}^* - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}. \end{aligned} \tag{3.10}$$

Next, if $z_t|\mathcal{F}_{t-1} \sim SNG(a)$, the dynamics under $\mathbb{Q}$ are

$$\begin{aligned} y_t &= \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t^*, \\ \epsilon_t^* &= \sqrt{ah_t} + \xi_t^*,\ \xi_t^* \sim -G\left(a, \bar{b}_t\right), \\ h_t &= \omega + \alpha\left(\epsilon_{t-1}^* - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}, \end{aligned} \tag{3.11}$$

where

$$\bar{b}_t = \left[\exp\left(\frac{\bar{\mu}_t^* - r + \sqrt{ah_t}}{a}\right) - 1\right]^{-1}$$

with $\bar{\mu}_t^* = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i}$.

Finally, if $z_t|\mathcal{F}_{t-1} \sim SNIG(a)$, then under $\mathbb{Q}$ it follows that

$$\begin{aligned} y_t &= \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t^*, \\ \epsilon_t^* &= \sqrt{dh_t} + \frac{1}{\bar{c}_t}\sqrt{\frac{h_t}{d}}\xi_t^*,\ \xi_t^* \sim -IG\left(d\sqrt{\bar{c}_t}\right), \\ h_t &= \omega + \alpha\left(\epsilon_{t-1}^* - \gamma\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}, \end{aligned} \tag{3.12}$$

where

$$\bar{c}_t = \frac{1}{4}\left(\frac{r - \bar{\mu}_t^* - \sqrt{dh_t}}{2d} - \frac{2\sqrt{dh_t}}{r - \bar{\mu}_t^* - \sqrt{dh_t}}\right)^2.$$

Given the dynamics under $\mathbb{Q}$, we next turn how to simulate the return series under $\mathbb{Q}$ and obtain a simulation based option prices.

4 Simulation based GARCH option prices

In this section, we discuss a simulation based procedure to price V_t. More precisely, the option price, V_t, in eq. (2.5) has no closed form in general. Hence, we rely on Monte Carlo simulations to approximate V_t at current time t given some parameter estimates. We follow Zhu and Ling (2015) and use the following algorithm:

Pricing Algorithm:

1. Obtain $S_\tau = S_t \exp\left(\sum_{i=t+1}^{\tau} y_i\right)$, where $\{y_i\}_{i=t+1}^{\tau}$ is simulated under $\mathbb{Q}$ according to eq. (3.2), (3.4), (3.5), (3.6), (3.8), (3.10), (3.11), and (3.12); respectively.

2. Compute

$$v_t = e^{-r(\tau-t)} \max\{S_\tau - K, 0\}.$$

3. Repeat Step 1 and 2 N times to get a sequence $\left\{v_t^{(n)}\right\}_{n=1}^{N}$ and obtain V_t by $\tilde{V}_t = \frac{1}{N}\sum_{n=1}^{N} v_t^{(n)}$.

Note that in our simulations we rely on empirical martingale simulation (EMS) as a variance reduction tool; see Duan and Simonato (1998). The EMS approach is based on the observation that option prices simulated using Monte Carlo methods to approximate expectations under one equivalent martingale measure (EMM) often violate arbitrage conditions. To overcome the preceding problem and to increase simulation accuracy, the EMS method adjusts Monte Carlo simulations such that the call-put parity is enforced. That is, they replace the n-th path of $S_\tau^{(n)}$ by

$$\hat{S}_\tau^{(n)} = \frac{S_\tau^{(n)}}{\frac{1}{N}\sum_{n=1}^{N} S_\tau^{(n)}} S_t e^{r(\tau-t)}. \tag{4.1}$$

Hence, the modified price estimates changes to

$$\hat{V}_t = \frac{e^{r(\tau-t)}}{N} \sum_{i=1}^{N} \max\left\{\hat{S}_\tau^{(n)} - K, 0\right\}. \tag{4.2}$$

Asymptotic properties of $\hat{V}_t$ have been studied in Duan and Simonato (1998). The EMS approach is widely applied to improve the numerical efficiency of Monte Carlo simulations in GARCH option pricing models; cf. Barone-Adesi, Engle, and Mancini (2008) and Chorro, Guégan, and Ielpo (2012).

5 Parameter estimation and estimation uncertainty

In the previous section, we discussed how to obtain simulation based option prices for the different GARCH specifications. However, to implement this Monte Carlo approach, we need to recover the unknown parameters from historical time series data. We do this by using a two-step approach. More precisely, in a first step, we estimate the GARCH model parameters using QMLE where we assume that the innovations are Gaussian. And in a second step, we estimate the distributional parameters a, and d by the method of moments approach. This two step approach is common in the literature and performs well; see Chapter 4 in Chorro, Guégan, and Ielpo (2015) for a large empirical application.

To estimate the model parameters, we assume as before that under $\mathbb{P}$,

$$\begin{aligned} y_t &= m_t(\theta) + \epsilon_t(\theta), \\ \epsilon_t(\theta) &= z_t\sqrt{h_t(\theta)},\ t = 1, \ldots, T, \end{aligned} \tag{5.1}$$

where $z_t \sim N(0,1)$, $T < \tau$, and θ contains the GARCH model parameters to be estimated. Let θ_0 denote the true, data-generating value.

Then, the conditional Gaussian log-likelihood function of θ in terms of the observations, given some initial values ϵ_0 and h_0, takes following form

$$L_T(\theta) = -\sum_{t=1}^{T} l_t(\theta),\ l_t(\theta) = \frac{\epsilon_t^2(\theta)}{h_t(\theta)} + \log h_t(\theta), \tag{5.2}$$

where we omitted any constant terms. The quasi maximum likelihood estimator is then computed as

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L_T(\theta), \tag{5.3}$$

where Θ denotes the parameter space. Under certain assumptions, it can be shown that $\hat{\theta}$ is asymptotically normal distributed, that is,

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \to_d N\left(0, \Sigma\right), \tag{5.4}$$

where Σ is some positive definite matrix, as $T \to \infty$. For instance, Francq and Zakoïan (2004) provide conditions for estimators of ARMA-GARCH to be asymptotically normal while Conrad and Mammen (2016) state conditions for the asymptotic normality of estimators in GIM models. We conjecture that the same result holds for the estimators of the NGARCH model.

Remark 5.1 *Observe that Σ is estimated by*

$$\hat{\Sigma} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}, \tag{5.5}$$

where $\hat{I} = T^{-1}\sum_{t=1}^{T}(\partial l_t(\hat{\theta})/\partial\theta)(\partial l_t(\hat{\theta})/\partial\theta')$ and $\hat{J} = T^{-1}\sum_{t=1}^{T}\partial^2 l_t(\hat{\theta})/\partial\theta\partial\theta'$.

In the second step, we estimate the distributional parameters a and d by methods of moments. That is, we exploit that $E\left(z_t^3|\mathcal{F}_{t-1}\right) = \frac{-2}{\sqrt{a}}$ if $z_t|\mathcal{F}_{t-1} \sim SNG(a)$ and $E\left(z_t^3|\mathcal{F}_{t-1}\right) = \frac{-3}{\sqrt{d}}$ if $z_t|\mathcal{F}_{t-1} \sim SNIG(d)$ such that

$$\hat{a} = \left(\frac{2\sum h_t^{3/2}(\hat{\theta})}{\sum \epsilon_t^3(\hat{\theta})}\right)^2 \text{ and } \hat{d} = \left(\frac{3\sum h_t^{3/2}(\hat{\theta})}{\sum \epsilon_t^3(\hat{\theta})}\right)^2, \tag{5.6}$$

see Zhu and Ling (2015), and Siu, Howell, and Yang (2004),

Naturally, the estimation of the model parameters introduces estimation uncertainty in the simulated GARCH option prices. In this paper, we follow Blasques, Koopman, Lasak, and Lucas (2015), and Bams, Lehnert, and Wolff (2005) to approximate parameter uncertainty by the covariance matrix of the parameter estimates and its asymptotic distribution. That is, we use eq. (5.4) to generate a new set of GARCH model parameters, θ^* as follows:

$$\theta^* \sim N\left(\hat{\theta}, \frac{1}{T}\hat{\Sigma}\right), \tag{5.7}$$

where $\hat{\theta}$ is estimated by QMLE. In addition, we obtain a new set of distributional parameters, a^*, and b^* through

$$a^* = \left(\frac{2\sum h_t^{3/2}(\theta^*)}{\sum \epsilon_t^3(\theta^*)}\right)^2 \text{ and } d^* = \left(\frac{3\sum h_t^{3/2}(\theta^*)}{\sum \epsilon_t^3(\theta^*)}\right)^2. \tag{5.8}$$

Remark 5.2 *As an alternative to using eq. (5.8), one could consider to estimate the standard errors of $\hat{a}$ and $\hat{b}$ such that we generate a^* by drawing from $N\left(\hat{a}, \frac{1}{T}\hat{\sigma}_a\right)$; respectively d^* by drawing from $N\left(\hat{d}, \frac{1}{T}\hat{\sigma}_d\right)$.*

6 Bootstrap algorithm to quantify estimation uncertainty

In this section, we follow Bams, Lehnert, and Wolff (2005), and Blasques, Koopman, Lasak, and Lucas (2015) to model parameter uncertainty. That is, we outline a bootstrap algorithm that allows us to assess the impact of parameter uncertainty on $\tilde{V}_t$. More precisely, we simulate a large number of $\tilde{V}_t$ via the bootstrap by repeating Pricing Algorithm 1 many times. For this bootstrap, we fix the innovations under $\mathbb{Q}$ and only allow for variation in the model parameters. That is, for each bootstrap repetition, we generate a new set of parameters by $\theta^* \sim N(\hat{\theta}, \frac{1}{T}\hat{\Sigma})$. we can only fix the innovations under $\mathbb{Q}$ if we assume that the innovations are Gaussian. In the other two cases, the distributional parameters b_t, and c_t; respectively $\bar{b}_t$, and $\bar{c}_t$ vary over time. Hence, we cannot eliminate innovation uncertainty by fixing the innovations in each bootstrap repetition. One way to minimize innovation uncertainty in the non-Gaussian cases is to use a large number of Monte Carlo repetitions. However, for the sole purpose of illustrating parameter uncertainty, we argue that it is better to use a Gaussian distribution such that we can exclude innovation uncertainty completely.

Bootstrap Algorithm 1:

1. Obtain $\hat{\theta}$ as eq. (5.3) and $\hat{\Sigma}$ as in eq. (5.5) using the historical data set $\{y_t\}_{t=1}^T$.

2. Draw $\left\{z_t^{(n)}\right\}_{t=T+1}^{\tau}$ i.i.d. from $N(0,1)$ for $n = 1, \ldots, N$.

3. Draw $\theta^* \sim N(\hat{\theta}, \frac{1}{T}\hat{\Sigma})$. If the parameter constraints are violated, then draw a new set of θ^*.

4. Obtain $S_\tau^{(n)} = S_T \exp\left(\sum_{t=T+1}^{\tau} y_t^{(n)}\right)$, where $\left\{y_t^{(n)}\right\}_{t=T+1}^{\tau}$ is simulated under $\mathbb{Q}$ according to eq. (3.2), (3.4), or (3.8), using $\left\{z_t^{(n)}\right\}_{t=T+1}^{\tau}$, and θ^*.

5. Compute

$$v_T = e^{-r(\tau-T)} \max\{S_\tau - K, 0\},$$

   where S_τ from Step 4 is inserted.

6. Repeat Step 4 and 5 N times to get a sequence $\left\{v_T^{(n)}\right\}_{n=1}^{N}$ and calculate $\tilde{V}_T = \frac{1}{N}\sum_{n=1}^{N} v_T^{(n)}$.

7. Repeat Step 3 to 6 B times to get a sequence $\left\{\tilde{V}_T^{*(b)}\right\}_{b=1}^{B}$.

Remark 6.1 *Note that we fix the innovations $\{z_t^{(n)}\}_{t=T+1}^{\tau}$ from Step 2 in each of the B repetitions. This design ensures that the only variation in $\{\tilde{V}_T^{*(b)}\}_{b=1}^{B}$ is due to variation in θ^*.*

Remark 6.2 *Note that in Step 3 we ensure that the parameter constraints are met when drawing new model parameters θ^*. For instance, we ensure that $\omega^* > 0$ such that conditional variance stays strictly positive.*

7 Monte Carlo experiment

Next, we conduct a Monte Carlo experiment to quantify estimation uncertainty in GARCH option prices. That is, we simulate a large number of $\tilde{V}_T^*$ via Bootstrap Algorithm 1 for call options written on the S&P500 index. We do this for the GIM-GARCH and AR(1)-GARCH model with Gaussian innovations; cf. eq. (3.2), and (3.8).

First, to determine the parameter values for our numerical experiment, we use QMLE to fit the GIM-GARCH and AR(1)-GARCH model to historical daily log-return series of the S&P500 index. Our data contains $T = 1437$ observations ranging from January 4, 2010 to September 7, 2015.

Second, we obtain a large number of GARCH option prices via Bootstrap Algorithm 1 where we set $r = 3\%$, $S_T = 1990.20$, TM $:= (\tau - T) \in \{21, 45, 64, 82, 125\}$, $N = 50,000$, and $B = 999$. Finally, the strike is $K \in \{1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250\}$. And hence the moneyness, denoted M, ranges from 0.85 to 1.13.

Table 7.1: GIM-GARCH(1,1) parameter estimators and their robust standard errors for S&P500 obtained via QMLE.

ν	ω	α	β	$\alpha+\beta$
0.102	4.18×10^{-6}	0.1551	0.807	0.959
(0.026)	(0.000)	(0.025)	(0.026)	

Once the we simulated a series of $\tilde{V}_T^*$, $\left\{V_T^{*(b)}\right\}_{b=1}^B$, we calculate the median, the 2.5%- and 97.5%-quantile along with $\mu = \frac{1}{B}\sum_{b=1}^B \tilde{V}_T^{*(b)}$, and $\sigma = \sqrt{\frac{1}{B-1}\sum_{b=1}^B \left(\tilde{V}_T^{*(b)} - \mu\right)^2}$ as summary statistics. In addition, we also state the length of this prediction interval as $\Delta = 97.5\%\text{-quantile} - 2.5\%\text{-quantile}$. The results are discussed next.

7.1 GIM-GARCH

The estimation results are reported in Table 7.1. The persistence of the model implied by these parameters is $\alpha + \beta = 0.959$ which is in line with what is usually found in the empirical literature.

In Figure 7.1, we plot σ of $\left\{V_T^{*(b)}\right\}_{b=1}^B$ against moneyness and time to maturity. We make the following observations when looking at Figure 7.1. First, for a given time to maturity, we observe that the variation in option prices, measured by σ, is bigger for options at the money or close to being at the money.

This observation is in line with the fact that options at the money have the largest sensitivity wrt the variance of the price of the underlying stock (vega). More precisely, the variance is modeled as a GARCH process which is subject to estimation uncertainty. Hence, the impact of estimation uncertainty is largest for option prices at the money. Second, for a given moneyness, the variation in the option prices increases as the time to maturity increases. This finding is also intuitive since with an increasing time horizon, the model parameters could be subject to structural change which increases parameter uncertainty. This in turn leads to a larger variation in option prices. For instance, standard deviation of option prices with 21 days to maturity is less than the level of options with 45 days to maturity. Further summary statistics for $\left\{\tilde{V}_T^{*(b)}\right\}_{b=1}^B$ are tabulated in Table B.3 in Appendix B.

7.2 AR(1)-GARCH

Compared to the GIM-GARCH model, we introduce one additional parameter in the conditional mean part of the return process. The AR structure in the conditional mean is chosen to model the autocorrelation in the return process. We use the same data and sample size as in the GIM-GARCH case to estimate the AR(1)-GARCH model. The

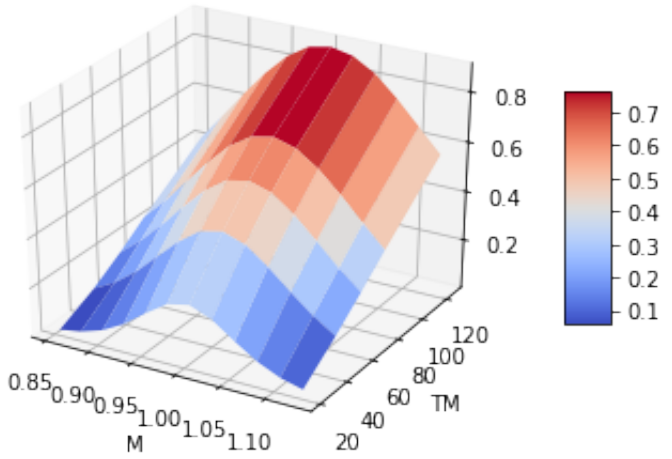

Figure 7.1: Standard deviation plotted against moneyness (M) and time to maturity (TM) for the GIM-GARCH with normal innovations.

Table 7.2: AR(1)-GARCH(1,1) parameter estimators and their robust standard errors for S&P500 obtained via QMLE

ϕ_0	ϕ_1	ω	α	β	$\alpha+\beta$
2.762×10^{-4}	-0.026	4.069×10^{-6}	0.140	0.817	0.958
(0.000)	(0.028)	(0.000)	(0.025)	(0.026)	

results are presented in Table 7.2. Compared to the previous case, the parameter estimates of the volatility parameters remain almost unchanged. Also, the persistence of the model is almost the same as before; $\alpha + \beta = 0.958$.

In general, we draw the same conclusions from Figure 7.2 as in the GIM-GARCH case. That is, for a given moneyness, the variation of the option prices is highest for at the money options. Moreover, option prices vary more as the time to maturity increases for a given moneyness. In general it holds that option prices with a longer maturity have a higher variation. However, it is surprising that the AR(1)-GARCH model has not in general a higher level of variation than the simpler GIM-GARCH model. A plausible observation would be that the AR(1)-GARCH model has a higher level of variation since this model has one additional parameters leading to more estimation uncertainty which should also be reflected in the variation of the simulated option prices. Further summary statistics for $\left\{\tilde{V}_T^{*(b)}\right\}_{b=1}^{B}$ are tabulated in Table B.4 in the Appendix B.

8 Bootstrap algorithm for GARCH option prices

In the previous two sections, we demonstrated that option prices are subject to estimation uncertainty which varies with moneyness and time to maturity. To acknowledge this uncertainty in GARCH option prices, we propose in the following another bootstrap

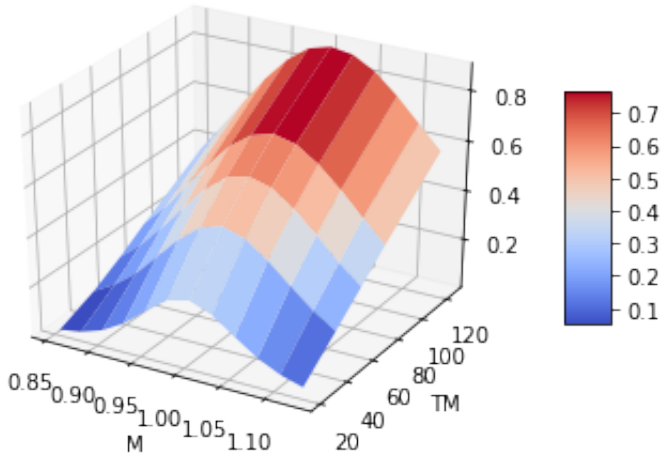

Figure 7.2: Standard deviation plotted against moneyness (M) and time to maturity (TM) for the AR(1)-GARCH(1,1) model

algorithm. For this second bootstrap, we additionally introduce variation in the innovation paths in each bootstrap repetition. This modification allows us to simulate a large number of option prices for various GARCH model specifications where the conditional distribution under $\mathbb{Q}$ is non-normal; cf. Section 3. The variation in model parameters is again introduced by using $\theta^* \sim N\left(\hat{\theta}, \frac{1}{T}\hat{\Sigma}\right)$.

Bootstrap Algorithm 2:

1. Obtain $\hat{\theta}$ as eq. (5.3) and $\hat{\Sigma}$ as in eq. (5.5) using the historical data set $\{y_t\}_{t=1}^{T}$.

2. Draw $\theta^* \sim N(\hat{\theta}, \frac{1}{T}\hat{\Sigma})$ and calculate a^* when $z_t|\mathcal{F}_{t-1} \sim SNG(a)$ as in eq. (5.8); respectively d^* when $z_t|\mathcal{F}_{t-1} \sim SNIG(d)$. If the parameter constraints are violated, then draw a new set of θ^*.

3. Obtain $S_\tau = S_T \exp\left(\sum_{t=T+1}^{\tau} y_t\right)$, where $\{y_t\}_{t=T+1}^{\tau}$ is simulated under $\mathbb{Q}$ according to eq. (3.2), (3.4), (3.5), (3.6), (3.8), (3.10), (3.11), or (3.12), and using θ^* in combination with a^* or d^* if needed.

4. Compute

$$v_T = e^{-r(\tau-T)} \max\{S_\tau - K, 0\},$$

   where S_τ from Step 3 is inserted.

5. Repeat Step 3 and 4 N times to get a sequence $\left\{v_T^{(n)}\right\}_{n=1}^{N}$ and calculate $\tilde{V}_T = \frac{1}{N}\sum_{n=1}^{N} v_T^{(n)}$.

6. Repeat Step 2 to 5 B times to get a sequence $\left\{\tilde{V}_T^{*(b)}\right\}_{b=1}^{B}$.

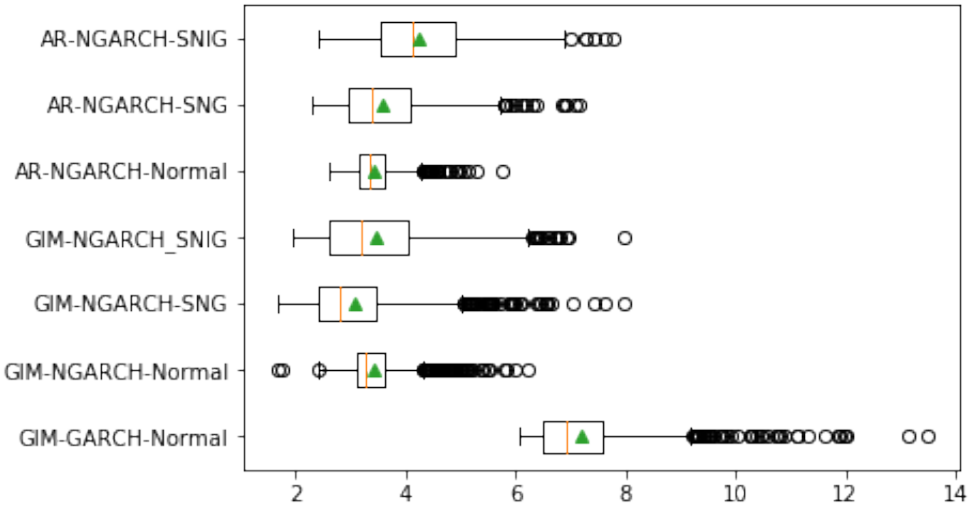

Figure 9.1: Box plot of total ARE_B for all considered specifications. The green triangle depicts the sample mean of the corresponding ARE_B while the red line is the median. The results are based on $N = 50,000$ and $B = 999$.

In the next section, we will see how our Bootstrap Algorithm 2 performs when applied to real data.

9 Empirical illustration

In this section, we implement Bootstrap Algorithm 2 on a real data set to illustrate the variation in GARCH option prices. The empirical study is performed using European Call option data available in Schoutens (2003). We use this data set because it makes our findings comparable to results in Badescu and Kulperger (2008) and Zhu and Ling (2015). The data set contains 54 European Call options written on S&P500 stock index at the close of the market on April 18, 2002. The closing prices on this day was $S_0 = 1124.47$, and the annual risk-free rate r is 1.9%. The strike prices range from \$975 to \$1325 while the time to maturity is $TM := \tau - T \in \{22, 46, 109, 173, 234\}$ days. The parameters of all pricing models are estimated using the log-return of daily closing prices of the index from January 04, 1988 to April, 17, 2002. In total we have $T = 3606$ observations.

For our analysis, we follow Zhu and Ling (2015) and estimate three different models: First, we obtain estimates for the simple GIM-GARCH model which will serve as a benchmark. Second, we also estimate a GIM-NGARCH model. Finally, we obtain parameter estimates for a AR(3)-NGARCH model. The estimates are available in Table B.1 in Appendix B. In terms of misspecification, we calculate the Ljung and Box test for the residuals; respectively squared residuals; cf. Table B.2 in Appendix B. From this table, we conclude that the AR-type model provide a good fit for the log-return series. This is not true for the GIM-type specifications which cannot fit the conditional mean of the log-return series properly.

Table 9.1: Summary statistics of ARE for SNG innovations

Model	κ	TM	$\{ARE^{(b)}\}_{b=1}^{B}$			
			μ	2.5%	97.5%	Δ
GIM-NGARCH	0.7	22	2.36	1.14	4.24	3.10
AR-NGARCH	0.7		1.20	0.71	2.35	1.64
		46	2.58	0.54	5.80	5.26
			2.03	0.48	4.60	4.12
		109	2.00	0.75	4.58	3.83
			2.70	1.62	4.42	2.81
		173	3.23	2.22	5.73	3.51
			4.45	3.31	6.44	3.12
		234	5.03	3.68	8.79	5.11
			6.48	4.75	9.33	4.58
		All	3.06	2.01	5.56	3.55
			3.58	2.52	5.42	2.91

All results are based on $N = 50,000$ and $B = 999$. Note that $\mu = \frac{1}{B}\sum_{b=1}^{B} ARE^{(b)}$ and $\Delta = 97.5\%$-quantile $-$ 2.5%-quantile

Based on our estimation results, we next implement Bootstrap Algorithm 2 to simulate B option prices, $\left\{\tilde{V}_T^{*(b)}\right\}_{b=1}^{B}$. We set $N = 50,000$ and $B = 999$ for the simulations. To get an understanding of the variation in the simulated option prices, we obtain the 95% confidence intervals for each price and tabulate it together with the true option price. For instance, we show the confidence intervals for the GIM-NGARCH model with SNG innovations in Table B.7 in Appendix B. From this table, we see that the length of a given prediction interval increases as the time to maturity increases. Hence, this implies that our simulated option prices vary more as the time horizon increases. This finding is qualitatively the same for all models as seen from the remaining tables in Appendix B.

Uncertainty in the simulated option prices also implies uncertainty in the metrics used to evaluate the performance of the different model implied prices. In our setting, we choose the average relative error (ARE) to highlight this uncertainty. Specifically, for all simulated option price, $\tilde{V}_T^{*(b)}$, we calculate the ARE as

$$ARE = \frac{1}{K}\sum_{j=1}^{K} \frac{\left|V_j^{\text{model}} - V_j^{\text{market}}\right|}{V_j^{\text{market}}} \times 100, \tag{9.1}$$

where K is the total number of options, V_j^{market} is the market price of the jth option while V_j^{model} is the jth option price implied by Bootstrap Algorithm 2. Note that V_j^{model} depends on the initial variance, h_1, which we set to $(\kappa\sigma_\epsilon)$, where σ_ϵ is the estimated volatility of

Table 9.2: Summary statistics of ARE for SNIG innovations

Model	κ	TM	$\{ARE^{(b)}\}_{b=1}^{B}$			
			μ	2.5	97.5	Δ
GIM-NGARCH	0.7	22	2.55	1.38	4.03	2.65
AR-NGARCH	0.7		2.67	1.63	3.88	2.24
		46	3.22	0.59	5.92	5.33
			3.49	1.29	5.59	4.30
		109	2.47	1.01	4.98	3.97
			3.35	1.68	5.21	3.53
		173	3.49	2.63	5.95	3.32
			4.47	3.38	6.53	3.14
		234	5.41	4.17	9.02	4.86
			6.69	4.91	9.78	4.86
		All	3.47	2.21	6.01	3.81
			4.25	2.67	6.25	3.57

All results are based on $N = 50,000$ and $B = 999$. Note that $\mu = \frac{1}{B}\sum_{b=1}^{B} ARE^{(b)}$ and Δ = 97.5%-quantile − 2.5%-quantile

the last trading day of the return series and κ is a user chosen tuning parameter. In our setting, we take the same κ's as proposed in Zhu and Ling (2015). Following this procedure, provides us with a series of ARE denoted by $ARE_B := \left\{ARE^{(b)}\right\}_{b=1}^{B}$. We next calculate the sample mean, $\mu = \frac{1}{B}\sum_{b=1}^{B} ARE^{(b)}$, the 95% confidence interval, and the length of the confidence interval, Δ = 97.5%-quantile − 2.5%-quantile. Summary statistics for ARE_B of the GIM-GARCH with normal innovations is shown in Table 9.3. The results for the GIM-NGARCH and AR(3)-NGARCH model with SNG or SNIG innovations are shown in Table 9.1; respectively Table 9.2. In Figure 9.1, we show a box plot of the total ARE_B for all models.

We make the following observations: First, the sample mean of the total ARE_B for all considered specifications is comparable to the results in Badescu and Kulperger (2008) and Zhu and Ling (2015). Second, the confidence intervals of ARE_B of the GIM-NGARCH are longer than the interval for the AR-NGARCH model regardless of the chosen innovations for all maturities. Third, in the case of normal innovations, the confidence intervals of ARE_B of the standard GIM-GARCH model are widest over all maturities. Fourth, in terms of the lowest mean of the total ARE_B, the GIM-GARCH model with SNG innovations performs best; cf. Table 9.1. However, we see from Figure 9.1 that the variation of ARE_B for the GIM-NGARCH model with SNG innovations is larger than in the AR-NGARCH model with normal innovations which has a comparable mean in terms

Table 9.3: Summary statistics of ARE for normal innovations

Model	κ	TM	$\{ARE^{(b)}\}_{b=1}^{B}$			
			μ	2.5	97.5	Δ
GIM-NGARCH	0.7	22	2.08	0.66	3.85	3.19
AR-NGARCH	0.7		2.17	0.93	3.38	2.45
GIM-GARCH	0.6		2.36	1.40	4.11	3.18
		46	1.75	0.62	4.44	3.81
			1.71	0.64	3.51	2.87
			4.01	3.52	5.48	4.83
		109	2.51	2.00	3.93	1.93
			2.56	2.19	3.53	1.34
			6.01	5.37	7.71	5.52
		173	4.27	3.76	5.53	1.77
			4.31	3.97	4.88	0.91
			8.72	7.48	11.36	7.39
		234	6.12	5.29	8.40	3.11
			5.92	5.48	6.85	1.37
			12.77	9.31	19.43	13.95
		All	3.44	2.91	4.86	1.95
			3.43	2.97	4.38	1.42
			7.20	6.20	9.74	6.77

All results are based on $N = 50,000$ and $B = 999$. Note that $\mu = \frac{1}{B}\sum_{b=1}^{B} ARE^{(b)}$ and $\Delta = 97.5\%$-quantile $-$ 2.5%-quantile

of total ARE_B.

Overall, we find that, when only looking at the sample mean of the total ARE_B, the GIM-NGARCH model with SNG innovations performs best. This is the same conclusion as in Zhu and Ling (2015). Yet, when we also take the variation in the total ARE_B into account, the conclusion changes. That is, the AR-NGARCH model with normal innovations has a slighlty higher sample mean of 3.44 compared to 3.04, however, the length of the 95% confidence is 1.95 compared to 3.55 for the GIM-NGARCH model with SNG innovations. Hence, no single specification outperforms the other models when taking estimation uncertainty into account.

10 Conclusion

In this article, we discuss two aspects of the impact of estimation uncertainty on GARCH option prices. First, we design a bootstrap algorithm that allows us to generate a large number of option prices where the only source of variation stems from parameter uncertainty. We choose to model parameter uncertainty by the asymptotic distribution of the model parameters and their corresponding covariance matrix. In a numerical example, we show that the impact of parameter uncertainty is higher for options at the money.

Second, based on our first observation, we propose another bootstrap algorithm that allows us to adapt option prices, based on general GARCH models, to parameter uncertainty. In an empirical application, we demonstrate that this parameter uncertainty is also transmitted to the metric used to evaluate option prices. More precisely, the GIM-NGARCH model with SNG innovations has the lowest total ARE on average. However, compared to the comparable average total ARE of the AR(3)-NGARCH model with normal innovations, the total ARE of this GIM-NGARCH is subject to higher variation. As a result, no single GARCH specification dominates when we take parameter uncertainty into account.

Further issues are left for future research. First, it would be interesting to increase our sample size and see how our method performs on a larger data set. Second, one could try to implement the bootstrap approach of Pascual, Romo, and Ruiz (2006) to account for parameter uncertainty. For this approach the model parameters are re-estimated in each bootstrap step which is a way to non parametrically approximate the distribution of the model parameters. However, this method also comes with a high computational cost which might be even more time consuming in the context of option pricing. Finally, an interesting question is to find a metric for option pricing evaluation which takes explicitly estimation uncertainty into account or is robust against it.

Appendix

A Processes for a risky asset under $\mathbb{Q}$

In this section we follow Zhu and Ling (2015) and provide an overview of the processes of an asset return under $\mathbb{Q}$ when z_t is conditionally (i) normal, (ii) shift negative gamma (SNG), or (iii) shift negative inverse Gaussian (SNIG).

A.1 normal innovations

First, we study the return process of an asset under $\mathbb{Q}$ when the innovations follow a standard normal distribution. That is, suppose that $z_t|\mathcal{F}_{t-1} \sim N(0,1)$. Then, the moment generating function under measure $\mathbb{P}$ is

$$M^{\mathbb{P}}_{y_t|\mathcal{F}_{t-1}}(z) = \exp\left(z\mu_t + \frac{z^2 h_t}{2}\right).$$

Due to eq. (2.4), we find that

$$\delta_t = \frac{1-\mu_t}{h_t} - \frac{1}{2}$$

such that

$$M^{\mathbb{Q}}_{y_t|\mathcal{F}_{t-1}}(z) = \exp\left[z\left(r - \frac{h_t}{2}\right) + \frac{z^2 h_t}{2}\right] \tag{A.1}$$

by eq. (2.3). Finally, under $\mathbb{Q}$,

$$y_t = r - \frac{h_t}{2} + \epsilon^*_t, \tag{A.2}$$

where $\epsilon^*_t|\mathcal{F}_{t-1} \sim N(0, h_t)$; cf. Zhu and Ling (2015). Note that this result is the same as in the seminal paper of Duan (1995) where the local risk-neutral measure was used instead of the Esscher transform.

A.2 Shift Negative Gamma innovations

We also consider the case where the innovations follow a Shift Negative Gamma (SNG) distribution. That is, $z_t = (\xi_t + a)/\sqrt{a}$, where $a > 0$ and $\xi_t|\mathcal{F}_{t-1} \sim -\mathrm{G}(a,1)$ with $\mathrm{G}(a,b)$ being a Gamma distribution with shape parameter a and scale parameter b. Then, we say that z_t is conditionally SNG which is denoted by $z_t|\mathcal{F}_{t-1} \sim \mathrm{SNG}(a)$. Compared to the Gaussian distribution, the SNG distribution is skewed and has fatter tails; cf Siu, Howell, and Yang (2004).

Assume that $z_t|\mathcal{F}_{t-1} \sim \text{SNG}(a)$ in eq. (2.1), then

$$y_t = \mu_t + \epsilon_t \tag{A.3}$$

where $\epsilon_t = \sqrt{ah_t} + \sqrt{h_t/a}\xi_t$ under $\mathbb{P}$. Following Zhu and Ling (2015) and using eq. (2.2), we find that

$$M^{\mathbb{P}}_{y_t|\mathcal{F}_{t-1}}(z) = \frac{a^{a/2}\exp\left[z\left(\mu_t + \sqrt{ah_t}\right)\right]}{\left(\sqrt{a} + z\sqrt{h_t}\right)^a}, \tag{A.4}$$

for $z > -\frac{a}{h_t}$. Next, by eq. (2.4), we have that $\delta_t = b_t - \sqrt{ah_t}$, where

$$b_t = \left[\exp\left(\frac{\mu_t - r + \sqrt{ah_t}}{a}\right) - 1\right]^{-1}. \tag{A.5}$$

Given δ_t and eq. (2.3), it follows that

$$M^{\mathbb{Q}}_{y_t|\mathcal{F}_{t-1}}(z) = \frac{\exp\left[z\left(\mu_t + \sqrt{ah_t}\right)\right]}{(1 + z/b)^a}, \tag{A.6}$$

for $z > -b_t$ if $b_t > 0$. Consequently, we have that under $\mathbb{Q}$,

$$y_t = \mu_t + \epsilon^*_t, \tag{A.7}$$

where $\epsilon^*_t = \sqrt{ah_t} + \xi^*_t$, where $\xi^*_t|\mathcal{F}_{t-1} \sim -\text{G}(a, b)$; cf. Zhu and Ling (2015).

A.3 Shift negative inverse Gaussian innovations

As a last alternative, we assume that $z_t = (\xi_t + d)/\sqrt{d}$ with $d > 0$, $\xi_t|\mathcal{F}_{t-1} \sim -\text{IG}(d)$, where $\text{IG}(d)$ denotes an inverse Gaussian parameter with parameter d. Here, we say that z_t is conditionally shift negative inverse Gaussian (SNIG) and denote this by $z_t|\mathcal{F}_{t-1} \sim \text{SNIG}(d)$. Again, as the SNG distribution, the SNIG distribution also has fatter tails and is more skewed than the normal distribution.

Assume that $z_t|\mathcal{F}_{t-1} \sim \text{SNIG}(d)$ in eq. (2.1), then

$$y_t = \mu_t + \epsilon_t \tag{A.8}$$

with $\epsilon_t = \sqrt{dh_t} + \sqrt{h_t/d}\xi_t$ under $\mathbb{P}$. As before, we use eq. (2.2), (2.3), and (2.4) to see that

$$M^{\mathbb{Q}}_{y_t|\mathcal{F}_{t-1}}(z) = \exp\left\{\left(\mu_t + \sqrt{dh_t}\right)z + d\left[\sqrt{1 + 2\sqrt{h_t/d}\delta_t} - \sqrt{1 + 2\sqrt{h_t/d}\,(z + \delta_t)}\right]\right\} \tag{A.9}$$

where for $\delta_t \in \mathcal{F}_{t-1}$ it holds that

$$1 + 2\sqrt{h_t/d}\delta_t = \frac{1}{4}\left(\frac{r - \mu_t - \sqrt{dh_t}}{2d} - \frac{2\sqrt{dh_t}}{r - \mu_t - \sqrt{dh_t}}\right)^2 =: c_t; \tag{A.10}$$

see Zhu and Ling (2015). As a result, it follows under $\mathbb{Q}$ that

$$y_t = \mu_t + \epsilon_t^*, \tag{A.11}$$

where $\epsilon_t^* = \sqrt{dh_t} + c_t^{-1}\sqrt{h_t/d}\xi_t^*$ with $\xi_t^*|\mathcal{F}_{t-1} \sim -\text{IG}\left(\sqrt{c_t}d\right)$.

B Additional simulation and empirical results

Table B.1: Estimation results of different GARCH models. Robust standard errors are given in parenthesis.

	GIM-GARCH	GIM-NGARCH	AR-NGARCH
ν	0.064 (0.016)	0.034 (0.016)	
ϕ_0			$1.4e^{-4}$ (0.015)
ϕ_1			0.042 (0.018)
ϕ_2			$1.3e^{-4}$ (0.023)
ϕ_3			-0.041 (0.022)
ω	$3.7e^{-6}$ (0.000)	$5.68e^{-6}$ (0.000)	$6.99e^{-6}$ (0.000)
α	0.123 (0.061)	0.127 (0.054)	0.131 (0.039)
β	0.847 (0.085)	0.787 (0.081)	0.762 (0.061)
θ		0.532 (0.226)	0.599 (0.227)
a		39.22	68.43
d		88.25	153.76

Table B.2: p-values of the Ljung-Box test for serial autocorrelation in the residuals $\hat{z}_t$; respectively squared residuals $\hat{z}_t$

Model		5	10	20	50
GIM-GARCH	$\hat{z}_t$	0.010	0.027	0.017	0.015
	$\hat{z}_t^2$	0.639	0.917	0.993	0.999
GIM-NGARCH	$\hat{z}_t$	0.013	0.024	0.020	0.014
	$\hat{z}_t^2$	0.425	0.821	0.973	0.964
AR-NGARCH	$\hat{z}_t$	0.633	0.458	0.198	0.058
	$\hat{z}_t^2$	0.504	0.864	0.947	0.797

Table B.3: Summary statistics for the GIM-GARCH model

TM	K	μ	median	σ	2.5%	97.5%	Δ
21	1700	291.286	291.286	0.017	291.250	291.321	0.071
	1750	241.814	241.814	0.033	241.746	241.881	0.135
	1800	193.018	193.020	0.062	192.890	193.141	0.251
	1850	145.793	145.796	0.111	145.560	146.009	0.449
	1900	102.035	102.040	0.177	101.654	102.382	0.727
	1950	64.527	64.532	0.239	64.018	64.992	0.974
	2000	36.305	36.311	0.258	35.752	36.818	1.065
	2050	18.350	18.355	0.222	17.875	18.791	0.916
	2100	8.651	8.654	0.157	8.314	8.965	0.651
	2150	3.963	3.963	0.099	3.750	4.161	0.411
	2200	1.853	1.853	0.057	1.731	1.968	0.237
	2250	0.893	0.892	0.034	0.822	0.960	0.138
45		293.748	293.749	0.089	293.567	293.928	0.361
		245.523	245.525	0.136	245.243	245.797	0.554
		198.734	198.737	0.203	198.315	199.137	0.823
		154.559	154.564	0.287	153.957	155.128	1.170
		114.491	114.499	0.370	113.710	115.224	1.514
		80.208	80.217	0.434	79.292	81.067	1.775
		53.094	53.102	0.454	52.133	53.993	1.859
		33.348	33.354	0.424	32.450	34.187	1.737
		20.066	20.070	0.357	19.310	20.774	1.464
		11.754	11.758	0.279	11.162	12.308	1.147
		6.831	6.832	0.205	6.397	7.242	0.845
		4.017	4.018	0.146	3.709	4.313	0.603
64		296.161	296.162	0.163	295.830	296.489	0.658
		249.000	249.004	0.234	248.521	249.467	0.946
		203.740	203.745	0.320	203.078	204.379	1.300
		161.394	161.400	0.414	160.534	162.219	1.684
		123.180	123.192	0.502	122.135	124.180	2.045
		90.265	90.277	0.564	89.089	91.389	2.299
		63.496	63.508	0.583	62.280	64.657	2.377
		42.940	42.956	0.556	41.781	44.049	2.269
		28.036	28.045	0.491	27.012	29.015	2.003
		17.831	17.833	0.407	16.980	18.644	1.665
		11.227	11.229	0.319	10.557	11.867	1.309
		7.058	7.059	0.244	6.546	7.549	1.003
82		298.868	298.873	0.242	298.379	299.356	0.978
		252.681	252.687	0.328	252.011	253.337	1.326
		208.642	208.648	0.427	207.765	209.497	1.732
		167.703	167.711	0.525	166.621	168.748	2.128
		130.766	130.774	0.613	129.500	131.989	2.488
		98.684	98.699	0.671	97.298	100.026	2.728
		72.065	72.085	0.690	70.640	73.444	2.804
		50.993	51.012	0.664	49.620	52.323	2.703
		35.136	35.151	0.604	33.887	36.346	2.458
		23.630	23.637	0.521	22.550	24.675	2.125
		15.643	15.647	0.429	14.754	16.504	1.750
		10.266	10.268	0.341	9.558	10.954	1.395
125		305.376	305.384	0.415	304.533	306.214	1.681
		261.237	261.244	0.521	260.174	262.286	2.112
		219.544	219.555	0.630	218.258	220.802	2.544
		180.979	180.989	0.733	179.482	182.432	2.950
		146.113	146.129	0.820	144.440	147.740	3.300
		115.531	115.548	0.878	113.742	117.275	3.534
		89.422	89.439	0.895	87.596	91.206	3.609
		67.753	67.771	0.877	65.964	69.503	3.539
		50.379	50.399	0.826	48.696	52.032	3.336
		36.870	36.888	0.746	35.349	38.366	3.016
		26.679	26.690	0.655	25.344	27.994	2.650
		19.100	19.103	0.562	17.953	20.229	2.276

All results are based on $N = 50,000$ and $B = 999$. Note that $\mu = \frac{1}{B}\sum_{b=1}^{B} \tilde{V}_T^{*(b)}$, $\sigma = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\tilde{V}_T^{*(b)} - \mu\right)^2}$, and Δ = 97.5%-quantile − 2.5%-quantile.

Table B.4: Summary statistics for the AR(1)-GARCH model

TM	K	μ	median	σ	2.5%	97.5%	Δ
21	1700	291.186	291.186	0.015	291.157	291.217	0.060
	1750	241.651	241.650	0.030	241.593	241.712	0.119
	1800	192.773	192.773	0.059	192.659	192.894	0.235
	1850	145.489	145.488	0.111	145.274	145.712	0.437
	1900	101.795	101.792	0.183	101.437	102.158	0.720
	1950	64.550	64.545	0.250	64.059	65.040	0.981
	2000	36.694	36.692	0.273	36.155	37.228	1.073
	2050	18.943	18.942	0.237	18.477	19.413	0.936
	2100	9.207	9.204	0.173	8.872	9.552	0.680
	2150	4.370	4.370	0.111	4.157	4.597	0.440
	2200	2.110	2.109	0.067	1.980	2.249	0.269
	2250	1.049	1.048	0.041	0.971	1.133	0.162
45		293.303	293.302	0.076	293.157	293.460	0.304
		244.916	244.915	0.125	244.675	245.174	0.498
		197.978	197.975	0.193	197.608	198.375	0.767
		153.723	153.721	0.283	153.177	154.304	1.126
		113.730	113.727	0.375	113.008	114.493	1.485
		79.708	79.705	0.446	78.849	80.612	1.763
		52.955	52.951	0.471	52.046	53.910	1.864
		33.534	33.529	0.443	32.678	34.434	1.755
		20.455	20.452	0.377	19.730	21.227	1.497
		12.204	12.203	0.299	11.630	12.820	1.189
		7.245	7.245	0.223	6.817	7.710	0.893
		4.352	4.351	0.163	4.038	4.694	0.656
64		295.385	295.382	0.141	295.113	295.674	0.561
		248.001	247.998	0.214	247.590	248.437	0.847
		202.559	202.556	0.304	201.972	203.177	1.204
		160.130	160.127	0.405	159.354	160.948	1.594
		121.988	121.980	0.501	121.031	122.998	1.967
		89.299	89.293	0.571	88.204	90.449	2.245
		62.879	62.872	0.595	61.735	64.081	2.346
		42.692	42.687	0.571	41.596	43.851	2.254
		28.091	28.085	0.508	27.118	29.129	2.010
		18.068	18.061	0.426	17.249	18.943	1.694
		11.530	11.527	0.338	10.886	12.229	1.344
		7.366	7.363	0.262	6.865	7.914	1.048
82		297.761	297.757	0.212	297.357	298.196	0.838
		251.317	251.314	0.299	250.744	251.926	1.181
		207.074	207.070	0.404	206.301	207.893	1.592
		166.047	166.042	0.511	165.069	167.075	2.006
		129.192	129.187	0.606	128.036	130.410	2.374
		97.325	97.318	0.671	96.043	98.672	2.629
		71.040	71.035	0.696	69.709	72.440	2.731
		50.345	50.341	0.675	49.055	51.709	2.654
		34.826	34.822	0.617	33.651	36.079	2.428
		23.577	23.570	0.536	22.560	24.674	2.114
		15.740	15.731	0.445	14.891	16.659	1.768
		10.446	10.440	0.357	9.767	11.190	1.423
125		303.587	303.578	0.366	302.897	304.338	1.440
		259.162	259.153	0.478	258.265	260.140	1.875
		217.278	217.268	0.594	216.161	218.484	2.323
		178.620	178.610	0.706	177.292	180.049	2.757
		143.792	143.782	0.800	142.284	145.406	3.122
		113.405	113.392	0.866	111.774	115.151	3.377
		87.595	87.584	0.890	85.920	89.391	3.471
		66.283	66.271	0.876	64.634	68.058	3.424
		49.292	49.278	0.828	47.735	50.975	3.240
		36.115	36.103	0.751	34.703	37.651	2.948
		26.187	26.175	0.662	24.945	27.547	2.603
		18.810	18.798	0.571	17.742	19.992	2.250

All results are based on $N = 50{,}000$ and $B = 999$. Note that $\mu = \frac{1}{B}\sum_{b=1}^{B} \tilde{V}_T^{*(b)}$, $\sigma = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\tilde{V}_T^{*(b)} - \mu\right)^2}$, and Δ = 97.5%-quantile − 2.5%-quantile.

Table B.5: 95% prediction intervals for GIM-GARCH with normal innovations

K	TM = 22	TM = 46	TM = 109	TM = 173	TM = 234
975			161.60	173.30	
			[156.64, 158.48]	[163.54, 166.96]	
995			144.80	157.00	
			[138.78, 141.03]	[146.87, 150.81]	
1025			120.10	133.10	146.50
			[113.40, 116.28]	[123.39, 128.03]	[132.16, 138.27]
1050		84.50	100.70	114.80	
		[81.15, 82.26]	[93.93, 97.31]	[105.43, 110.54]	
1075		64.30	82.50	97.60	
		[60.82, 62.28]	[76.33, 80.08]	[89.05, 94.53]	
1090	43.10				
	[41.68, 42.39]				
1100	35.60		65.50	81.20	
	[34.28, 35.09]		[60.72, 64.77]	[74.30, 80.04]	
1110		39.50			
		[37.14, 38.94]			
1120	22.90	33.50			
	[21.82, 22.73]	[31.63, 33.44]			
1125	20.20	30.70	51.00	66.90	81.70
	[19.23, 20.15]	[29.09, 30.90]	[47.36, 51.49]	[61.29, 67.17]	[72.35, 79.80]
1130		28.00			
		[26.69, 28.51]			
1135		25.60	45.50		
		[24.44, 26.25]	[42.64, 46.76]		
1140	13.30	23.20		58.90	
	[12.77, 13.66]	[22.32, 24.12]		[54.31, 60.24]	
1150		19.10	38.10	53.90	68.30
		[18.50, 20.26]	[36.21, 40.30]	[50.01, 55.92]	[61.00, 68.45]
1160		15.30			
		[15.22, 16.90]			
1170		12.10			
		[12.42, 14.01]			
1175		10.90	27.70	42.50	56.60
		[11.20, 12.74]	[27.22, 31.07]	[40.40, 46.12]	[51.06, 58.45]
1200			19.60	33.00	46.10
			[20.11, 23.66]	[32.26, 37.73]	[42.42, 49.67]
1225			13.20	24.90	36.90
			[14.66, 17.84]	[25.53, 30.68]	[35.02, 42.03]
1250				18.30	29.30
				[20.10, 24.83]	[28.75, 35.42]
1275				13.20	22.50
				[15.74, 19.98]	[23.49, 29.74]
1300					17.20
					[19.08, 24.93]
1325					12.80
					[15.47, 20.81]

All results are based on $N = 50,000$ and $B = 999$.

Table B.6: 95% prediction intervals for GIM-NGARCH with normal innovations

K	TM = 22	TM = 46	TM = 109	TM = 173	TM = 234
975			161.60	173.30	
			[159.06, 161.03]	[166.67, 170.00]	
995			144.80	157.00	
			[141.41, 143.63]	[150.14, 153.73]	
1025			120.10	133.10	146.50
			[116.18, 118.79]	[126.60, 130.57]	[135.63, 140.58]
1050		84.50	100.70	114.80	
		[82.83, 84.01]	[96.56, 99.51]	[108.34, 112.58]	
1075		64.30	82.50	97.60	
		[62.43, 63.82]	[78.55, 81.73]	[91.50, 95.94]	
1090	43.10				
	[42.52, 43.22]				
1100	35.60		65.50	81.20	
	[34.98, 35.74]		[62.40, 65.67]	[76.19, 80.74]	
1110		39.50			
		[38.04, 39.59]			
1120	22.90	33.50			
	[21.99, 22.81]	[32.19, 33.75]			
1125	20.20	30.70	51.00	66.90	81.70
	[19.24, 20.07]	[29.46, 31.03]	[48.29, 51.58]	[62.48, 67.02]	[73.88, 79.41]
1130		28.00			
		[26.89, 28.44]			
1135		25.60	45.50		
		[24.45, 26.00]	[43.23, 46.51]		
1140	13.30	23.20		58.90	
	[12.33, 13.08]	[22.17, 23.69]		[55.06, 59.55]	
1150		19.10	38.10	53.90	68.30
		[18.03, 19.48]	[36.33, 39.51]	[50.45, 54.88]	[61.87, 67.34]
1160		15.30			
		[14.47, 15.82]			
1170		12.10			
		[11.44, 12.67]			
1175		10.90	27.70	42.50	56.60
		[10.11, 11.30]	[26.55, 29.46]	[40.10, 44.32]	[51.27, 56.55]
1200			19.60	33.00	46.10
			[18.82, 21.38]	[31.37, 35.28]	[42.02, 47.04]
1225			13.20	24.90	36.90
			[12.92, 15.09]	[24.11, 27.68]	[34.07, 38.75]
1250				18.30	29.30
				[18.22, 21.39]	[27.31, 31.61]
1275				13.20	22.50
				[13.54, 16.31]	[21.68, 25.56]
1300					17.20
					[17.00, 20.49]
1325					12.80
					[13.19, 16.26]

All results are based on $N = 50,000$ and $B = 999$.

Table B.7: 95% prediction intervals for GIM-NGARCH with SNG innovations

K	TM = 22	TM = 46	TM = 109	TM = 173	TM = 234
975			161.60	173.30	
			[159.88, 162.24]	[167.74, 171.46]	
995			144.80	157.00	
			[142.24, 144.87]	[151.19, 155.16]	
1025			120.10	133.10	146.50
			[116.96, 119.95]	[127.53, 131.97]	[136.64, 142.17]
1050		84.50	100.70	114.80	
		[83.28, 84.75]	[97.29, 100.51]	[109.12, 113.84]	
1075		64.30	82.50	97.60	
		[62.80, 64.49]	[79.11, 82.52]	[92.08, 97.00]	
1090	43.10				
	[42.74, 43.68]				
1100	35.60		65.50	81.20	
	[35.13, 36.12]		[62.70, 66.26]	[76.57, 81.51]	
1110		39.50			
		[38.09, 39.94]			
1120	22.90	33.50			
	[21.94, 22.96]	[32.11, 33.97]			
1125	20.20	30.70	51.00	66.90	81.70
	[19.13, 20.14]	[29.33, 31.19]	[48.31, 51.86]	[62.65, 67.55]	[74.13, 80.03]
1130		28.00			
		[26.70, 28.54]			
1135		25.60	45.50		
		[24.23, 26.02]	[43.19, 46.64]		
1140	13.30	23.20		58.90	
	[12.01, 12.98]	[21.89, 23.64]		[55.05, 59.96]	
1150		19.10	38.10	53.90	68.30
		[17.65, 19.32]	[36.10, 39.49]	[50.34, 55.20]	[61.90, 67.69]
1160		15.30			
		[13.99, 15.54]			
1170		12.10			
		[10.89, 12.32]			
1175		10.90	27.70	42.50	56.60
		[9.54, 10.91]	[26.04, 29.16]	[39.71, 44.39]	[51.07, 56.69]
1200			19.60	33.00	46.10
			[18.12, 20.92]	[30.76, 35.06]	[41.53, 46.94]
1225			13.20	24.90	36.90
			[12.13, 14.53]	[23.37, 27.22]	[33.39, 38.45]
1250				18.30	29.30
				[17.35, 20.76]	[26.48, 31.22]
1275				13.20	22.50
				[12.62, 15.59]	[20.77, 25.06]
1300					17.20
					[16.06, 19.85]
1325					12.80
					[12.25, 15.57]

All results are based on $N = 50,000$ and $B = 999$.

Table B.8: 95% prediction intervals for GIM-NGARCH with SNIG innovations

K	TM = 22	TM = 46	TM = 109	TM = 173	TM = 234
975			161.60	173.30	
			[159.69, 161.53]	[167.45, 170.37]	
995			144.80	157.00	
			[142.03, 144.12]	[150.86, 154.01]	
1025			120.10	133.10	146.50
			[116.74, 119.14]	[127.19, 130.69]	[136.23, 140.61]
1050		84.50	100.70	114.80	
		[83.29, 84.37]	[97.04, 99.67]	[108.78, 112.55]	
1075		64.30	82.50	97.60	
		[62.77, 64.08]	[78.81, 81.66]	[91.76, 95.69]	
1090	43.10				
	[42.79, 43.48]				
1100	35.60		65.50	81.20	
	[35.18, 35.93]		[62.42, 65.33]	[76.21, 80.20]	
1110		39.50			
		[38.07, 39.53]			
1120	22.90	33.50			
	[21.98, 22.79]	[32.10, 33.57]			
1125	20.20	30.70	51.00	66.90	81.70
	[19.17, 19.99]	[29.32, 30.78]	[48.05, 50.98]	[62.31, 66.32]	[73.68, 78.62]
1130		28.00			
		[26.68, 28.15]			
1135		25.60	45.50		
		[24.20, 25.64]	[42.90, 45.80]		
1140	13.30	23.20		58.90	
	[12.08, 12.84]	[21.86, 23.29]		[54.76, 58.74]	
1150		19.10	38.10	53.90	68.30
		[17.62, 19.00]	[35.86, 38.67]	[50.07, 53.99]	[61.46, 66.34]
1160		15.30			
		[13.97, 15.26]			
1170		12.10			
		[10.88, 12.04]			
1175		10.90	27.70	42.50	56.60
		[9.54, 10.64]	[25.86, 28.46]	[39.49, 43.23]	[50.62, 55.39]
1200			19.60	33.00	46.10
			[17.98, 20.28]	[30.54, 34.06]	[41.20, 45.76]
1225			13.20	24.90	36.90
			[12.04, 13.95]	[23.17, 26.36]	[33.09, 37.37]
1250				18.30	29.30
				[17.22, 20.05]	[26.26, 30.14]
1275				13.20	22.50
				[12.52, 14.97]	[20.56, 24.06]
1300					17.20
					[15.88, 18.99]
1325					12.80
					[12.10, 14.81]

All results are based on $N = 50,000$ and $B = 999$.

Table B.9: 95% prediction intervals for AR-NGARCH with normal innovations

K	TM = 22	TM = 46	TM = 109	TM = 173	TM = 234
975			161.60	173.30	
			[158.82, 160.11]	[166.32, 168.35]	
995			144.80	157.00	
			[141.19, 142.66]	[149.80, 152.01]	
1025			120.10	133.10	146.50
			[116.02, 117.72]	[126.33, 128.80]	[135.21, 138.31]
1050		84.50	100.70	114.80	
		[82.87, 83.71]	[96.49, 98.37]	[108.14, 110.80]	
1075		64.30	82.50	97.60	
		[62.53, 63.52]	[78.55, 80.58]	[91.39, 94.18]	
1090	43.10				
	[42.61, 43.13]				
1100	35.60		65.50	81.20	
	[35.08, 35.64]		[62.47, 64.56]	[76.17, 79.05]	
1110		39.50			
		[38.23, 39.34]			
1120	22.90	33.50			
	[22.10, 22.70]	[32.40, 33.51]			
1125	20.20	30.70	51.00	66.90	81.70
	[19.36, 19.96]	[29.69, 30.79]	[48.44, 50.52]	[62.56, 65.43]	[73.75, 77.24]
1130		28.00			
		[27.12, 28.20]			
1135		25.60	45.50		
		[24.70, 25.76]	[43.42, 45.48]		
1140	13.30	23.20		58.90	
	[12.42, 12.98]	[22.41, 23.45]		[55.19, 58.04]	
1150		19.10	38.10	53.90	68.30
		[18.27, 19.27]	[36.55, 38.54]	[50.61, 53.42]	[61.80, 65.24]
1160		15.30			
		[14.68, 15.62]			
1170		12.10			
		[11.63, 12.50]			
1175		10.90	27.70	42.50	56.60
		[10.29, 11.11]	[26.75, 28.60]	[40.28, 42.95]	[51.28, 54.58]
1200			19.60	33.00	46.10
			[19.00, 20.66]	[31.55, 34.04]	[42.11, 45.21]
1225			13.20	24.90	36.90
			[13.08, 14.49]	[24.31, 26.58]	[34.19, 37.07]
1250				18.30	29.30
				[18.44, 20.43]	[27.46, 30.10]
1275				13.20	22.50
				[13.75, 15.46]	[21.82, 24.20]
1300					17.20
					[17.16, 19.29]
1325					12.80
					[13.34, 15.25]

All results are based on $N = 50,000$ and $B = 999$.

Table B.10: 95% prediction intervals for AR-NGARCH with SNG innovations

K	TM = 22	TM = 46	TM = 109	TM = 173	TM = 234
975			161.60	173.30	
			[160.04, 161.79]	[168.14, 170.71]	
995			144.80	157.00	
			[142.54, 144.48]	[151.77, 154.52]	
1025			120.10	133.10	146.50
			[117.56, 119.77]	[128.43, 131.47]	[137.80, 141.40]
1050		84.50	100.70	114.80	
		[83.73, 85.00]	[98.13, 100.58]	[110.32, 113.56]	
1075		64.30	82.50	97.60	
		[63.48, 64.89]	[80.27, 82.77]	[93.59, 96.99]	
1090	43.10				
	[43.19, 44.01]				
1100	35.60		65.50	81.20	
	[35.68, 36.53]		[64.17, 66.71]	[78.29, 81.80]	
1110		39.50			
		[39.19, 40.65]			
1120	22.90	33.50			
	[22.62, 23.48]	[33.30, 34.78]			
1125	20.20	30.70	51.00	66.90	81.70
	[19.84, 20.69]	[30.55, 32.00]	[50.01, 52.54]	[64.57, 68.05]	[76.35, 80.36]
1130		28.00			
		[27.96, 29.36]			
1135		25.60	45.50		
		[25.47, 26.85]	[44.94, 47.42]		
1140	13.30	23.20		58.90	
	[12.75, 13.56]	[23.16, 24.48]		[57.12, 60.54]	
1150		19.10	38.10	53.90	68.30
		[18.90, 20.18]	[37.93, 40.32]	[52.47, 55.83]	[64.27, 68.23]
1160		15.30			
		[15.19, 16.39]			
1170		12.10			
		[12.03, 13.13]			
1175		10.90	27.70	42.50	56.60
		[10.63, 11.69]	[27.86, 30.12]	[41.98, 45.20]	[53.59, 57.37]
1200			19.60	33.00	46.10
			[19.86, 21.87]	[32.98, 36.05]	[44.13, 47.79]
1225			13.20	24.90	36.90
			[13.69, 15.45]	[25.47, 28.30]	[35.95, 39.41]
1250				18.30	29.30
				[19.33, 21.86]	[28.99, 32.18]
1275				13.20	22.50
				[14.40, 16.65]	[23.06, 26.03]
1300					17.20
					[18.17, 20.85]
1325					12.80
					[14.13, 16.54]

All results are based on $N = 50,000$ and $B = 999$.

Table B.11: 95% prediction intervals for AR-NGARCH with SNG innovations

K	TM = 22	TM = 46	TM = 109	TM = 173	TM = 234
975			161.60	173.30	
			[158.99, 160.24]	[166.36, 168.30]	
995			144.80	157.00	
			[141.32, 142.73]	[149.76, 151.89]	
1025			120.10	133.10	146.50
			[116.02, 117.67]	[126.13, 128.51]	[134.83, 137.83]
1050		84.50	100.70	114.80	
		[83.02, 83.86]	[96.35, 98.18]	[107.80, 110.33]	
1075		64.30	82.50	97.60	
		[62.57, 63.56]	[78.28, 80.25]	[90.86, 93.53]	
1090	43.10				
	[42.73, 43.27]				
1100	35.60		65.50	81.20	
	[35.14, 35.73]		[62.02, 64.04]	[75.44, 78.17]	
1110		39.50			
		[37.99, 39.11]			
1120	22.90	33.50			
	[22.01, 22.62]	[32.08, 33.21]			
1125	20.20	30.70	51.00	66.90	81.70
	[19.22, 19.83]	[29.33, 30.45]	[47.80, 49.84]	[61.68, 64.42]	[72.77, 76.17]
1130		28.00			
		[26.72, 27.83]			
1135		25.60	45.50		
		[24.26, 25.35]	[42.72, 44.73]		
1140	13.30	23.20		58.90	
	[12.16, 12.72]	[21.94, 23.02]		[54.22, 56.92]	
1150		19.10	38.10	53.90	68.30
		[17.74, 18.76]	[35.74, 37.69]	[49.59, 52.26]	[60.70, 64.04]
1160		15.30			
		[14.11, 15.06]			
1170		12.10			
		[11.04, 11.90]			
1175		10.90	27.70	42.50	56.60
		[9.70, 10.52]	[25.85, 27.65]	[39.15, 41.75]	[50.03, 53.28]
1200			19.60	33.00	46.10
			[18.06, 19.64]	[30.32, 32.78]	[40.73, 43.85]
1225			13.20	24.90	36.90
			[12.17, 13.50]	[23.08, 25.29]	[32.77, 35.67]
1250				18.30	29.30
				[17.22, 19.18]	[26.04, 28.67]
1275				13.20	22.50
				[12.59, 14.27]	[20.43, 22.78]
1300					17.20
					[15.82, 17.91]
1325					12.80
					[12.10, 13.91]

All results are based on $N = 50,000$ and $B = 999$.

References

Aastveit, K. A., C. Foroni, and F. Ravazzolo (2017): "Density Forecasts With Midas Models," *Journal of Applied Econometrics*, 32(4), 783–801.

Agosto, A., G. Cavaliere, D. Kristensen, and A. Rahbek (2016): "Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)," *Journal of Empirical Finance*, 38, 640–663.

Ahmad, A., and C. Francq (2016): "Poisson QMLE of Count Time Series Models," *Journal of Time Series Analysis*, 37(3), 291–314.

Angelini, G., and L. De Angelis (2017): "PARX model for football match predictions," *Journal of Forecasting*, 36(7), 795–807.

Badescu, A., R. J. Elliott, R. Kulperger, J. Miettinen, and T. K. Siu (2011): "A comparison of pricing kernels for garch option pricing with generalized hyperbolic distributions," *International Journal of Theoretical and Applied Finance*, 14(05), 669–708.

Badescu, A. M., and R. J. Kulperger (2008): "GARCH option pricing: A semiparametric approach," *Insurance: Mathematics and Economics*, 43(1), 69–84.

Bams, D., T. Lehnert, and C. C. P. Wolff (2005): "An evaluation framework for alternative VaR-models," *Journal of International Money and Finance*, 24(6), 944–958.

Barone-Adesi, G., R. F. Engle, and L. Mancini (2008): "A GARCH Option Pricing Model with Filtered Historical Simulation," SSRN Scholarly Paper ID 603382, Social Science Research Network, Rochester, NY.

Blasques, F., S. J. Koopman, K. Lasak, and A. Lucas (2015): "In-Sample Confidence Bands and Out-of-Sample Forecast Bands for Time-Varying Parameters in Observation Driven Models," SSRN Scholarly Paper ID 2629121, Social Science Research Network, Rochester, NY.

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

Chen, B., Y. R. Gel, N. Balakrishna, and B. Abraham (2011): "Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes," *Journal of Forecasting*, 30(1), 51–71.

Chorro, C., D. Guégan, and F. Ielpo (2012): "Option pricing for GARCH-type models with generalized hyperbolic innovations," *Quantitative Finance*, 12(7), 1079–1094.

——— (2015): *A Time Series Approach to Option Pricing: Models, Methods and Empirical Performances*. Springer-Verlag, Berlin Heidelberg.

Christoffersen, P., S. Heston, and K. Jacobs (2006): "Option valuation with conditional skewness," *Journal of Econometrics*, 131(1), 253–284.

Christoffersen, P. F. (1998): "Evaluating Interval Forecasts," *International Economic*

Review, 39(4), 841–862.

Christou, V., and K. Fokianos (2015a): "Estimation and testing linearity for non-linear mixed poisson autoregressions," *Electronic Journal of Statistics*, 9(1), 1357–1377.

——— (2015b): "On count time series prediction," *Journal of Statistical Computation and Simulation*, 85(2), 357–373.

Conrad, C., and E. Mammen (2016): "Asymptotics for parametric GARCH-in-Mean models," *Journal of Econometrics*, 194(2), 319–329.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*. OUP Oxford, Oxford.

Dotsis, G., and R. N. Markellos (2007): "The finite sample properties of the GARCH option pricing model," *Journal of Futures Markets*, 27(6), 599–615.

Doukhan, P., K. Fokianos, and D. Tjøstheim (2012): "On weak dependence conditions for Poisson autoregressions," *Statistics & Probability Letters*, 82(5), 942–948.

——— (2013): "Correction to On weak dependence conditions for Poisson autoregressions [Statist. Probab. Lett. 82 (2012) 942–948]," *Statistics & Probability Letters*, 83(8), 1926–1927.

Doukhan, P., and O. Wintenberger (2008): "Weakly dependent chains with infinite memory," *Stochastic Processes and their Applications*, 118(11), 1997–2013.

Duan, J.-C. (1995): "The Garch Option Pricing Model," *Mathematical Finance*, 5(1), 13–32.

Duan, J.-c., and J.-g. Simonato (1998): "Empirical Martingale Simulation for Asset Prices," *Management Science*, 44(9), 1218–1233.

Duan, J.-C., and H. Zhang (2001): "Pricing Hang Seng Index options around the Asian financial crisis – A GARCH approach," *Journal of Banking & Finance*, 25(11), 1989–2014.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

Engle, R. F., and V. K. Ng (1993): "Measuring and Testing the Impact of News on Volatility," *The Journal of Finance*, 48(5), 1749–1778.

Ferland, R., A. Latour, and D. Oraichi (2006): "Integer-Valued GARCH Process," *Journal of Time Series Analysis*, 27(6), 923–942.

Fokianos, K., and R. Fried (2010): "Interventions in INGARCH processes," *Journal of Time Series Analysis*, 31(3), 210–225.

Fokianos, K., and M. H. Neumann (2013): "A goodness-of-fit test for Poisson count processes," *Electronic Journal of Statistics*, 7, 793–819.

Fokianos, K., A. Rahbek, and D. Tjøstheim (2009): "Poisson Autoregression," *Journal of the American Statistical Association*, 104(488), 1430–1439.

Francq, C., and J.-M. Zakoïan (2004): "Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes," *Bernoulli*, 10(4), 605–637.

Franke, J., J.-P. Kreiss, and E. Mammen (2002): "Bootstrap of kernel smoothing in nonlinear time series," *Bernoulli*, 8(1), 1–37.

Franke, J., J.-P. Kreiss, E. Mammen, and M. H. Neumann (2002): "Properties of the nonparametric autoregressive bootstrap," *Journal of Time Series Analysis*, 23(5), 555–585.

Franke, J., M. H. Neumann, and J.-P. Stockis (2004): "Bootstrapping nonparametric estimators of the volatility function," *Journal of Econometrics*, 118(1–2), 189–218.

Gao, F., and F. Song (2008): "Estimation Risk in GARCH VaR and ES Estimates," *Econometric Theory*, 24(5), 1404–1424.

Gerber, H. U., and E. S. W. Shiu (1994): "Martingale Approach to Pricing Perpetual American Options," *ASTIN Bulletin: The Journal of the IAA*, 24(2), 195–220.

——— (1996): "Actuarial bridges to dynamic hedging and option pricing," *Insurance: Mathematics and Economics*, 18(3), 183–218.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, 48(5), 1779–1801.

Hall, P., and M. P. Wand (1988): "Minimizing L1 distance in nonparametric density estimation," *Journal of Multivariate Analysis*, 26(1), 59–88.

Hansen, B. E. (2006): "Interval forecasts and parameter uncertainty," *Journal of Econometrics*, 135(1–2), 377–398.

Hudecová, r., M. Hušková, and S. G. Meintanis (2015): "Tests for time series of counts based on the probability-generating function," *Statistics*, 49(2), 316–337.

Jacod, J., and A. Shiryaev (1998): "Local martingales and the fundamental asset pricing theorems in the discrete-time case," *Finance and Stochastics*, 2(3), 259–273.

Kristensen, D., and A. Rahbek (2005): "ASYMPTOTICS OF THE QMLE FOR A CLASS OF ARCH(q) MODELS," *Econometric Theory*, null(05), 946–961.

Kupiec, P. H. (1995): "Techniques for Verifying the Accuracy of Risk Measurement Models," *The Journal of Derivatives*, 3(2), 73–84.

Liboschik, T., K. Fokianos, and R. Fried (2017): "tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models," *Journal of Statistical Software*, 82(1), 1–51.

Liboschik, T., P. Kerschke, K. Fokianos, and R. Fried (2016): "Modelling interventions in INGARCH processes," *International Journal of Computer Mathematics*, 93(4), 640–657.

Mikosch, T., and C. Stărică (2000): "Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process," *The Annals of Statistics*, 28(5), 1427–1451.

Neumann, M. H. (2011): "Absolute regularity and ergodicity of Poisson count processes," *Bernoulli*, 17(4), 1268–1284.

Pagan, A., and A. Ullah (1999): *Nonparametric Econometrics*. Cambridge University

Press.

Pascual, L., J. Romo, and E. Ruiz (2006): "Bootstrap prediction for returns and volatilities in GARCH models," *Computational Statistics & Data Analysis*, 50(9), 2293–2312.

Phillips, P. C. B., and J. Yu (2005): "Jackknifing Bond Option Prices," *The Review of Financial Studies*, 18(2), 707–742.

Reeves, J. J. (2005): "Bootstrap prediction intervals for ARCH models," *International Journal of Forecasting*, 21(2), 237–248.

Resnick, S. I. (2001): *A probability path*. Birkhäuser, Boston, 2. pr. edn.

Robio, P. O. (1999): "Forecast intervals in ARCH models: bootstrap versus parametric methods," *Applied Economics Letters*, 6(5), 323–327.

Schoutens, W. (2003): *Lévy Processes in Finance: Pricing Financial Derivatives. Wiley Series in Probability and Statistics*. Wiley.

Shimizu, K. (2010): "Bootstrap Does not Always Work," in *Bootstrapping Stationary ARMA-GARCH Models*, ed. by K. Shimizu, pp. 9–17. Vieweg+Teubner, Wiesbaden.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC, 1 edition edn.

Silverman, B. W., and G. A. Young (1987): "The Bootstrap: To Smooth or Not to Smooth?," *Biometrika*, 74(3), 469–479.

Siu, T. K., T. Howell, and H. Yang (2004): "On Pricing Derivatives Under Garch Models: A Dynamic Gerber-Shiu Approach," *North American Actuarial Journal; Schaumburg*, 8(3), 17–31.

Tay, A. S., and K. F. Wallis (2000): "Density forecasting: a survey," *Journal of Forecasting*, 19(4), 235–254.

Thombs, L. A., and W. R. Schucany (1990): "Bootstrap Prediction Intervals for Autoregression," *Journal of the American Statistical Association*, 85(410), 486–492.

Trucíos, C., and L. K. Hotta (2016): "Bootstrap prediction in univariate volatility models with leverage effect," *Mathematics and Computers in Simulation*, 120, 91–103.

Vorbrink, J. (2014): "Financial markets with volatility uncertainty," *Journal of Mathematical Economics*, 53, 64–78.

White, H. (2001): *Asymptotic theory for econometricians*. Academic Press, San Diego, rev. ed edn.

Zhu, K., and S. Ling (2015): "Model-based pricing for financial derivatives," *Journal of Econometrics*, 187(2), 447–457.