Volatility modelling in time and space

Sondre Hølleland

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2020

UNIVERSITY OF BERGEN

Volatility modelling in time and space

Sondre Hølleland

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 06.11.2020

Year: 2020

Title: Volatility modelling in time and space

Name: Sondre Hølleland

Print: Skipnes Kommunikasjon / University of Bergen

Preface

It all started back in 2014, when I was enrolled in the master's program of statistics at the University of Bergen. I was appointed Hans Arnfinn Karlsen as my main supervisor and Bård Støve as co-supervisor. My master's thesis (Hølleland, 2016) was about a spatio-temporal extension of GARCH models. One month after presenting my work, I started my Ph.D. position with Karlsen as my supervisor, continuing in the disciplinary intersection of space, time and GARCH. Now, four years later, I am submitting my Ph.D. thesis with both great joy and gratitude.

The present work has been produced during my four years of employment as Ph.D. student at the Department of Mathematics, University of Bergen – from July 2016 to July 2020.

The thesis conventionally consist of two parts. The first part introduces univariate GARCH models with fundamental theory of stationarity and asymptotic properties of the quasi maximum likelihood estimator. We also present an empirical example modelling the S&P500 index series. Relating to Paper B, we have a brief discussion on GARCH in temperature modelling. The multivariate CCC-GARCH model is discussed and we give a bivariate empirical example. Spatial (G)ARCH is presented and we explain the boundary problem of spatial statistics. The circular model and half-space models are finally presented, before we introduce the papers and some computer code.

The second part consists of the following four papers:

Paper A Sondre Hølleland and Hans A. Karlsen, *A spatio-temporal GARCH model.* Journal of Time Series Analysis **41**, 2, 2020.

Paper B Sondre Hølleland and Hans A. Karlsen, *Decline in temperature variability on Svalbard.* Revision submitted to Journal of Climate, 2020.

Paper C Sondre Hølleland and Hans A. Karlsen, *Space-Time ARMA-GARCH models with applications.*

Paper D Hans A. Karlsen and Sondre Hølleland, *Spatial GARCH processes.*

Acknowledgements

I would like to thank you, my supervisor, co-author and mentor, associate professor Dr. philos Hans Arnfinn Karlsen, for your guidance, support, vital help, detailed feedback on manuscripts and profound enthusiasm for my work and ideas. Thanks for hour(s) long talks and discussions on politics, photography and crime TV shows, with statistics in-between it all. You are always available for a question or two, and the answers are detailed and precise. I admire your profound intuition for theoretical statistics and I think I speak for all participants when I say that department lunches would be dull without you popping in, exclaiming a provocative statement to spark a discussion. Besides providing statistical inspiration, you have also cured my ignorance for politics and TV debates.

To outsiders, statisticians may seem like a group a grey mice, but the statistics group at the Department of Mathematics must be the exception. I would like to thank the extended statistics group, including past and present employees, for contributing to a fun and interesting working environment during my time at the department. It has been my pleasure working with you all. Especially thanks to Jan Bulla and my co-supervisors Bård Støve and Yushu Li. I have appreciated your good advice on many aspects of academic life. To Bård, thanks for making my travelling budget add up. You are a magician.

During this period, I have been privileged with the opportunity of going to statistical conferences in Hong Kong, Pisa, Palermo, London and Sola, always with a great group of colleagues. Even though the conferences themselves not always live up to our high expectations, being miserable alone is always worse than together with others. Experiencing the world with good food and wine also helps. The trips to Sola and Palermo were sponsored by the Meltzer Research Fund. I also had the opportunity to go to Sitges in Spain for teaching R to students from fields related to tinnitus research. Thanks to Jan Bulla and Winfried Schlee for inviting me. It was fun and a pleasant experience.

I have taken great joy in joining the already mentioned department lunch breaks. This is a sanctuary for political discussions and relevant news debates and has taught me the importance of being updated on yesterdays news events. I would especially like to mention my personal friends and colleagues at the department, Håvard G. Frøysa and Kristian Gundersen, whom I have enjoyed many great coffee breaks and mountain hikes with. To Håvard, thanks for the roughly nine years we have spent together at UiB.

Thanks to my family and friends for support. Saving the best for last, I am grateful to my wife and life companion, Christi Elin. You have been patient, listening to my complaints and frustrations. You are my rock.

Sondre Hølleland
July 2020

Abstract

This thesis contributes to the scientific community in several aspects. We introduce both spatial- and spatio-temporal extensions to the family of GARCH and ARMA-GARCH models and present asymptotic statistics for the quasi maximum likelihood estimators [QMLE] for the GARCH extensions. An important property of these extensions are their spatial- and spatio-temporal stationarity, which is part of the model specifications. The models all exist on an equidistant d-dimensional grid, be it purely spatial or spatio-temporal. Volatility modelling is important in finance, but we also present applications from other fields of study, e.g. climate, meteorology and even cell biology. In stationary spatial statistics on infinite lattices, a boundary problem arises. This is dealt with, in two of the papers, by assuming a circular model. This means wrapping the spatial part of the grid of observation onto a torus surface by connecting opposing edges, and effectively removing the boundaries so that each site's neighbours are observed. The torus space is good for visualization and the point is that we regard sites on opposite sides of the rectangle we observe as neighbours. Circulation changes the area of observation from infinite to being closed and finite, and proving asymptotic results becomes easier. Consistency and asymptotic normality of the QMLE is established in the circular situation for GARCH models. The circular model can be used as an approximation of an infinite grid model, in which the circular estimator will be biased. In this setting, we suggest a parametric bootstrap bias correction to compensate for the false links between boundary sites due to circulation. In simulation studies, this approach provides good results for both GARCH and ARMA-GARCH models. For ARMA-GARCH, it is not uncommon to fit an ARMA model to data and a GARCH model to its residuals, but simultaneously estimating all parameters is better. We show by a simulation experiment that the variance of the ARMA-part of the QMLE can be reduced by doing this. The second paper of this thesis is an application of non-stationary GARCH modelling in climate research. We investigate how volatility has developed in a daily temperature series at Svalbard Airport over the last 44 years. During this period the temperature there has increased intensively. We model the volatility using a GARCH model with a trend, where the slope depends on the day of the year. Except for the summer, we find a decreasing temperature variability, i.e. a negative trend. The temperature on Svalbard is getting higher and more stable at the same time and we believe this is due to the reduced sea ice extent in the region. Without the circulation, on an infinite grid and in a potentially purely spatial setting, we turn to half-space GARCH models in the final paper. These models use an ordering of the spatial locations, extending non-deterministic time series to space. The MLE used is based on a modified likelihood, and we show that it is consistent and asymptotically Gaussian. Instead of the standard Lyapunov condition for existence of a stationary solution, a generalization of Nelson's criteria is used.

Contents

Part I

Introduction

Chapter 1

Theory review

The common thread of this thesis is GARCH, or *generalized autoregressive conditional heteroskedasticity*, models. Although this full name may seem like a mouthful, it is quite accurate. When Engle (1982) suggested the ARCH model, for which he won the 2003 Nobel Prize in economic sciences together with Clive W.J. Granger, it was an autoregressive model for conditional heteroskedasticity, i.e. time varying conditional volatility. Later, Bollerslev (1986) generalized the models by adding a recursive term, hence the G in GARCH.

The models are mainly applied in finance and econometrics, where modelling volatility of returns on e.g. stocks, currency rates and bonds are relevant. A reason for this is that the models capture many characteristics of financial returns. McNeil et al. (2005, p.117) list the following empirical observations about return series, which are so entrenched in econometric science that they are referred to as stylized facts:

SF1: Return series are not iid although they show little serial correlation.

SF2: Series of absolute or squared returns show profound serial correlation.

SF3: Conditional excepted returns are close to zero.

SF4: Volatility appears to vary over time.

SF5: Return series are leptokurtic or heavy-tailed.

SF6: Extreme returns appear in clusters.

GARCH models are designed to mimic all of the above. A GARCH series is uncorrelated (SF1), while the squared series has a correlation structure similar to that of an ARMA model (SF2). The conditional expectation of the observations is zero (SF3) and the conditional volatility is time dependent (SF4). Since the conditional volatility depends on previous squared returns, an extreme return yesterday will give a high volatility for tomorrow as well, leading to *volatility clusters* (SF6). Volatility clustering is the tendency for extreme returns to be followed by other extreme returns. GARCH models are heavy-tailed (SF5), even when the innovations are Gaussian. Using a Student's t-distribution instead will further strengthen the leptokurtic property.

We start by considering relevant theory for univariate GARCH models.

1.1 Univariate GARCH

Let $\{X_t\}$ denote a time series of uncorrelated variables with expectation zero. Then a GARCH(p, q) model is defined by

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{1}$$

where $\{\sigma_t\}$ is the conditional volatility series, $\{Z_t\}$ the innovations and $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ is the parameter vector, contained in the parameter space Θ. The restrictions on Θ include that all the parameters must be non-negative and $\omega > 0$. Let $\theta_0 \in \Theta$ denote the true parameter vector generating the observations. The innovations are independent and identically distributed (iid) with $\mathbb{E}\, Z_t = 0$ and $\mathbb{E}\, Z_t^2 = 1$. If $\beta_j \equiv 0$, (1) is an ARCH(p) model.

A squared GARCH process has an ARMA representation, which can be useful for studying the fourth order properties of the model. Let $U_t = \sigma_t^2(Z_t^2 - 1)$, then $X_t^2 = \sigma_t^2 Z_t^2 = \sigma_t^2 + \sigma_t^2(Z_t^2 - 1) = \sigma_t^2 + U_t$, and

$$\begin{aligned} X_t^2 &= \omega + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + U_t \\ &= \omega + \sum_{j=1}^{p \vee q} (\alpha_j + \beta_j) X_{t-j}^2 - \sum_{j=1}^{q} \beta_j U_{t-j} + U_t, \end{aligned} \tag{2}$$

which we recognize as an ARMA$(p \vee q, q)$ with a non-zero intercept. Here we have defined $\alpha_j = 0$ and $\beta_j = 0$ for $j > p$ and $j > q$, respectively.

Second order stationarity If we assume that the GARCH process is second order stationary, this implies that $\operatorname{Var} X_t = \mathbb{E}\, X_t^2 = \mathbb{E}\, \sigma_t^2 = \tau^2$ is constant, and we get that

$$\begin{aligned} \tau^2 &= \omega + \sum_{j=1}^{p} \alpha_j \mathbb{E}\, X_{t-j}^2 + \sum_{j=1}^{q} \beta_j \mathbb{E}\, \sigma_{t-j}^2 = \omega + \Big(\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j \Big) \tau^2, \\ \tau^2 &= \frac{\omega}{1 - \sum \alpha_j - \sum \beta_j}. \end{aligned}$$

Clearly this indicates that for $\tau^2 < \infty$, it is necessary that

$$\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1. \tag{3}$$

In fact, (3) is equivalent to second order stationarity.

THEOREM 1. The GARCH(p, q) process defined by (1) is a white noise process if and only if (3) holds.

Proof. See Bollerslev (1986, Theorem 1). □

Nelson and the Lyapunov exponent Nelson (1990) showed that a GARCH$(1,1)$ model is strictly stationary if and only if

$$\mathbb{E}\,\log(\alpha_1 Z_t^2 + \beta_1) < 0. \tag{4}$$

Applying Jensen's inequality and using that $\mathbb{E}\,Z_t^2 = 1$, we get

$$\mathbb{E}\,\log(\alpha_1 Z_t^2 + \beta_1) \le \log(\alpha_1 + \beta_1) < 0,$$

which is equivalent to (3) for $p = q = 1$. That is, (3) $\Rightarrow$ (4). As noted by Straumann and Mikosch (2006), the same condition can be derived using a stochastic recurrence equation [SRE] argument. The SRE for σ_t^2 is given by

$$\sigma_t^2 = (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2 + \omega = A_t \sigma_{t-1}^2 + B_t,$$

where $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$ and $B_t = \omega$. Then, the sequence (A_t, B_t) is iid and it follows from Brandt (1986) that $\mathbb{E}\,\log|A_0| < 0$ and $\mathbb{E}\,\log^+|B_0| < \infty$ guarantee existence and uniqueness of a strictly stationary solution to the SRE $Y_t = A_t Y_{t-1} + B_t$. In a sense, Nelson (1990) reinvented the wheel, although he was probably the first to put this in a GARCH context.

When $p \vee q \ge 2$, stationarity conditions go from a univariate formulation to a multivariate one and thus become harder to establish. A necessary and sufficient condition for existence of a unique, non-anticipative, strictly stationary solution to (1) was shown by Bougerol and Picard (1992a, Theorem 1.3). The condition is that the top Lyapunov exponent of the sequence of matrices $\{\mathbb{Q}_t\}$ is strictly negative at θ_0, where

$$\mathbb{Q}_t(\theta_0) = \begin{bmatrix} \alpha_1 Z_t^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ & & \mathbb{I}_{q-1} & & \mathbf{0} & & \mathbb{0}_{(q-1)\times(p-2)} & & \mathbf{0} \\ Z_t^2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ & & \mathbb{0}_{(p-2)\times(q-1)} & & \mathbf{0} & & \mathbb{I}_{p-2} & & \mathbf{0} \end{bmatrix}_{(p+q-1)\times(p+q-1)}.$$

When $\mathbb{E}\,\log^+\|\mathbb{Q}_0\| < \infty$, the top Lyapunov exponent, $\gamma_{\mathbb{Q}}$, is defined by

$$\gamma_{\mathbb{Q}} = \lim_{n\to\infty} n^{-1} \log\|\mathbb{Q}_n \mathbb{Q}_{n-1} \cdots \mathbb{Q}_1\| \text{ a.s.}$$

For $p = q = 1$, the condition $\gamma_{\mathbb{Q}} < 0$ reduces to (4). That $\gamma_{\mathbb{Q}} < 0$ is sufficient follows immediately from Brandt (1986). Bougerol and Picard (1992b) establish that it is also necessary, though with a rather involved proof.

1.1.1 Quasi maximum likelihood estimation

Parameter estimation of GARCH models is often done by quasi maximum likelihood. This means using a Gaussian distribution for the innovations, irrespective of their true distribution. It is done by defining a likelihood process

$$h_t(\theta) = \omega + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{j=1}^{q} \beta_j h_{t-j},$$

with $h_t(\theta_0) \equiv \sigma_t^2$. Let $n \in \mathbb{N}_+$ denote the sample size. The theoretical quasi log-likelihood is then given by

$$L_n(\theta) = \sum_{t=1}^{n} \ell_t(\theta), \quad \ell_t(\theta) = -\frac{1}{2}\left(\log h_t + \frac{X_t^2}{h_t} \right), \tag{5}$$

where X_t^2 does not depend on θ and is generated under θ_0. Corresponding to h_t and ℓ_t, we define $\widehat{h}_t$ and $\widehat{\ell}_t$ in the same manner, except that these are to be initiated for $t \leq 0$, while h_t assumes an infinite history for $\{X_t\}$. By replacing h_t and ℓ_t in (6), we get $\widehat{L}_n$. The quasi maximum likelihood estimators (QMLEs) are the parameters that maximize $\widehat{L}_n$ and (5), respectively, i.e.

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \widehat{L}_n(\theta) \quad \text{and} \quad \widetilde{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\, L_n(\theta). \tag{6}$$

The empirical likelihood, $\widehat{L}_n$, is an approximation of the theoretical L_n. We denote $L(\theta) = \mathbb{E}_{\theta_0} \ell(\theta)$ as the asymptotic likelihood, where $\ell = \ell_t(\theta)$ for an arbitrary $t \in \mathbb{Z}$. We use the same notation for $h = h_t$ and $\sigma^2 = \sigma_t^2$.

Consistency and asymptotic normality

Under specific regularity conditions, $\widehat{\theta}_n$ is consistent and asymptotic normally distributed. This was first established by Weiss (1986) for ARCH models under fourth order moment conditions on X_t. Lee and Hansen (1994) and Lumsdaine (1996) developed asymptotic statistics for GARCH(1, 1), while Boussama (1998, 2000), Jeantheau (1998), Berkes et al. (2003), Straumann and Mikosch (2003, 2006) and Francq and Zakoïan (2004) cover the general GARCH(p, q). The number of publications in this relatively short time period indicates a strong competitive fight to be the first to rigorously prove precise asymptotic results for GARCH(p, q). Jeantheau (1998) considers consitency for multivariate GARCH, while Straumann and Mikosch (2003, 2006) prove consistency and asymptotic normality for more general conditionally heteroskedastic time series, with GARCH(p, q) as a special case. We present consistency and asymptotic normality results as given by Francq and Zakoïan (2004), which could be seen as an accumulation of knowledge from the others.

Let $\mathcal{A}(z) = \sum_{j=1}^{p} \alpha_j z^j$ and $\mathcal{B}(z) = 1 - \sum_{j=1}^{q} \beta_j z^j$, where $\mathcal{A}(z) = 0$ if $p = 0$ and $\mathcal{B}(z) = 1$ if $q = 0$. The list of assumptions are:

A1: $\theta_0 \in \Theta$ and Θ is compact.

A2: $\gamma_{\mathbb{Q}} < 0$ and $\sum_{j=1}^{q} \beta_j < 1$ on Θ.

A3: Z_t^2 has a non-degenerate distribution with $\mathbb{E}\, Z_t^2 = 1$.

A4: At θ_0, if $q > 0$, $\mathcal{A}(z)$ and $\mathcal{B}(z)$ have no common root, $\mathcal{A}(1) \neq 0$ and $\alpha_p + \beta_q \neq 0$.

A5: $\theta_0 \in \Theta^\circ$, where Θ° is the interior of Θ.

A6: $\kappa_Z = \operatorname{Var} Z_t^2 < \infty$.

THEOREM 2. Under A1-A4, we have strong consistency; $\widehat{\theta}_n \to \theta_0$ almost surely.

Proof. See Francq and Zakoïan (2004) or Straumann and Mikosch (2006). □

For their consistency conditions, Berkes et al. (2003) assume that $\mathbb{E}\, Z_t^{2+\delta} < \infty$ for some $\delta > 0$. The additional δ is assumed here, because the authors use the uniform convergence approach for their consistency proof. We will return to this shortly. As the corresponding consistency result of Straumann and Mikosch (2006) covers a more general situation, their assumptions are less specific to the GARCH(p, q) model.

THEOREM 3. Under A1-A6, $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ is asymptotically distributed as $\mathcal{N}(0, 2^{-1}\kappa_z \mathbb{I}_0^{-1})$, where

$$\mathbb{I}_0 = 2^{-1}\mathbb{E}_{\theta_0} \nabla \log h_t \nabla' \log h_t.$$

Proof. See Francq and Zakoïan (2004) or Straumann and Mikosch (2006). □

The assumption A5 is clearly necessary, since the asymptotic distribution cannot be Gaussian with expectation at a boundary point. Compared to A6, Berkes et al. (2003) assume that $\mathbb{E}\,|Z_t|^{4+\epsilon} < \infty$ for some $\epsilon > 0$, while Boussama (2000) assume $\mathbb{E}\, Z_t^6 < \infty$. Straumann and Mikosch (2006) avoid this epsilon by using SRE theory.

Proving consistency For proving strong consistency of the QMLE in a GARCH setting there is a recipe. You first prove that $\widetilde{\theta}_n$ is strongly consistent and then you show that $\widehat{\theta}_n$ also must be. The last connection is established by showing that $n^{-1}\|\widehat{L}_n - L_n\|_\Theta \to 0$ a.s. where $\|\cdot\|_\Theta = \sup_{\theta\in\Theta} |\cdot|$ denotes the sup norm. In addition, you need to show that the maximum is identifiable and unique, meaning that $L(\theta) < L(\theta_0)$ for all $\theta \in \Theta\backslash\{\theta_0\}$, with equality if and only if $\theta \equiv \theta_0$.

There are two main strategies of consitency proofs used in the GARCH literature – with and without uniform convergence of $n^{-1}L_n$. This is illustrated by Straumann and Mikosch (2006), who consider both approaches. With uniform convergence, consistency is established by applying a version of the uniform strong law of large numbers (Rao, 1962). It requires the finiteness of $\mathbb{E}\,\|\ell\|_\Theta < \infty$, which Straumann and Mikosch (2006) solve by assuming that $\mathbb{E}\,\|\sigma^2/h\|_\Theta < \infty$ and $\mathbb{E}\,\|\log h\|_\Theta < \infty$. Finite variance is sufficient for the former condition.

The other approach uses a result established by Pfanzagl (1969, Lemma 3.11) and avoid this moment condition. The basic idea dates back to Wald (1949) and the lemma of Pfanzagl is essentially the same as Lemma 5 of LeCam (1953), although LeCam does not address some of the relevant measurability questions. The lemma shows that under weak conditions, $n^{-1}L_n \to L$ a.s., L is upper semi-continuous and that

$$\overline{\lim}\, \|n^{-1}L_n\|_{\Theta'} \leq \|L\|_{\Theta'} \tag{7}$$

holds for any compact subset $\Theta' \subseteq \Theta$. With (7) established and since an upper semicontinuous function attains its maximum on compact sets, one can use standard arguments as presented by e.g. Ferguson (1996, pp. 114-115) for arguing that $\widetilde{\theta}_n$ is consistent.

The two approaches divide the literature of consitency proofs in the GARCH setting with Lee and Hansen (1994), Lumsdaine (1996) and Berkes et al. (2003), who use the first approach, while Jeantheau (1998), Straumann and Mikosch (2003, 2006) and Francq and Zakoïan (2004) rely directly on Pfanzagl (1969). In this context, Francq and Zakoïan (2004) construct their own version of that lemma. This is somewhat surprising, since the approach for consistency based on Pfanzagl (1969) was established prior to their publication.

Proving CLT Asymptotic normality of the QMLE typically rest on a Taylor expansion of the score function at θ_0. Since $\widetilde{\theta}_n$ satisfies the likelihood equations $\nabla L_n(\widetilde{\theta}_n) = 0$, and by dividing by $n^{1/2}$, we get that

$$(\overline{\mathbb{J}}_n + \overline{\mathbb{R}}_n)\, n^{1/2}(\widetilde{\theta}_n - \theta_0) = n^{-1/2}\nabla L_n(\theta_0), \quad \overline{\mathbb{J}}_n = n^{-1}\mathbb{J}_n, \quad \overline{\mathbb{R}}_n = n^{-1}\mathbb{R}_n,$$

where $\mathbb{J}_n = -\nabla^2 L_n(\theta_0)$ and $\mathbb{R}_n$ is the remainder term from the Taylor expansion. Proving asymptotic normality of $\widehat{\theta}_n$ then follows by showing that;

i) The remainder $\overline{\mathbb{R}}_n$ can be neglected.

ii) The observed information matrix, $\overline{\mathbb{J}}_n$ converges to a positive definite matrix; $\mathbb{I}_0$.

iii) The observable estimator $\widehat{\theta}_n$ and its theoretical counterpart $\widetilde{\theta}_n$ are square root n-equivalent.

iv) The score function, $\nabla L_n(\theta_0)$, is asymptotically normal.

As often is the case; *the devil is in the details*, but this list, or variations of it, cover the main ingredients. An obstacle for (i) is to verify the existence of higher order moments of the score function.

Let us consider an empirical example illustrating GARCH modelling.

Example 1. The classical application for univariate GARCH models is stock- or financial index data. We consider the Standard & Poor's 500 stock index, acquired from Yahoo Finance using the R package *quantmod* (Ryan and Ulrich, 2020), from January 1st 2005 to April 30th 2020 giving 3 857 observations, presented in Figure 1. This period includes the financial crisis of 2008 and the SARS-CoV-2 (Covid-19) pandemic of 2020, for which governments around the world shut down large parts of their activities to stop the spread. There are also some events which seem to correspond well with the US presidential elections. We estimate a GARCH$(1,1)$ model to the return data. Let V_t denote the value of the index at time t. The return series is then calculated as

$$X_t = \frac{V_t - V_{t-1}}{V_{t-1}} = \frac{V_t}{V_{t-1}} - 1, \tag{8}$$

and modelled by

$$X_t = \sigma_t\, Z_t, \quad \sigma_t^2 = \omega + \alpha\, X_{t-1}^2 + \beta\, \sigma_{t-1}^2.$$

We use the QMLE given by (6) to estimate the parameters and the results are presented in Table 1. We see that $\alpha+\beta = 0.978 < 1$, indicating weak stationarity and the variance, τ^2, is estimated to $\widehat{\omega}/(1-\widehat{\alpha}-\widehat{\beta}) = 1.17 \cdot 10^{-4}$, which is a bit lower than the empirical variance of $1.55 \cdot 10^{-4}$.

In Figure 2A we have plotted the return series with $\{\pm 1.96\, h_t^{1/2}(\widehat{\theta}_n)\}$. This seems to cover the data quite well and you may especially notice how it captures the increased volatility in the volatile periods, e.g. leading up to the 2008 financial crisis. In Figure 2B and C we consider the standardized residuals, $\widehat{Z}_t = X_t\, h_t^{-1/2}(\widehat{\theta}_n)$. The sample autocorrelation function indicates only smaller deviations from insignificant serial correlation (B), while the qq-normality plot (C) especially indicates a heavier lower tail than the Gaussian. The asymmetry may be an indication of the tendency of losses having heavier tails than gains. The economic interpretation is that a fall in equity value increases the

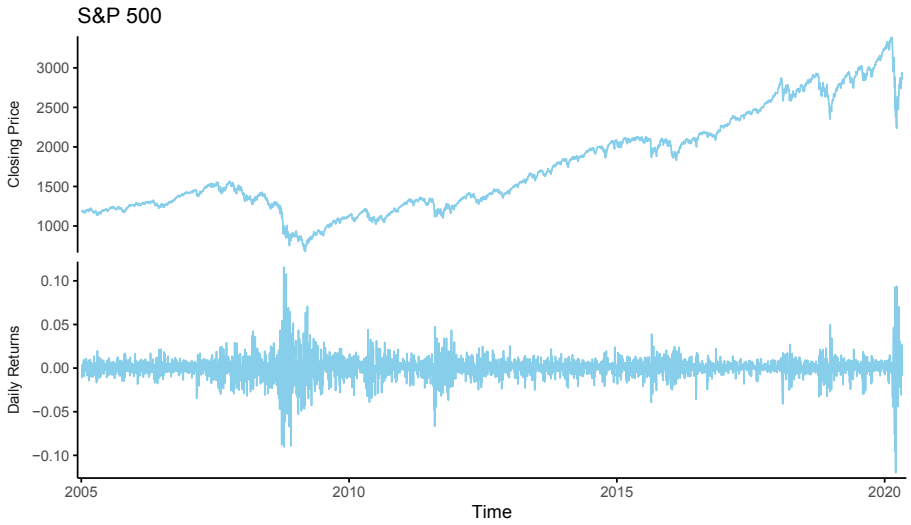

Figure 1: Closing price and daily returns of the Standard & Poor's 500 index.

	Estimate	Standard errors	T-score	P-value
ω	$2.56 \cdot 10^{-6}$	$0.35 \cdot 10^{-6}$	7.34	0.00
α	$1.31 \cdot 10^{-1}$	$0.11 \cdot 10^{-1}$	11.70	0.00
β	$8.47 \cdot 10^{-1}$	$0.12 \cdot 10^{-1}$	71.97	0.00

Table 1: Parameter estimates with standard errors, t-scores and p-values for the S&P 500 return series.

debt-to-equity ratio, i.e. the leverage of a company, and should consequently make the stock more volatile (McNeil et al., 2005, p. 149). This characteristic has therefore been called a *leverage effect*, for which an exponential- (Nelson, 1991) or threshold GARCH (Glosten et al., 1993) could be better suited.

1.1.2 GARCH in temperature modelling

As we have just seen an example of, GARCH models are mainly applied to financial- or econometric time series, but not exclusively. GARCH can also be used in other contexts, often as a residual model for situations where the residuals are heteroskedastic or show signs of volatility clustering.

One application outside finance is temperature modelling. Tol (1996) uses an AR(2)-GARCH(1,1) model on a daily temperature series from De Bilt, The Netherlands. He concludes that the heteroskedastic models outperform their homoskedastic versions. Franses et al. (2001) model weekly mean temperatures, also with an AR-GARCH and in The Netherlands, including seasonal components. Their model allows for asymmetry in the conditional volatility, which they conclude is present in the data. Taylor and Buizza (2004) compare the predicting performance of these models against ensemble prediction methods, but conclude that the ensemble approach is preferred. In preprocessing temperature series for studying their tail distribution, Dupuis (2012, 2014) uses

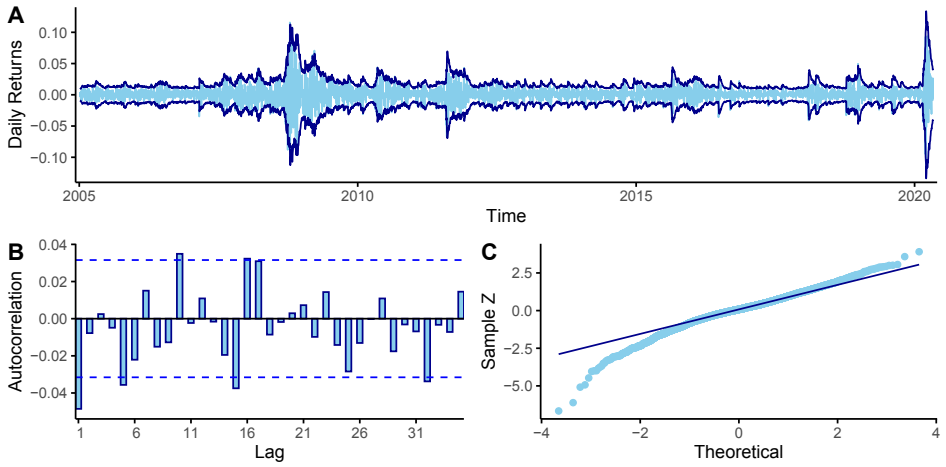

Figure 2: Model evaluation plots for the S&P 500 return series. See text for details.

a deterministic seasonal- and an exponential GARCH model, respectively.

In the intersection between finance and meteorology, we find weather derivatives. To set a price on a weather derivate, you need a model for the weather. Campbell and Diebold (2005) suggest a temperature model for predicting the temperature in four major US cities; Atlanta, Chicago, Las Vegas and Philadelphia. Their model consists of an autoregressive term, a trend and a seasonal part. The residuals are modelled as a GARCH with seasonal intercept using sine and cosine functions. In Paper B, we take a similar approach to model development in temperature variability on Svalbard, although our volatility model includes a trend that depends on the day of the year.

1.2 CCC-GARCH

For multivariate GARCH, the applications have traditionally been in the same areas as univariate; finance and econometrics. In the GARCH literature there exits a large number of alternative formulations with their very own acronym. In fact, Bollerslev (2008) wrote a 42 pages long glossary for ARCH and GARCH acronyms with explanations, many of which are multivariate models.

We will focus our attention on the subclass of multivariate GARCH [MGARCH] models that is most relevant for this thesis; the constant correlation coefficient [CCC] model. This model was introduced by Bollerslev (1990) and extended by Jeantheau (1998) to the form we use today. Let

$$\mathbf{X}_t = \begin{bmatrix} X_{t,1} \\ X_{t,2} \\ \vdots \\ X_{t,m} \end{bmatrix}, \quad \boldsymbol{\sigma}_t = \begin{bmatrix} \sigma_{t,1} \\ \sigma_{t,2} \\ \vdots \\ \sigma_{t,m} \end{bmatrix}, \quad \mathbf{Z}_t = \begin{bmatrix} Z_{t,1} \\ Z_{t,2} \\ \vdots \\ Z_{t,m} \end{bmatrix} \quad \text{and} \quad \mathbb{D}_t = \begin{bmatrix} \sigma_{t,1} & 0 & \dots & 0 \\ 0 & \sigma_{t,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{t,m} \end{bmatrix},$$

with $m \in \mathbb{N}_+$. When we write a squared vector, we mean this in a Hadamard way, i.e. for $\mathbf{a} = (a_1, a_2, \dots, a_m) \in \mathbb{R}^m$, let $\mathbf{a}^2 = \mathbf{a} \circ \mathbf{a} = (a_1^2, a_2^2, \dots, a_m^2)$. A CCC-GARCH$(p, q)$

process, $\{\mathbf{X}_t \in \mathbb{R}^m\}$, is defined by

$$\begin{aligned} \mathbf{X}_t &= \mathbb{H}_t^{1/2}\mathbf{Z}_t, \qquad \mathbb{H}_t = \mathbb{D}_t\,\mathbb{R}\,\mathbb{D}_t, \\ \boldsymbol{\sigma}_t^2 &= \boldsymbol{\omega} + \sum_{j=1}^{p} \mathbb{A}_j \mathbf{X}_{t-j}^2 + \sum_{j=1}^{q} \mathbb{B}_j \boldsymbol{\sigma}_{t-1}^2, \end{aligned} \tag{9}$$

where $\mathbb{R}$ is a correlation matrix, $\boldsymbol{\omega}$ a $m \times 1$ vector with positive coefficients, $\mathbb{A}_j$ and $\mathbb{B}_j$ are $m \times m$ matrices with non-negative coefficients, $\{\mathbf{Z}_t \in \mathbb{R}^m\}$ is a sequence of iid standardized variables. We can write $\mathbf{X}_t = \mathbb{D}_t\mathbf{W}_t$, where $\mathbf{W}_t = \mathbb{R}^{1/2}\mathbf{Z}_t$ is a centred vector with covariance matrix $\mathbb{R}$. The components of $\mathbf{X}_t$ will then have the usual form $X_{t,j} = \sigma_{t,j}W_{t,j}$, but the conditional variance $\sigma_{t,j}^2$ depends on the past of all the components of $\mathbf{X}_t^2$.

By requiring that $\mathbb{A}_j$ and $\mathbb{B}_j$ consist of nonnegative coefficients and that $\mathbb{R}$ is positive definite, $\mathbb{H}_t$ is positive definite. However, less restrictive conditions can be found (Conrad and Karanasos, 2010). Strict stationarity conditions based on a Lyapunov condition generalize directly from the univariate case (Bougerol and Picard, 1992a), which cannot be said for general MGARCH models. There also exist representations of (9) where the matrices $\mathbb{B}_j$ are diagonal. Limitations of these models are non-stability by aggregation and the arbitrary nature of assuming constant conditional correlations (Francq and Zakoian, 2019, pp. 279-280), which is often violated in applications. For instance, when the entire market crashes, correlations are typically higher than in a normal state. Thus, a CCC-GARCH has too little flexibility in the conditional volatility.

Jeantheau (1998) gave general conditions for strong consistency of the QMLE for MGARCH, while Ling and McAleer (2003) established consistency and asymptotic normality of vector ARMA-GARCH for the CCC formulation. For the pure CCC-GARCH case, see Francq and Zakoian (2019, Thm 10.8 and 10.9). Aue et al. (2009) established a sufficient condition for strict stationarity and the existence of fourth-order moments of the process (9).

We illustrate usage of CCC-GARCH by an empirical example.

Example 2. We model two series of stock returns simultaneously using a CCC-GARCH model. The companies we consider are Amazon.com Inc. and Microsoft Corporation and the data is acquired from Yahoo Finance using the R package *quantmod* (Ryan and Ulrich, 2020). The returns are calculated by (8) and mean subtracted prior to modelling. With a bivariate model the correlation matrix $\mathbb{R}$ is reduced to one parameter ρ to be estimated. The model can be formulated as

$$\boldsymbol{\sigma}_t^2 = \boldsymbol{\omega} + \mathbb{A}\mathbf{X}_{t-1}^2 + \mathbb{B}\boldsymbol{\sigma}_{t-1}^2, \quad \mathbb{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbb{B} = \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}, \quad \mathbb{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

with $\boldsymbol{\omega} = (\omega_1, \omega_2)'$. We have assumed $\mathbb{B}$ diagonal to simplify the model. In Figure 3, we have plotted the two series with the fitted volatility as confidence bands. Parameter estimation is done by a QMLE similar to (6). We start out with the parameter vector $\theta = (\rho, \omega_1, \omega_2, a_{11}, a_{12}, a_{21}, a_{22}, b_{11}, b_{22})'$ and gradually remove insignificant parameters with a significance level of 5%. We use AIC for model selection and a sandwich estimator to estimate the covariance matrix of the estimator. With AIC, the initial model is the

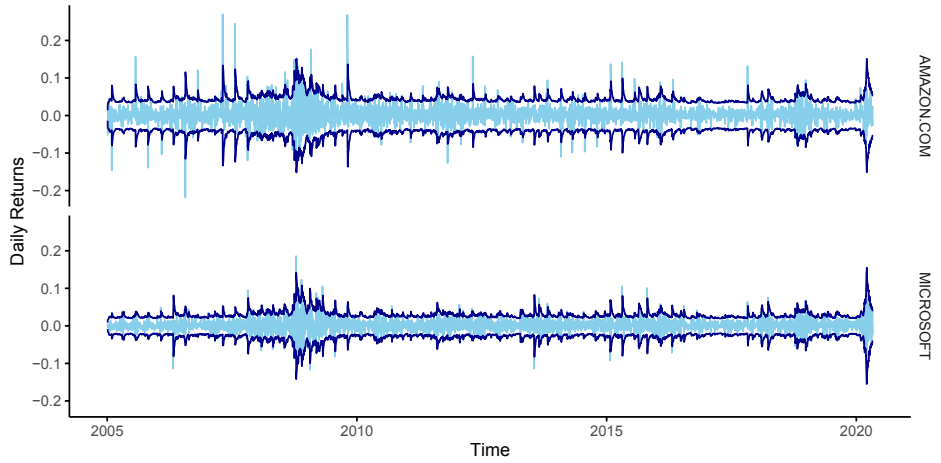

Figure 3: Return series of Amazon.com Inc and Microsoft Corporation from 2005 to April 2020 with $\pm 1.96\,\widehat{\sigma}_{t,j}$ from the fitted CCC-GARCH model.

	Estimates	Standard errors	T-score	P-value
ρ	$4.43 \cdot 10^{-1}$	$2.34 \cdot 10^{-2}$	18.91	0.00
ω_1	$5.03 \cdot 10^{-5}$	$1.54 \cdot 10^{-5}$	3.27	0.00
ω_2	$1.93 \cdot 10^{-5}$	$6.03 \cdot 10^{-6}$	3.20	0.00
a_{11}	$5.93 \cdot 10^{-2}$	$3.83 \cdot 10^{-2}$	1.55	0.12
a_{12}	$9.34 \cdot 10^{-2}$	$5.41 \cdot 10^{-2}$	1.73	0.08
a_{21}	$8.15 \cdot 10^{-3}$	$7.54 \cdot 10^{-3}$	1.08	0.28
a_{22}	$1.23 \cdot 10^{-1}$	$2.99 \cdot 10^{-2}$	4.10	0.00
b_{11}	$8.15 \cdot 10^{-1}$	$3.37 \cdot 10^{-2}$	24.17	0.00
b_{22}	$7.90 \cdot 10^{-1}$	$4.61 \cdot 10^{-2}$	17.14	0.00

Table 2: Estimation results for Example 2.

preferred one. The resulting parameter estimates with t-scores and p-values are reported in Table 2. If we were to use the inverted Hessian as estimate for the covariance matrix, all estimates would be significant at the 5% level, but using the more robust sandwich estimator gives p-values above 5% for a_{11}, a_{12} and a_{21}. Removing any of these will however increase the AIC. The sandwich estimator does not require knowledge of the residual distribution. Note that the constant correlation coefficient, ρ, is estimated to 0.443, between Microsoft and Amazon returns.

These multivariate models are well suited for applications such as the one in Example 2, but if we were to use them directly in a spatio-temporal setting with many spatial locations, the number of parameters to be estimated would be large. This can be solved by assuming $\mathbb{A}_j$ and $\mathbb{B}_j$ to be sparse matrices, where only neighbouring locations influence each other. This will work in the spatio-temporal setting and is what we do in Paper A and Paper C. In the purely spatial setting, the link to CCC-GARCH is no longer relevant.

1.3 GARCH in space

The title of this thesis includes both time and space. Up until now we have only considered volatility modelling in time, but spatial volatility modelling is also possible.

Let $X = \{X(s),\, s \in \mathcal{R}\}$ be a spatial process defined on a discrete spatial network $\mathcal{R}$ that have a neighbourhood geometry. We classify these processes in two bins, when the spatial area of observation, $\mathcal{R}$, is irregular or regular. The irregular cases are often found in domains like econometrics, environmental studies, epidemiology or geography, with sites $s \in \mathcal{R}$ representing centroids of geographical units, such as municipals, counties or countries. The variable of interest at site s is $X(s)$. In the other class, where $\mathcal{R}$ is regular, we find domains such as imaging, radiography and remote sensing. Here $\mathcal{R}$ will typically be a subset of $\mathbb{Z}^d$ and this enables the process to be stationary. For $d = 1$, these are the closest to time series. The fundamental difference is the lack of a natural ordering of $\mathcal{R}$ that is inherent in time series.

Otto et al. (2018, 2019) introduce a spatial ARCH model formulated as

$$\mathbf{X} = \boldsymbol{\sigma} \circ \mathbf{Z}, \quad \boldsymbol{\sigma}^2 = \boldsymbol{\omega} + \mathbb{W}\,\mathbf{X}^2, \tag{10}$$

with $\mathbf{X} = (X(s_1), \ldots, X(s_m))$ and correspondingly for $\boldsymbol{\sigma}$, $\mathbf{Z}$ and $\boldsymbol{\omega} \geq 0$. The m sites $\{s_j \in \mathcal{R}\}$ denote locations in a finite d dimensional space and the $m \times m$ spatial weight matrix $\mathbb{W}$, is assumed to have zeros on the diagonal, be non-stochastic and nonnegative. The spatial weight matrix is further parametrized in some way, e.g. $\mathbb{W} = \alpha\widetilde{\mathbb{W}}$ or $\mathbb{W} = \text{diag}(\alpha_1, \ldots, \alpha_m)\widetilde{\mathbb{W}}$, where $\widetilde{\mathbb{W}}$ is a known spatial weight matrix and α and $\{\alpha_j\}$ are parameters. For existence of a unique solution of (10), the support of the squared innovations has to be bounded, i.e. $Z^2(s_j) \leq a$ for all $j = 1, \ldots, m$, where a is a positive real number (Otto et al., 2018, Theorem 1). This is a strong restriction.

The spatial weight matrices are often used as a convenient notation for vectorized spatial models. If the observations are on a regular grid $\widetilde{\mathbb{W}}$ can be defined by contiguity matrices, such as the queen- or rook contiguity matrices. These are constructed based on the movement patterns of the respective chess pieces as illustrated in Figure 4 and are used in Paper A and Paper C. If $\mathcal{R}$ is irregular, $\widetilde{\mathbb{W}}$ can be a k-nearest neighbours matrix, an inverse distance matrix or a contiguity matrix in the Hotelling sense.

If you use a contiguity neighbourhood on $\mathcal{R} \subset \mathbb{Z}^d$, such as the rook or queen in Figure 4, and you desire for a stationary model, this implies that the dependence structure should be the same no matter which site you are considering. That is; the neighbourhood structure is translation invariant. If you are considering a point at the boundary, say h1 on the chessboard in Figure 4 with a first order rook contiguity neighbourhood. The process at h1 depends on g1 and h2, but also i1 and h0, which both are outside the chessboard. This is *the boundary problem* of spatial statistics.

1.3.1 The boundary problem and circulation

The boundary problem in spatial statistics arise in stationary spatial- and spatio-temporal processes with a translation invariant neighbourhood structure. In the center of the spatial region you will most likely have observations for all dependent variables, but as you move your focus closer to the boundary, dependent neighbours will eventually be outside your area of observation. For $d \geq 2$ this is a potentially big problem, as the edge effect will be the dominating source of bias (Guyon, 1982).

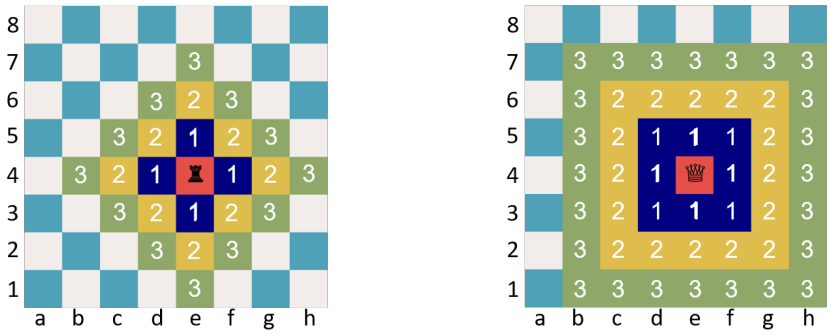

Figure 4: Rook- and queen contiguity neighbourhoods up to order three on a chessboard. The numbers and colours indicate the order of neighbourhood, defined by the number of moves necessary by the respective chess pieces if you move one tile at the time.

In Paper A we introduce the terms circular, circulation and circulating. By this we mean treating a spatial rectangle as if it was a torus when it comes to neighbourhood relations. This is a common approach in a purely spatial context and is sometimes referred to as *wrapping the space onto a torus*. The benefit of having a toroidal surface is the lack of boundary effects – simply because a torus surface has no boundary. This can be seen in Figure 5, where we have circulated the rook and queen contiguities from Figure 4. This does not imply any changes to the observations, but in e.g. (10), we will use a circular spatial weight matrix, $\widetilde{\mathbb{W}}$, that connects opposing edges in the spatial area of observation. The weight matrix is particularly useful for the torus model, since the circulation method can be implemented by a careful specification of this matrix. For $d = 1$, circulating implies connecting the edges of a line – forming a circle.

In the spatial statistics literature the circular assumption is mentioned by Cressie (2015, p. 438) for dealing with the boundary problem on infinite lattices. Moran (1973a) considered a stationary Gaussian process with first-order neighbourhood structure on a square torus lattice and let the size of the torus tend to infinity. He later extended to certain non-Gaussian processes (Moran, 1973b). Others who have assumed a circular model are e.g. Rue and Tjelmeland (2002), Allcroft and Glasbey (2003) and Thon et al. (2012) for spatial processes and Glasbey and Allcroft (2008) for spatio-temporal. Circulation is also used in convolutional neural networks (Goodfellow et al., 2016; Maron

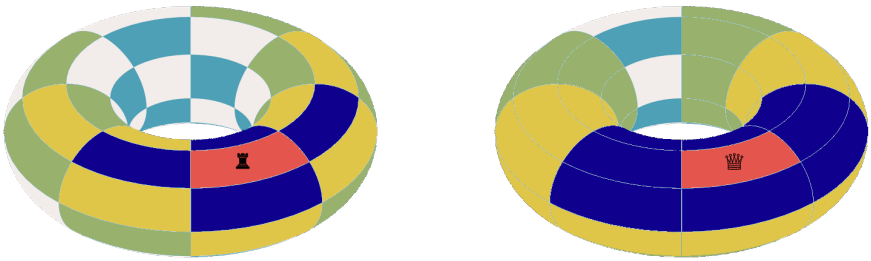

Figure 5: Circular rook- and queen contiguity neighbourhoods, corresponding to Figure 4, illustrated on the 8×8 torus surface.

et al., 2017). The *spdep* R package (Bivand and Wong, 2018) contains functions for making spatial weight matrices with a torus option for circulation.

As alternatives to circulation, Cressie (2015, p. 422) also suggest integrating out the unobserved data from the conditioning event or constructing a *guard area* inside the perimeter of the area of interest, where observations contribute to the likelihood only through their neighbourhood relations with internal locations. Griffith (1983) and Griffith and Amrhein (1983) compare different strategies for dealing with the boundary problem and in short conclude that no method is all that satisfactory. A version of the guard area approach is used in Paper D, where we introduce half-space GARCH models.

1.3.2 Half-space models

Half-space models oppose the lack of ordering in spatial processes. By defining an ordering of the spatial locations, one can define a causal relationship much like the one we know from time series. In a spatial situation it is rare that a process only is influenced by other variables from a certain direction, but you can imagine for instance a process on a river where the stream floats in a certain direction. Observations up-stream influences those further down the stream, but not the other way around. However, this should not be the selling-point of half-space models. These models should probably not be used in settings where the interpretation of the parameters is a vital part of the modelling. If the goal is a parametrization of the correlation structure on a grid, the half-space method can be an efficient way to achieve this. The autocorrelation function itself does not recognize its half-space heritage. A characteristic of half-space models is that the spatial area is infinite, i.e. $\mathcal{R} = \mathbb{Z}^d$.

Half-space representations were introduced by Whittle (1954), who considered the two dimensional case. Helson and Lowdenslager (1958, 1961) developed a mathematical theory for half-space representations and Tjøstheim (1983) established asymptotic inference for half-space autoregressive models. A half-space $\mathcal{S}_+ \subset \mathbb{Z}^d$, is defined by $\mathbb{Z}^d = \mathcal{S}_+ \cup \mathcal{S}_- \cup \{0\}$ being a partition, $\mathcal{S}_- = -\mathcal{S}_+$ and $\mathcal{S}_+$ being closed under addition. A half-space structure defines a total ordering giving a past, present and future at each site. For $s, t \in \mathbb{Z}^d$, we write $s \prec t$ if $t - s \in \mathcal{S}_+$ and $s \preceq t$ including the case $s = t$. A common choice for $\mathcal{S}_+$ is the lexicographical half-space. If $d = 2$, $s = (s_1, s_2) \in \mathbb{Z}^2$ and $t = (t_1, t_2) \in \mathbb{Z}^2$, we have that $s \prec t$ if $s_1 < t_1$ or if $s_1 = t_1$ and $s_2 < t_2$. The lexicographical half-space is visualized in Figure 6 for $d = 2$. An autoregressive half-space spatial model at location t only depends on variables at locations s for which $s \prec t$, i.e. observations in the past. This is the natural generalization of time series to spatial models and, for $d = 1$, a half-space model is a univariate time series. For $d \geq 2$, one can define the dominant index as time and get a spatio-temporal series. Half-space models belong to the class of unilateral models and causal quadrant models are a subclass of half-space models for which dependent variables are located at points where $s_j \leq t_j$, $j = 1, \ldots, d$, except for $s \equiv t$. The unilaterial processes have been studied by Tjøstheim (1978, 1983), Korezlioglu and Loubaton (1986), Basu and Reinsel (1993) and Yao and Brockwell (2006) to mention a few. On a more practicle note, Mojiri et al. (2018) compare the use of unilateral spatial models for predicting unobserved areas to traditional kriging in a simulation study and conclude that these models are able to compete.

Although the half-space model formulation signifies a directional dependence, the following example, inspired by Gaetan and Guyon (2010, Example 1.9-1), illustrates

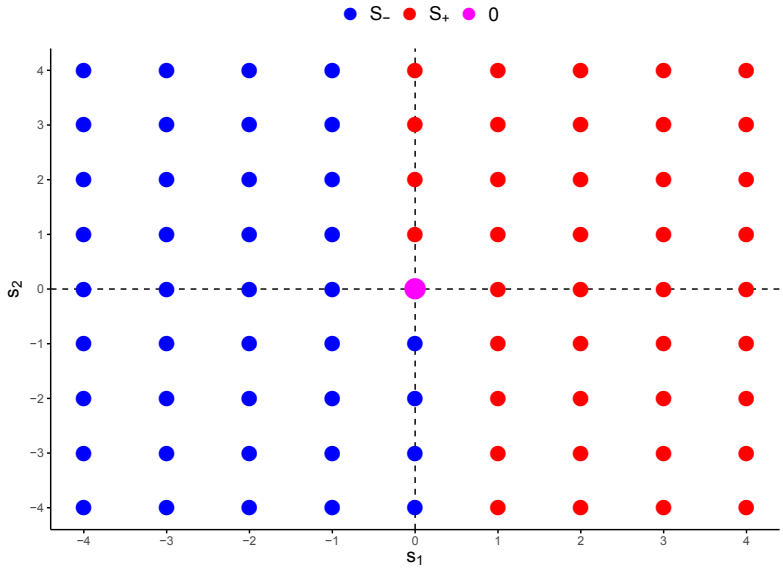

Figure 6: Lexicographic half-spaces $\mathcal{S}_+$ and $\mathcal{S}_- = -(\mathcal{S}_+)$ for $\mathbb{Z}^2$.

that a half-plane ($d = 2$) SAR model can be expressed as a CAR model with more symmetries.

EXAMPLE 3. Let $s = (s_1, s_2) \in \mathcal{R} = \mathbb{Z}^2$ and

$$X(s_1, s_2) = \alpha X(s_1 - 1, s_2) + \beta X(s_1, s_2 - 1) + Z(s_1, s_2), \quad Z(s_1, s_2) \sim \text{iid WN}(0, \sigma_Z^2),$$

be a causal quadrant SAR model. Note that $(s_1 - 1, s_2) \prec (s_1, s_2)$ and $(s_1, s_2 - 1) \prec (s_1, s_2)$, i.e. this is a half-space model under the lexicographical ordering. We will find a symmetric CAR representation of this process. With (s_1, s_2) fixed, we have three equations involving $X(s_1, s_2)$;

$$\begin{aligned} X(s_1, s_2) &= \alpha X(s_1 - 1, s_2) + \beta X(s_1, s_2 - 1) + Z(s_1, s_2), \\ X(s_1 + 1, s_2) &= \alpha X(s_1, s_2) + \beta X(s_1 + 1, s_2 - 1) + Z(s_1 + 1, s_2), \\ X(s_1, s_2 + 1) &= \alpha X(s_1 - 1, s_2 + 1) + \beta X(s_1, s_2) + Z(s_1, s_2 + 1). \end{aligned}$$

Solving this system for $X(s_1, s_2)$, we get, with $\kappa^2 = (1 + \alpha^2 + \beta^2)^{-1}$,

$$\begin{aligned} X(s_1, s_2) = \alpha\kappa^2(X(s_1 - 1, s_2) + X(s_1 + 1, s_2)) + \beta\kappa^2(X(s_1, s_2 - 1) + X(s_1, s_2 + 1)) \\ - \alpha\beta\kappa^2(X(s_1 + 1, s_2 - 1) + X(s_1 - 1, s_2 + 1)) + W(s_1, s_2), \end{aligned}$$

where $\sigma_W^2 = \kappa^2\sigma_Z^2$ and $W(s_1, s_2) = \kappa^2(Z(s_1, s_2) - \alpha Z(s_1 + 1, s_2) - \beta Z(s_1, s_2 + 1))$ is a so-called coloured noise. The two sets of variables are illustrated in Figure 7. The half-plane SAR (green ellipse) has a more symmetric CAR representation (blue ellipse).

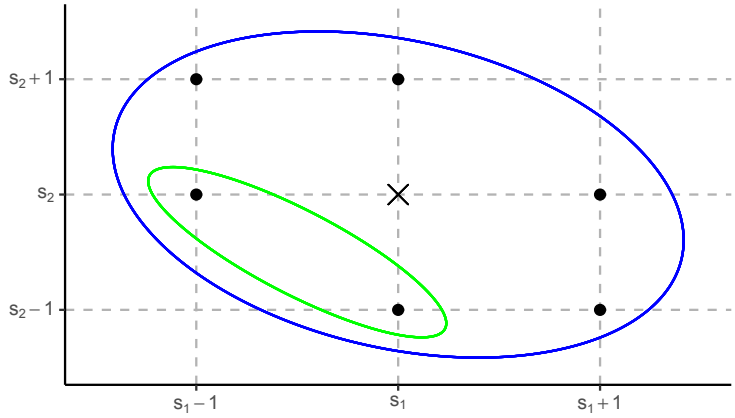

Figure 7: The points in the green ellipse are those included in the SAR representation of Example 3, while the ones in the blue are in the CAR.

Chapter 2

Introduction to the papers

Paper A: *A spatio-temporal GARCH model*

In this paper we introduce a stationary spatio-temporal GARCH model on an infinite gridded spatial surface. The spatio-temporal stationarity is achieved due to a translation invariant neighbourhood system. The model and its neighbourhood system is wrapped around a torus to get a model with finite spatial domain, referred to as a circular model, avoiding the boundary problem of spatial statistics (see Section 1.3.1). The circular process is related to the corresponding non-circular, but their relationship is non-trivial and partly unsolved. The circular estimator can be used as an approximation for non-circular data with a parametric bootstrap bias correction (PBBC). A simulation experiment is presented to illustrate that the PBBC may successfully compensate for the wrongful circulation due to the approximation, when the data is generated non-circularly. However, it is a challenge to establish all theoretical arguments supporting this method.

Corresponding results to those for univariate GARCH given in Section 1.1 are presented for the circular spatio-temporal setting. That is, we generalize the Lyapunov condition for stationarity, and prove consistency and asymptotic normality of the QMLE. The proofs are inspired by Straumann and Mikosch (2006) and Francq and Zakoïan (2004) and follow the recipes in Section 1.1.1. As the paper shows, the CSTGARCH is a special case of CCC-GARCH (Section 1.2), but the stationary context makes this less relevant. We also derive an ARMA representation corresponding to (2) and use it for studying the spatio-temporal correlation structure of the squared processes.

Finally, the paper demonstrates usage of these models on a real data example. Namely, the classical dataset of sea surface temperature anomalies in the Pacific Ocean, used as example by Cressie and Wikle (2011) and others (see references in Paper A). Here a circular spatio-temporal ARMA model is used as preprocessing to obtain an uncorrelated residual series. A CSTGARCH model is then fitted to these residuals.

Paper B: *Decline in variability on Svalbard*

Here we consider an application of GARCH models in a somewhat new setting; climate. For the most part, GARCH models are applied to financial time series, but here we apply them to a temperature time series, namely the Svalbard Airport daily mean temperature series. The data used span over 44 years, from 1976 to 2019. Over this period, the mean temperature has increased extensively on Svalbard, by a rate of 12.1°C per century,

which is much higher than the global mean. This is a well-known phenomenon called Arctic amplification and many have looked into why this happens. The most obvious reason is due to sea ice melting. When the sea ice melts, it is replaced by open sea water, which absorbs more heat from the sun than the ice and snow did. Open sea water also increases heat transfer to the Arctic regions and to the atmosphere there. We believe that this sea ice reduction also has consequences for the variability in temperature on Svalbard.

Usually, extreme temperature anomalies are large deviations from an expected temperature from a reference period. We refer to these as climatic extremes. Opposed to these, are distributional extremes, which are relative to the marginal distribution of the temperature. After preprocessing the data using an ARMA model, we use a GARCH model with seasonal components and a trend, where the slope parameter depends on the day of the year, to model conditional and unconditional temperature volatility on Svalbard. For comparison, we also fit a nonstochastic model. The climate change of volatility is estimated to zero during summer and $-0.1(^\circ\mathrm{C})^2$ per year in February, which has the largest decline. Our main conclusion is that the volatility has been decreasing at Svalbard Airport except for during summer, likely due to reduced sea ice extent. Open, ice free water makes the temperature more stable, transitioning the climate to a more coastal one.

Paper C: *Space-Time ARMA-GARCH models*

We extend the circular spatio-temporal GARCH models from paper A to CSTARMA-GARCH. Spatio-temporal ARMA models have a wider application potential compared to pure GARCH models and this extension provides a family of models with high flexibility. In the real data example of Paper A, we fitted a CSTARMA model as preprocessing before fitting a CSTGARCH to the residuals. Here, we redo the analysis and estimate the two parts simultaneously. Although the two parts are asymptotically uncorrelated, the inclusion of GARCH to an ARMA influences the asymptotic covariance matrix of the ARMA part. We illustrate this by a simulation experiment, where the inclusion of GARCH reduces the asymptotic variance of the ARMA part of the estimator. We also investigate how the parametric bootstrap bias correction from paper A performs on the STARMA-GARCH situation for both one and two dimensional space by simulation experiments. To simulate non-circular data, we use an efficient half-space simulation procedure inspired by Paper D. Finally a real data example from cell biology is considered.

Paper D: *Spatial GARCH processes*

This paper studies a half-space formulation (see Section 1.3.2) of spatial GARCH models, abbreviated as SGARCH. When the spatial region is infinite, the STGARCH introduced in Paper A is no longer a CCC-GARCH type of model. In particular, the Lyapunov condition is apparently not relevant and has to be replaced. In this purely spatial setting, where spatio-temporal is a special case, existence of a solution is guaranteed by a generalization of Nelson's condition (see (4) in Section 1.1). Similar to Paper A and Paper C, the SGARCH model is spatially stationary with a half-space neighbourhood

structure. Statistical inference is based on a modified likelihood and we prove consistency and asymptotic normality of this maximum likelihood estimator. The conditions are pretty close to the time series list of assumptions presented in Section 1.1.1, although a spatial setting is necessarily more complicated. To deal with the boundary problem, a guard area similar to the one mentioned in Section 1.3.1 is used. The paper opens up for interesting applications. In particular, empirical residuals can be tested for heteroskedasticity. We also believe this work quite easily can be extended to SARMA-GARCH. This is an ongoing work. Other extensions are also possible.

Chapter 3

Computer Code

All the code used in this thesis has been written in or for the programming language R and I want to pay tribute to the developers behind some of the great software packages that I have used. I have developed an R package called *starmagarch*, published on github, for simulating and estimating spatio-temporal ARMAGARCH models. This can be used for reproducing results of Paper A and Paper C, although Paper A does not use the ARMA part. The package can be installed using the following lines of code, employing the devtools package (Wickham et al., 2020):

```
library(devtools)
install_github("holleland/starmagarch")
```

The package uses the TMB (template model builder) package developed by Kristensen et al. (2016). TMB enables users to easily formulate their model, i.e. their objective function, in C++ and it calculates the first and second order derivatives of the objective function using automatic differentiation. The objective function and its derivatives can be called by the user in R and estimation of parameters is easily done using built-in optimizers of R, such as the `nlminb`. TMB was made for fitting statistical latent variable models to data, making use of the Laplace approximation, but we only use it as a convenient way of implementing a likelihood estimation routine for R, with C++ implementation for efficiency.

The *spdep* package, developed by Bivand and Wong (2018), is used in starmagarch for generating rook- and queen contiguity matrices. It conveniently has an option for circular matrices, which is useful for our purposes. The current implementation of starmagarch does not exploit the sparsity of these matrices, and can be further improved by accounting for this.

Almost every graphic or figure in this thesis and its containing papers are produced using the *ggplot2* package of Wickham (2016). This is an extremely powerful, useful and flexible R package for creating figures and illustrating data and is hereby recommended to all users of R.

The computer code for the other papers can also be found on github for reproducibility. These are not made as R packages, but as a collection of scripts and functions. For Paper C the mentioned starmagarch package is used, but the specific code examples can be found at

`https://github.com/holleland/Space-time-ARMAGARCH`.

Also the cell datasets are published here, by courtesy of Lee et al. (2015), along with

the spatially differenced sea surface temperature anomalies dataset. For the latter, the original dataset can be found here:

`ftp://ftp.wiley.com/public/sci_tech_med/spatio_temporal_data`.

While working on Paper C, I implemented simulation routines using the Rcpp package (Eddelbuettel et al., 2011). This package allows the user to write C++ functions executable from R. R is known for being a slow software for doing serious computations and by writing subroutines executable from R, one can save time and resources. By simulating more effectively, the computational costs of the simulation experiments in this paper was reduced notably.

For the analysis in Paper B we also use TMB for the likelihood estimation and the R- and C++ code are available at

`https://github.com/holleland/VariabilityOnSvalbard`.

The temperature time series from Svalbard Airport can be downloaded here, but is also available from

`http://eklima.met.no`.

The same holds for the reconstructed monthly series going back to 1898 (Nordli, 2010; Nordli et al., 2014). These are both published with permission of the Norwegian Meteorological Institute under a Creative Commons BY 3.0 licence (see Paper B for details).

References

Allcroft, D. J. and Glasbey, C. A. (2003). A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):487–498.

Aue, A., Hörmann, S., Horváth, L., Reimherr, M., et al. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087.

Basu, S. and Reinsel, G. C. (1993). Properties of the spatial unilateral first-order ARMA model. *Advances in applied Probability*, 25(3):631–648.

Berkes, I., Horváth, L., and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli*, 9(2):201–227.

Bivand, R. and Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST*, 27(3):716–748.

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

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, pages 498–505.

Bollerslev, T. (2008). Glossary to ARCH (GARCH). *CREATES Research Paper*, 49.

Bougerol, P. and Picard, N. (1992a). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics*, 52(1-2):115–127.

Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, 20(4):1714–1730.

Boussama, F. (1998). *Ergodicité, mélange et estimation dans les modeles GARCH*. Université Paris 7. Doctoral thesis.

Boussama, F. (2000). Normalité asymptotique de l'estimateur du pseudo-maximum de vraisemblance d'un modèle GARCH. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 331(1):81–84.

Brandt, A. (1986). The stochastic equation Yn+ 1= AnYn+ Bn with stationary coefficients. *Advances in Applied Probability*, pages 211–220.

Campbell, S. D. and Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 100(469):6–16.

Conrad, C. and Karanasos, M. (2010). Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory*, 26(3):838–862.

Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.

Cressie, N. A. C. (2015). *Statistics for spatial data*. Wiley Classics Library. John Wiley & Sons, Inc., New York, revised edition.

Dupuis, D. J. (2012). Modeling waves of extreme temperature: the changing tails of four cities. *Journal of the American Statistical Association*, 107(497):24–39.

Dupuis, D. J. (2014). A model for nighttime minimum temperatures. *Journal of Climate*, 27(19):7207–7229.

Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., Chambers, J., and Bates, D. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Ferguson, T. S. (1996). *A course in large sample theory*. Texts in Statistical Science Series. Chapman & Hall, London.

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

Francq, C. and Zakoian, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

Franses, P. H., Neele, J., and van Dijk, D. (2001). Modeling asymmetric volatility in weekly Dutch temperature data. *Environmental Modelling & Software*, 16(2):131–137.

Gaetan, C. and Guyon, X. (2010). *Spatial statistics and modeling*, volume 90. Springer.

Glasbey, C. and Allcroft, D. (2008). A spatiotemporal auto-regressive moving average model for solar radiation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(3):343–355.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (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.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Griffith, D. and Amrhein, C. (1983). An evaluation of correction techniques for boundary effects in spatial statistical analysis: traditional methods. *Geographical Analysis*, 15(4):352.

Griffith, D. A. (1983). The boundary value problem in spatial statistical analysis. *Journal of regional science*, 23(3):377–387.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, 69(1):95–105.

Helson, H. and Lowdenslager, D. (1958). Prediction theory and Fourier series in several variables. *Acta mathematica*, 99:165–202.

Helson, H. and Lowdenslager, D. (1961). Prediction theory and Fourier series in several variables. II. *Acta Mathematica*, 106(3-4):175–213.

Hølleland, S. (2016). Spatio-temporal generalized autoregressive conditional heteroskedasticity models. Master's thesis, University of Bergen.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14(1):70–86.

Korezlioglu, H. and Loubaton, P. (1986). Spectral factorization of wide sense stationary processes on Z^2. *Journal of multivariate analysis*, 19(1):24–47.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. (2016). TMB: automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. California Pub. Statist.*, 1:277–330.

Lee, K., Elliott, H. L., Oak, Y., Zee, C.-T., Groisman, A., Tytell, J. D., and Danuser, G. (2015). Functional hierarchy of redundant actin assembly factors revealed by fine-grained registration of intrinsic image fluctuations. *Cell systems*, 1(1):37–50.

Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10(01):29–52.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric theory*, 19(2):280–310.

Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH$(1, 1)$ and covariance stationary GARCH$(1, 1)$ models. *Econometrica*, 64(3):575–596.

Maron, H., Galun, M., Aigerman, N., Trope, M., Dym, N., Yumer, E., Kim, V. G., and Lipman, Y. (2017). Convolutional neural networks on surfaces via seamless toric covers. *ACM Trans. Graph.*, 36(4):71–1.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton Series in Finance. Princeton University Press, Princeton, NJ.

Mojiri, A., Waghei, Y., Sani, H. N., and Borzadaran, G. M. (2018). Comparison of predictions by kriging and spatial autoregressive models. *Communications in Statistics-Simulation and Computation*, 47(6):1785–1795.

Moran, P. A. P. (1973a). A Gaussian Markovian process on a square lattice. *J. Appl. Probability*, 10:54–62.

Moran, P. A. P. (1973b). Necessary conditions for Markovian processes on a lattice. *J. Appl. Probability*, 10:605–612.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory*, 6(3):318–334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.

Nordli, Ø. (2010). The Svalbard Airport temperature series. *Bulletin of Geography. Physical Geography Series*, 3(1):5–25.

Nordli, Ø., Przybylak, R., Ogilvie, A. E., and Isaksen, K. (2014). Long-term temperature trends and variability on Spitsbergen: the extended Svalbard Airport temperature series, 1898–2012. *Polar Research*, 33(1):21349.

Otto, P., Schmid, W., and Garthoff, R. (2018). Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity. *Spatial Statistics*, 26:125–145.

Otto, P., Schmid, W., and Garthoff, R. (2019). Stochastic properties of spatial and spatiotemporal ARCH models. *Statistical Papers*, pages 1–16.

Pfanzagl, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika*, 14(1):249–272.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, pages 659–680.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian journal of Statistics*, 29(1):31–49.

Ryan, J. A. and Ulrich, J. M. (2020). *quantmod: Quantitative Financial Modelling Framework*. R package version 0.4.17.

Straumann, D. and Mikosch, T. (2003). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. Technical report, University of Copenhagen.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Ann. Statist.*, 34(5):2449–2495.

Taylor, J. W. and Buizza, R. (2004). A comparison of temperature density forecasts from GARCH and atmospheric models. *Journal of Forecasting*, 23(5):337–355.

Thon, K., Rue, H., Skrøvseth, S. O., and Godtliebsen, F. (2012). Bayesian multiscale analysis of images modeled as Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 56(1):49–61.

Tjøstheim, D. (1978). Statistical spatial series modelling. *Advances in Applied Probability*, 10(1):130–154.

Tjøstheim, D. (1983). Statistical spatial series modelling II: some further results on unilateral lattice processes. *Advances in Applied Probability*, 15(3):562–584.

Tol, R. S. (1996). Autoregressive conditional heteroscedasticity in daily temperature measurements. *Environmetrics*, 7(1):67–75.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601.

Weiss, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric theory*, 2(01):107–131.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., Hester, J., and Chang, W. (2020). *devtools: Tools to Make Developing R Packages Easier*. R package version 2.2.2.

Yao, Q. and Brockwell, P. J. (2006). Gaussian maximum likelihood estimation for ARMA models II: spatial processes. *Bernoulli*, 12(3):403–429.

Part II

Papers

Paper A

A spatio-temporal GARCH model

Sondre Hølleland & Hans A. Karlsen

Journal of Time Series Analysis, **41**(2), 177-209, 2020

JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **41**: 177–209 (2020)
Published online 22 August 2019 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12498

ORIGINAL ARTICLE

A STATIONARY SPATIO-TEMPORAL GARCH MODEL

SONDRE HØLLELAND* AND HANS ARNFINN KARLSEN
Department of Mathematics, University of Bergen, Bergen, Norway

We introduce a lagged nearest-neighbour, stationary spatio-temporal generalized autoregressive conditional heteroskedasticity (GARCH) model on an infinite spatial grid that opens for GARCH innovations in a space-time ARMA model. This is illustrated by a real data application to a classical dataset of sea surface temperature anomalies in the Pacific Ocean. The model and its translation invariant neighbourhood system are wrapped around a torus forming a model with finite spatial domain, which we call circular spatio-temporal GARCH. Such a model could be seen as an approximation of the infinite one and simulation experiments show that the circular estimator with a straightforward bias correction performs well on such non-circular data. Since the spatial boundaries are tied together, the well-known boundary issue in spatial statistical modelling is effectively avoided. We derive stationarity conditions for these circular processes and study the spatio-temporal correlation structure through an ARMA representation. We also show that the matrices defined by a vectorized version of the model are block circulants. The maximum quasi-likelihood estimator is presented and we prove its strong consistency and asymptotic normality by generalizing results from univariate GARCH theory.

Received 14 May 2018; Accepted 29 June 2019

Keywords: Circulant; circular; CSTGARCH; GARCH; space-time; STGARCH

MOS subject classifications: 62F12; 62M10; 62M30.

1. INTRODUCTION

A spatio-temporal generalized autoregressive conditional heteroskedasticity (Spatio-temporal GARCH (STGARCH)) model is a time-space extension of the univariate GARCH models (Engle, 1982; Bollerslev, 1986), but why do we need GARCH in space-time models? For the same reasons we would in time series: whenever we suspect that a white noise series does not have a constant conditional volatility and we want our model to capture this. Characteristics of GARCH are little or no autocorrelation, yet profound correlation of the squared or absolute series. Varying conditional volatility, heavy tailed marginal distribution and clustering of extremes are other traits. The fact that we have a spatial component means that these features also appear in space, that is, the conditional volatility depends on where and when, and extremes cluster in space-time. Utilizing GARCH errors for another model, for example, as innovations in an ARMA model, is a common practice in time series and this can surely also be done in space-time. The models can be used for volatility forecasting and thus improve the quality of prediction intervals by accounting for conditional heteroskedasticity.

The STGARCH model we present has translation invariant neighbourhoods at different time lags that determine the set of variables influencing the future volatility in a traditional GARCH manner. The model is defined on the infinite lattice $\mathbb{Z}^d$, which leads to a boundary problem for a statistical analysis with data confined to a fixed finite spatial region. When a neighbourhood is translation invariant, you lack observations for some of the neighbours at the boundary. A circular modification of the model suggested here both solves the boundary problem and retains a projected version of the translation invariant neighbourhood system. We call this new model circular spatio-temporal GARCH (CSTGARCH). It is clear that it is closely related to the original model and

* Correspondence to: Sondre Hølleland, Department of Mathematics, University of Bergen, Allégaten 41, 5020 Bergen, Norway.
E-mail: sondre.holleland@uib.no

CSTGARCH could be seen as an approximation of STGARCH. As far as we know, alternative approaches also rely on approximations and our simulation experiments indicate that the circular one is a good choice.

The circularity of a CSTGARCH is a feature of the spatial part of the model. A circular spatial model means that points on opposite sides of the area of interest are considered neighbours. The term circular comes from the one-dimensional situation, where we regard spatial locations as points on a line. For a circular model, the two endpoints are neighbours and this can be seen as bending the line into a circle. Certainly, this will have consequences for the structure of the spatial dependencies, but here we will assume that the circular model is the actual situation. In two dimensions, this means wrapping a rectangular grid onto the surface of a torus. For purely spatial processes, Cressie (2015, p. 438) mentioned the circular approach as a possibility for dealing with the boundary problem on infinite lattices. Moran (1973a) constructed an initial circular model on a square torus lattice and let the size of the torus tend to infinity. He considered a stationary Gaussian process with first-order neighbourhood structure, but later extended to certain non-Gaussian processes (Moran, 1973b). The stationary STARMA models considered by both Ali (1979) and Pfeifer and Deutsch (1980) can be circular by specifying a circular neighbourhood matrix. This is done in our real data example in Section 6. The circulation is however more important in the STGARCH case, because we do not observe the conditional volatility process directly. This leads to a more severe boundary problem.

The circular model is elegant in its own right, where a neighbourhood can be translated everywhere on a torus without disturbing its geometrical shape. In this way, the model is within a finite region, but it is still spatially stationary because the gridded torus surface is a group with respect to addition. Moreover, if we view the circulation as a projection of the STGARCH model onto a finite spatial region, while keeping the time dimension, this is the one projection that maintain stationarity. It is difficult to find an alternative projection sharing this property.

Others have proposed models for volatility in a spatial- and spatio-temporal context. Sato and Matsuda (2017) suggested a spatial ARCH model with applications toward land prices in the Tokyo area of Japan, while the spatio-temporal GARCH models of temporal order one proposed by Borovkova and Lopuhaä (2012) is more like a spatially weighted constant correlation coefficient (CCC) GARCH model (Bollerslev, 1990). The ARCH models considered by Otto *et al.* (2018) are primarily spatial, but can also be formulated as spatio-temporal by defining one of the spatial dimensions as time. Such a spatio-temporal formulation is related to the models we consider. They applied the model as innovations of a spatial simultaneous autoregressive model for mortality of lung or bronchus cancer across U.S. counties in their real data example. Robinson (2009) uses a spatial version of stochastic volatility models to demonstrate that lack of spatial correlation does not in general imply spatial independence. He also establishes asymptotic theory for the pseudo-Gaussian maximum likelihood estimator. However, the above mentioned models are not expandable to the infinite space case and not formulated as stationary models with translation invariant neighbourhood systems. In this respect, what we suggest is fundamentally different.

The theory of GARCH models is vast and extensive, so we focus our attention on the most relevant theoretical works here. Nelson (1990) found a criteria for the existence of a unique, ergodic and stationary solution for GARCH(1,1) models and Bougerol and Picard (1992) generalized these results to GARCH(p, q) models. Berkes *et al.* (2003) proved consistency and asymptotic normality of maximum quasi-likelihood estimators for GARCH(p, q). Francq and Zakoïan (2004) removed a smoothness- and a moment restriction on the innovation process, while Straumann and Mikosch (2006) established similar results for a wider class of conditional heteroskedastic time series models. Jeantheau (1998) established strong consistency of the minimum contrast estimator for multivariate GARCH models, especially for the CCC-GARCH. Ling and McAleer (2003) proved consistency and asymptotic normality of the QMLE for a vector ARMA-GARCH using the CCC-formulation. A review of CCC-GARCH was given by Francq and Zakoïan (2011, pp. 279–307), where they also presented proofs of consistency and asymptotic normality for CCC-GARCH models.

Compared to related multivariate GARCH models, CSTGARCH is substantially simpler in several aspects and, in particular, it has fewer parameters due to the specific stationary spatial dependency structure with circular boundaries. Since the spatial region is finite, the model could be motivated as a subclass of CCC-GARCH, but our perspective is more towards the infinite STGARCH model. In addition, it is not a CCC-model in the sense that the number of parameters is independent of the dimension. In C- and STGARCH only quadratic terms are explicitly

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

expressed in the model and thereby avoiding cross terms, same as for the univariate- and CCC models. This is an essential advantage both for revealing theoretical properties and for estimation of the model. It is also inherent in the model as an extension of the univariate GARCH.

We will, in Section 2, introduce both STGARCH and CSTGARCH, and discuss some features especially of the latter. An illustrating example is presented. Then we derive the Gaussian quasi-likelihood in Section 3. In Section 4 we present consistency and asymptotic normality of the maximum quasi-likelihood estimators. A simulation experiment with circular and non-circular data is carried out in Section 5. Following this, a real data application to sea surface temperature anomalies in the Pacific Ocean is presented in Section 6. We make some concluding remarks in Section 7 and the more technical parts are put together in Section 8.

2. SPATIO-TEMPORAL GARCH MODELS

We introduce the spatio-temporal GARCH before turning to the circular version. Neighbourhood systems are defined and we discuss some interesting features of the circular model, such as conditions for stationarity and the spatio-temporal dependence structure imposed by the model.

2.1. Spatio-temporal GARCH

Let $\alpha : \mathbb{Z} \times \mathbb{Z}^d \rightsquigarrow \mathbb{R}_0$ be a function with finite support. For fixed i, $\alpha_i : \mathbb{Z}^d \rightsquigarrow \mathbb{R}_0$. The function β is defined in the same way. We refer to α, β as the parameter functions. The STGARCH model is given by

$$\begin{aligned} X_t(u) &= \sigma_t(u) Z_t(u), \quad u \in \mathbb{Z}^d \\ \sigma_t^2(u) &= \omega + \sum_i \sum_v \alpha_i(v) X_{t-i}^2(u-v) + \sum_i \sum_v \beta_i(v) \sigma_{t-i}^2(u-v), \quad u \in \mathbb{Z}^d, \end{aligned} \tag{2.1}$$

for $t \in \mathbb{Z}$. The modelled process is $\{X_t(u)\}$, while $\{Z_t(u)\}$ is a residual process and $\{\sigma_t(u)\}$ is the volatility process. Let $\mathbb{A}_{1i} = \{v \in \mathbb{Z}^d : \alpha_i > 0\}$ and $\mathbb{A}_{2i} = \{v \in \mathbb{Z}^d : \beta_i > 0\}$ for $i \geq 1$. For $i < p$, the model allows for some zero-valued $\alpha_i(v)$ for $v \in \mathbb{A}_{1i}$ and likewise for the β's. The order of the model is defined as the largest (p, q) so that $\mathbb{A}_{1p}$ and $\mathbb{A}_{2q}$ are non-empty. With the order defined, the second part of (2.1) is expressed more specifically as

$$\sigma_t^2(u) = \omega + \sum_{i=1}^{p} \sum_{v \in \mathbb{A}_{1i}} \alpha_i(v) X_{t-i}^2(u-v) + \sum_{i=1}^{q} \sum_{v \in \mathbb{A}_{2i}} \beta_i(v) \sigma_{t-i}^2(u-v). \tag{2.2}$$

Let $\boldsymbol{\alpha} \stackrel{\text{def}}{=} \{\alpha_i(v),\ v \in \mathbb{A}_{1i},\ i = 1, \ldots, p\}$, $\boldsymbol{\beta} \stackrel{\text{def}}{=} \{\beta_i(v),\ v \in \mathbb{A}_{2i},\ i = 1, \ldots, q\}$ and write $\theta \stackrel{\text{def}}{=} (\omega, \boldsymbol{\alpha}, \boldsymbol{\beta})$ for the parameter vector contained in the parameter space Θ, with the restriction that $\omega > 0$.

2.2. Circular Spatio-Temporal GARCH

Let $\mathbf{m} = (m_1, \ldots, m_d) \in \mathbb{Z}_+^d$ and $\mathcal{R} = \mathcal{R}(\mathbf{m}) = \mathbb{Z}^d / (\mathbf{m}\mathbb{Z}^d)$ be the quotient group of order $\mathbf{m}$. We get the circular version of (2.1) by replacing the infinite spatial index part $\mathbb{Z}^d$ with $\mathcal{R}$. In doing this we do not change the parameter functions but restrict the model so that for $u, v \in \mathcal{R}$, the difference $(u - v)$ and the sum $(u + v)$ are to be understood modulus $\mathbf{m}$ and are therefore also points in $\mathcal{R}$. This means that the stochastic processes involved are indexed on $\mathbb{Z} \times \mathcal{R}$. We will identify $[\mathbf{0}, \mathbf{m} - \mathbf{1}]$ with $\mathcal{R}$. For $v \in \mathbb{Z}^d$, we use the notation $(v|\mathbf{m}) \stackrel{\text{def}}{=} v - \lfloor v/\mathbf{m} \rfloor \circ \mathbf{m} \in \mathcal{R}$, where $\circ$ and $/$ are elementwise or Hadamard multiplication and division respectively, and $\lfloor \cdot \rfloor$ means rounding down to the nearest integer elementwise. We write $v' \sim v$ if $(v'|\mathbf{m}) = (v|\mathbf{m})$. In the circular model, we replace $X_{t-i}^2(u - v)$

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

with $X^2_{t-i}(u - v|\mathbf{m})$, and likewise for the last part of (2.2). Thus, the circular model is given by

$$\begin{aligned} X_t(u) &= \sigma_t(u)Z_t(u), \quad u \in \mathcal{R}, \\ \sigma^2_t(u) &= \omega + \sum_{i=1}^{p} \sum_{v \in \mathbb{\Delta}_{1i}} \alpha_i(v) X^2_{t-i}(u - v|\mathbf{m}) + \sum_{j=1}^{q} \sum_{v \in \mathbb{\Delta}_{2i}} \beta_j(v) \sigma^2_{t-j}(u - v|\mathbf{m}), \quad u \in \mathcal{R}, \end{aligned} \tag{2.3}$$

for $t \in \mathbb{Z}$. Note that in general, $\mathbb{\Delta}_{ki} \not\subseteq \mathcal{R}$. When the indexes are irrelevant, we refer to the process dummies $X = \{X_t(u),\ (t, u) \in \mathbb{Z} \times \mathcal{R}\}$ and likewise for σ and Z.

Some characteristics of the circular model are:

(i) The circular model may be seen as a torus approximation of the infinite model.
(ii) The advantage of the circular model comes from the group structure of $\mathcal{R}$ which retains stationarity in a beneficial way.
(iii) The infinite model and the finite model are different, but share the same parameter vector and residual process.
(iv) If $\mathbf{m}$ is not very small, a window of the infinite process confined to $\mathcal{R}$ would be quite similar to the circularly defined process. The differences that may be seen are mainly at the boundaries of the rectangle.
(v) Both models are designed to be strictly stationary.
(vi) The circular model could be conceived as a mathematical construction to handle the boundaries that are generated from a finite window of the infinite model in a smooth way.
(vii) Estimate of a circular model from data generated by an infinite model can be bias corrected to a meaningful estimate of the true model (cf. Section 5).
(viii) Estimation obstacles created by the boundaries of finite samples from the infinite model is a main motivation for the circular model. In that sense, it could be viewed more as a method than an alternative model. On the other hand, a stationary solution of the circular model does not necessarily rely on a corresponding stationary solution of the infinite model.

Let $\mathbf{X}_t = \{X_t(u),\ u \in \mathcal{R}\}$ be a vector of size m, where $m \stackrel{\text{def}}{=} |\mathcal{R}| = \prod_{i=1}^{d} m_i$ is the number of spatial locations, and likewise for $\mathbf{Z}_t$ and $\boldsymbol{\sigma}^2_t$. We also need $\mathbf{X}^2_t = \mathbf{X}_t \circ \mathbf{X}_t$. Throughout the article, we use the convention that all vectors are column vectors.

In spatial statistics, a neighbourhood set is a collection of sites that influence a given point. There are two requirements to a neighbourhood set: A site cannot be its own neighbour and neighbourhood relations are mutual. In spatio-temporal statistics with time-lagged nearest neighbour dependence, the first condition is not necessary, because we do not have instantaneous spatial dependency. In fact, the opposite is encouraged. Hence, for our neighbourhood systems, which are collections of neighbourhood sets for different temporal lags, we only require that neighbour relations are mutual. In the following theorem, we need a condition on each $\mathbb{\Delta}_{ki}$ to ensure that our neighbourhood systems fulfil this property.

A1: Each parameter domain $\mathbb{\Delta}_{ki}$ is symmetric in the sense that $-\mathbb{\Delta}_{ki} = \mathbb{\Delta}_{ki}$.

The neighbourhood systems for STGARCH are defined by $\mathcal{M}_{ki}(u) \stackrel{\text{def}}{=} u \ominus \mathbb{\Delta}_{ki}$. The spatial part of X that explicitly influences $\sigma_t(u)$ is located at $\cup_i \mathcal{M}_{1i}(u)$ and, in parallel, the direct effect from previous spatial values of the volatility have sites $\cup_i \mathcal{M}_{2i}(u)$. We see that these systems are translation invariant, that is, $\mathcal{M}_{ki}(u + h) = \mathcal{M}_{ki}(u) \oplus h$.

Theorem 2.1. The circular version of the STGARCH model is a CCC-GARCH model of dimension m,

$$\begin{aligned} \mathbf{X}_t &= \boldsymbol{\sigma}_t \circ \mathbf{Z}_t, \\ \boldsymbol{\sigma}^2_t &= \omega \mathbf{1}_m + \sum_{i=1}^{p} \mathbb{A}_i \mathbf{X}^2_{t-i} + \sum_{i=1}^{q} \mathbb{B}_i \boldsymbol{\sigma}^2_{t-i}, \end{aligned} \tag{2.4}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

with components

$$\sigma_t^2(u) = \omega + \sum_{i=1}^{p} \sum_{v \in \mathcal{N}_{1i}(u)} a_i(u,v) X_{t-i}^2(v) + \sum_{i=1}^{q} \sum_{v \in \mathcal{N}_{2i}(u)} b_i(u,v) \sigma_{t-i}^2(v),$$
$$a_i(u,v) \stackrel{\text{def}}{=} \sum_{v' \sim v} \alpha_i(u - v'), \quad \mathbb{A}_i = \{a_i(u,v),\, (u,v) \in \mathcal{R}^2\}, \tag{2.5}$$

and likewise for the b_i's. For the spatial neighbourhood systems we have

(i) Same structure as $\mathcal{M}_{ki}(u)$: $\mathcal{N}_{ki}(u) \stackrel{\text{def}}{=} u \ominus \mathbb{A}_{ki}$, $u \in \mathcal{R}$ with $\mathbb{A}_{ki} \stackrel{\text{def}}{=} (\mathbb{\Delta}_{ki}|\mathbf{m}) \subseteq \mathcal{R}$.
(ii) Translation invariance: $\mathcal{N}_{ki}(u+h) = \mathcal{N}_{ki}(u) \oplus h$ for $u, h \in \mathcal{R}$.
(iii) Mutual neighbourhood relations under A1: $v \in \mathcal{N}_{ki}(u) \Leftrightarrow u \in \mathcal{N}_{ki}(v)$.
(iv) Relation to STGARCH: $\mathcal{N}_{ki}(u) = (\mathcal{M}_{ki}(u)|\mathbf{m})$.

Proof. Let $u \in \mathcal{R}$ be fixed. Since $X_{t-i}^2(\,\cdot\,) = X_{t-i}^2(\,\cdot\,|\mathbf{m})$,

$$\sum_{v \in \mathbb{Z}} \alpha_i(v) X_{t-i}^2(u - v) = \sum_{v \in \mathbb{Z}} \alpha_i(u - v) X_{t-i}^2(v) = \sum_{v \in \mathcal{R}} \sum_{v' \sim v} \alpha_i(u - v') X_{t-i}^2(v')$$
$$= \sum_{v \in \mathcal{R}} \Big[\sum_{v' \sim v} \alpha_i(u - v') \Big] X_{t-i}^2(v) = \sum_{v \in \mathcal{R}} a_i(u,v) X_{t-i}^2(v).$$

The neighbourhood system at $1i$ is for $u \in \mathcal{R}$,

$$\{a_i(u,\cdot) > 0\} = \Big\{ v \in \mathcal{R} : \sum_{v' \sim v} \alpha_i(u - v') > 0 \Big\}$$
$$= \{v \in \mathcal{R},\, \exists h \in \mathbb{Z}^d :\ u - v + h \circ \mathbf{m} \in \mathbb{\Delta}_{1i}\}$$
$$= \{v \in \mathcal{R},\, \exists h \in \mathbb{Z}^d :\ v \in u \ominus \mathbb{\Delta}_{1i} \oplus h \circ \mathbf{m}\}$$
$$= \{v \in \mathcal{R} :\ v \in (u \ominus \mathbb{\Delta}_{1i}|\mathbf{m})\} = u \ominus \mathbb{A}_{1i} = \mathcal{N}_{1i}(u).$$

The translation invariance holds since

$$\mathcal{N}_{ki}(u+h) = (u \oplus h \ominus \mathbb{A}_{ki}|\mathbf{m}) = (\mathcal{N}_{ki}(u) \oplus h|\mathbf{m}) = \mathcal{N}_{ki}(u) \oplus h, \quad u, h \in \mathcal{R},$$

where the addition in the last equality is the group addition since we consider a neighbourhood system on $\mathcal{R}$.

Remains to show that neighbourhood relations are mutual. Let $u \in \mathcal{R}$ and $v \in \mathcal{N}_{ki}(u) = (u \ominus \mathbb{\Delta}_{ki}|\mathbf{m})$. Then, there exists $h \in \mathbb{\Delta}_{ki}$ such that $(u - h|\mathbf{m}) = v$. Thus $u = (v + h|\mathbf{m}) \in (v \oplus \mathbb{\Delta}_{ki}|\mathbf{m}) = \mathcal{N}_{ki}(v)$ due to A1. □

Remark 2.1. It is possible that $\mathbb{\Delta}_{ki}$ does not depend on (k,i), that is, $\mathbb{\Delta}_{ki} = \mathbb{\Delta}$ and $\mathcal{N}_{ki} = \mathcal{N}$ for all (k,i). This is the case in Example 1.

Remark 2.2. If the projection of $\mathbb{\Delta}_{1i} \rightsquigarrow \mathbb{A}_{1i}$ into $\mathcal{R}$ is one-to-one, the sum $\sum_{v' \sim v} \alpha_i(u - v')$ in (2.5) contains at most one non-zero term. Otherwise, the dimension of the parameter space is reduced. It is likewise for $\mathbb{\Delta}_{2i}$.

Before continuing to more theoretical aspects of the model, we present a simple example to illustrate what we have discussed so far.

Example 1. Let $d = 2$ and let the spatial region $\mathcal{R}$ be a 4×4 grid with a circular neighbourhood structure. This means that $\mathbf{m} = (4,4)$, $m = 16$ and the index set $\mathcal{R} = [0,3] \times [0,3]$. As a quotient group, $\mathcal{R}$ is a toroidal surface

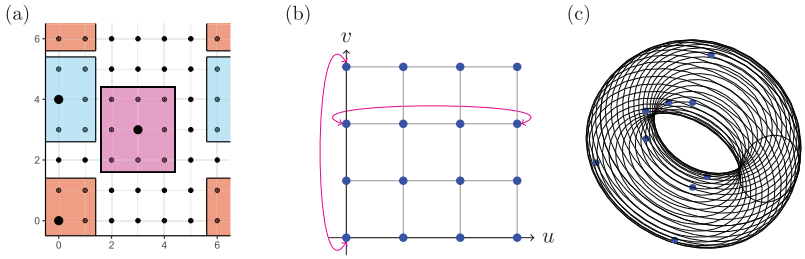

Figure 1. (a) Illustration of the circular neighbourhoods $\mathcal{N}(u)$, on another index set $\mathcal{R} = [0, 6]^2$, for three different values of u. These are marked as the larger sized points and their neighbourhoods are indicated by the different coloured squares. Next, these are two ways of visualizing the area of interest: (b) An equidistant grid and (c) a quotient group, which is a toroidal surface when $d = 2$. [Color figure can be viewed at wileyonlinelibrary.com]

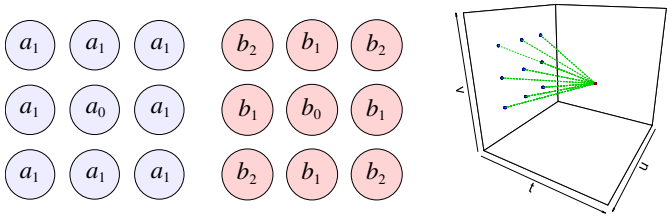

Figure 2. Parameter specification and dependency structure at one time lag. [Color figure can be viewed at wileyonlinelibrary.com]

for $d = 2$. In Figure 1(b,c) we have visualized both $\mathcal{R}$ as an index set and a quotient group respectively, where the 16 points on the equidistant grid corresponds to the points on the torus surface. The circular neighbourhood set is visualized on a larger area of interest, that is, $[0, 6]^2$, in Figure 1a for three different locations.

We present two different ways of parametrizing; one for $\boldsymbol{\alpha}$ and one for $\boldsymbol{\beta}$. Considering a CSTGARCH(1,1) model, we need to specify $\alpha(v)$ and $\beta(v)$. In the notation of (2.3) and Remark 2.1, let $\mathbb{\Delta} = \mathbb{\Delta}_{11} = \mathbb{\Delta}_{21} = \{(v_1, v_2) \in \mathbb{Z}^2 : |v_i| \leq 1, i = 1, 2\}$, or more explicitly

$$\mathbb{\Delta} = \big\{(-1,1), (0,1), (1,1), (-1,0), (0,0), (1,0), (-1,-1), (0,-1), (1,-1)\big\}.$$

It makes sense to have more or less symmetry in the spatial part of the model. Let therefore

$$\alpha(v) = \begin{cases} a_0, & \text{if } v = (0,0), \\ a_1, & \text{if } \max_i |v_i| = 1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad \beta(v) = \begin{cases} b_0, & \text{if } v = (0,0), \\ b_1, & \text{if } \sum_i |v_i| = 1, \\ b_2, & \text{if } |v_1| = 1, |v_2| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

which means that the parameter vector, $\theta = (\omega, a_0, a_1, b_0, b_1, b_2)'$, consists of in total 6 parameters. The parametrization of $\boldsymbol{\alpha}$ gives equal weight to each neighbour (a_1) and another weight to the site of the observation (a_0). For $\boldsymbol{\beta}$ we have one parameter for vertical and horizontal direction (b_1), one for all diagonal neighbours (b_2) and the point itself has its own weight (b_0). The specification of the model is illustrated in Figure 2. This is sometimes called a queen contiguity neighbourhood, as opposed to a rook contiguity, named after the possible movements of the respective chess pieces.

We can also specify the model using (2.5). Let $\mathcal{N} = \mathcal{N}_{11} = \mathcal{N}_{21}$, since $\mathbb{A}_{11} = \mathbb{A}_{21}$. In this case,

$$\mathcal{N}(u) = (u \ominus \mathbb{A}|\mathbf{m}) = \{u\} \cup \mathcal{N}^{(1)}(u) \cup \mathcal{N}^{(2)}(u), \quad u \in \mathcal{R},$$

where $\mathcal{N}^{(1)}(u) = \{(u+v|\mathbf{m}) : \sum_i |v_i| = 1\}$ and $\mathcal{N}^{(2)}(u) = \{(u+v|\mathbf{m}) : |v_1| = 1, |v_2| = 1\}$. Then,

$$a(u,v) = \begin{cases} a_0, & \text{if } v = u, \\ a_1, & \text{if } v \in \mathcal{N}(u)\backslash\{u\}, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad b(u,v) = \begin{cases} b_0, & \text{if } v = u, \\ b_1, & \text{if } v \in \mathcal{N}^{(1)}(u), \\ b_2, & \text{if } v \in \mathcal{N}^{(2)}(u), \\ 0, & \text{otherwise.} \end{cases}$$

In the vector notation of (2.4), $\boldsymbol{\sigma}_t^2 = \omega \mathbf{1}_m + \mathbb{A}\mathbf{X}_{t-1}^2 + \mathbb{B}\boldsymbol{\sigma}_{t-1}^2$, where $\mathbb{B}$ is given by

$$\mathbb{B} = \begin{bmatrix} \mathbb{S} & \mathbb{T} & \mathbb{0} & \mathbb{T} \\ \mathbb{T} & \mathbb{S} & \mathbb{T} & \mathbb{0} \\ \mathbb{0} & \mathbb{T} & \mathbb{S} & \mathbb{T} \\ \mathbb{T} & \mathbb{0} & \mathbb{T} & \mathbb{S} \end{bmatrix}, \quad \mathbb{S} = \begin{bmatrix} b_0 & b_1 & 0 & b_1 \\ b_1 & b_0 & b_1 & 0 \\ 0 & b_1 & b_0 & b_1 \\ b_1 & 0 & b_1 & b_0 \end{bmatrix}, \quad \mathbb{T} = \begin{bmatrix} b_1 & b_2 & 0 & b_2 \\ b_2 & b_1 & b_2 & 0 \\ 0 & b_2 & b_1 & b_2 \\ b_2 & 0 & b_2 & b_1 \end{bmatrix}. \tag{2.6}$$

The matrix $\mathbb{A}$ is obtained by setting $b_0 = a_0$ and $b_1 = b_2 = a_1$ in $\mathbb{B}$. Notice that the matrix also illustrates the neighbourhood structure. We have ordered the rows and columns of $\mathbb{B}$ lexicographically according to the coordinates of $\mathcal{R}$. The block matrix $\mathbb{S}$ represents relations between sites in the same row, while $\mathbb{T}$ represents relations between sites of adjacent rows.

Remark 2.3. If m becomes larger, the order of $\mathbb{A}$ and $\mathbb{B}$, which is $m \times m$, increases. However, the row and column sums will remain the same, $a_0 + 8a_1$ and $b_0 + 4b_1 + 4b_2$ respectively, and the number of non-zero terms in each row is constant. Therefore, the sparsity of the matrix will increase with m. The example illustrates that the number of parameters does not depend on the actual size of the spatial region. We consider m fixed, but in contrast to other multivariate GARCH models, a larger m is beneficial.

2.3. Generalized Circulant and Stationary Structure

It turns out that the model has an interesting circulant algebraic structure which is neither obvious nor intended. Circulant matrices of order n are matrices that can be written on the form

$$\mathbb{C} = \operatorname{circ}(c_0, \ldots, c_{n-1}) = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & \cdots & c_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & \ldots & c_0 \end{bmatrix} = \{c_{ij}\} = \{c(i-j|n)\}.$$

A block circulant is a circulant block matrix whose blocks again are circulants. This is also called a circulant of level 2. In general a circulant of level d is a block circulant whose blocks are circulants of level $d-1$ (Davis, 1994, pp. 184–91).

We will show that the parameter matrices $\mathbb{A}_i$ and $\mathbb{B}_i$ are circulant matrices for $d = 1$. Since we index the matrices on $\mathcal{R}^2$ we use a circulant concept that does not depend on dimension.

Definition 2.1. A matrix $\mathbb{A} = \{a(u,v), (u,v) \in \mathcal{R}^2\}$ is a generalized circulant on $\mathcal{R}$ if $a(u,v) = a(u-v|\mathbf{m})$.

Definition 2.2. A matrix $\mathbb{A}$ defined on $\mathcal{R}^2$ is stationary if for any element $a(u,v) = a(u+v', v+v')$ for all $v' \in \mathcal{R}$.

Proposition 2.1. The set of stationary matrices is closed under matrix addition and multiplication.

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

Proof. For addition, it is obvious. For multiplication, let $\mathbb{C} = \mathbb{A}\,\mathbb{B}$. Then

$$\begin{aligned}\mathbb{C}[u+v', v+v'] &= \sum_w \mathbb{A}[u+v', w]\mathbb{B}[w, v+v'] = \sum_w \mathbb{A}[u+v', w+v']\mathbb{B}[w+v', v+v'] \\ &= \sum_w \mathbb{A}[u, w]\mathbb{B}[w, v] = \mathbb{C}[u, v].\end{aligned}$$

□

Proposition 2.2. Generalized circulants are stationary matrices.

Proof. Let $\mathbb{A} = \{a(u, v), (u, v) \in \mathcal{R}^2\}$ be a generalized circulant. Then $a(u+v', v+v') = a(u+v'-v-v'|\mathbf{m}) = a(u-v|\mathbf{m}) = a(u, v)$ and $\mathbb{A}$ is stationary. □

Theorem 2.2. The model matrices in the CCC-GARCH representation (2.4) have specific algebraic structure. They are:

(i) Generalized circulants.
(ii) Circulants for $d = 1$.
(iii) Block circulants with circulant blocks of level $d-1$, for $d \geq 2$, if $\mathcal{R}$ is ordered lexicographically.

Proof. Let $\mathbb{A}$ be any of the model matrices. We have that $\mathbb{A}$ is a generalized circulant by definition and stationary by Proposition 2.2. For $d = 1$, $\mathbb{A}$ is therefore a circulant. For $d = 2$, we have $\mathbf{m} = (m_1, m_2)$ and we organize $\mathcal{R}$ lexicographically, that is, $u = (u_1, u_2) < v = (v_1, v_2)$ if and only if $u_1 < v_1$ or $u_1 = v_1$ and $u_2 < v_2$. Then $\mathbb{A}$ is a block circulant with circulant blocks, also called circulant of level 2. This can be seen in (2.6) of Example 1, where $\mathbb{S}$ and $\mathbb{T}$ are circulant blocks of $\mathbb{B}$, that is, $\mathbb{B} = \text{circ}(\mathbb{S}, \mathbb{T}, \mathbb{0}, \mathbb{T})$. For $d \geq 2$, $\mathbb{A}$ is a block circulant whose blocks are circulants of level $d-1$, when the rows and columns are ordered lexicographically, by an induction argument. □

2.4. Stationarity

For $p, q \geq 1$, let $\mathbb{Q}_t$ be a square block matrix of order $(p+q-1)$ defined as

$$\mathbb{Q}_t = \begin{bmatrix} \mathbb{A}_1\mathbb{Z}_t^2 + \mathbb{B}_1 & \mathbb{B}_2 & \cdots & \mathbb{B}_{q-1} & \mathbb{B}_q & \mathbb{A}_2 & \cdots & \mathbb{A}_{p-1} & \mathbb{A}_p \\ & \mathbb{I}_{q-1} \otimes \mathbb{I}_m & & & \mathbb{0} & \mathbb{0} & \cdots & \mathbb{0} & \mathbb{0} \\ \mathbb{Z}_t^2 & \mathbb{0} & \cdots & \mathbb{0} & \mathbb{0} & \mathbb{0} & \cdots & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \cdots & \mathbb{0} & \mathbb{0} & & \mathbb{I}_{p-2} \otimes \mathbb{I}_m & & \mathbb{0} \end{bmatrix}, \tag{2.7}$$

where each block is of dimension $m \times m$, $\mathbb{Z}_t^2 = \text{diag}\{\mathbf{Z}_t^2\}$, $\mathbb{I}_k$ is a $k \times k$ identity matrix and $\mathbb{0}$ is a null matrix. Apart from the first row, the matrix $\mathbb{Q}_t$ is defined by $\mathbb{Q}_t[q+1, 1] = \mathbb{Z}_t^2$, $\mathbb{Q}_t[i, i-1] = \mathbb{I}_m$ for $i \in [2, q] \cup [q+2, p+q-1]$ and $\mathbb{0}$ otherwise. Let $\mathbf{V}_t$ and $\mathbf{c}$ both be vectors of dimension $(p+q-1)m$, defined by

$$\mathbf{V}_t = \begin{bmatrix} \mathbf{S}_t \\ \mathbf{Y}_t \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \omega \mathbf{1}_m \\ \mathbf{0} \end{bmatrix}, \tag{2.8}$$

where $\mathbf{S}_t \overset{\text{def}}{=} (\boldsymbol{\sigma}_{t+1}^2, \ldots, \boldsymbol{\sigma}_{t-q+2}^2)$ and $\mathbf{Y}_t \overset{\text{def}}{=} (\mathbf{X}_t^2, \ldots, \mathbf{X}_{t-p+2}^2)$. If $p = 1$ and $q = 0$, $\mathbb{Q}_t = [\mathbb{A}_1\mathbb{Z}_t^2]$, $\mathbf{V}_t = \mathbf{S}_t = \boldsymbol{\sigma}_{t+1}^2$ and $\mathbf{c} = \omega \mathbf{1}_m$. The same is true for $p = q = 1$, except that $\mathbb{Q}_t = [\mathbb{A}_1\mathbb{Z}_t^2 + \mathbb{B}_1]$. Combining (2.7) and (2.8), we have that the CSTGARCH model can be expressed as a stochastic recurrence equation (SRE),

$$\mathbf{V}_t = \mathbb{Q}_t \mathbf{V}_{t-1} + \mathbf{c}, \quad t \in \mathbb{Z}. \tag{2.9}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

Under an i.i.d. assumption on the residuals, this is the state space representation of the CSTGARCH model since the sequence $\{(\mathbb{Q}_t, \mathbf{c})\}$ is i.i.d. and any reasonable solution $\{\mathbf{V}_t\}$ is Markov. From the SRE in (2.9), we formally get

$$\mathbf{V}_t = \sum_{k=0}^{\infty} \prod_{\ell=0}^{k-1} \mathbb{Q}_{t-\ell}\, \mathbf{c}.$$

When $\mathbb{E}\log^+\|\mathbb{Q}_0\| < \infty$, the Lyapunov exponent is defined and satisfy

$$\gamma_{\mathbb{Q}} = \inf\left\{ \mathbb{E}\left[n^{-1}\log\|\mathbb{Q}_0\mathbb{Q}_{-1}\cdots\mathbb{Q}_{-n+1}\|\right], n \in \mathbb{N}\right\} = \lim_{n\to\infty} n^{-1}\log\|\mathbb{Q}_0\mathbb{Q}_{-1}\cdots\mathbb{Q}_{-n+1}\|,$$

for any matrix operator norm. The maximum absolute row sum, defined below, is a convenient choice in this context.

Definition 2.3. Let $\|\mathbb{M}\| \stackrel{\text{def}}{=} \|\mathbb{M}\|_\infty = \sup\{\|\mathbb{M}\mathbf{x}\|_\infty : \|\mathbf{x}\|_\infty = 1\}$ for a matrix $\mathbb{M}$.

We will refer to the following assumptions, where θ_0 denotes the true parameter.

A2: The residual process Z is i.i.d. and $Z \in L^{2\delta}$ for some $\delta > 0$.
A3: At θ_0, the Lyapunov exponent is strictly negative, $\omega > 0$ and $\mathbb{E}\log^+\|\mathbb{Q}_0\| < \infty$.

Remark 2.4. The condition $\mathbb{E}\log^+\|\mathbb{Q}_0\| < \infty$ is implied by A2 together with $\omega > 0$.

An important submatrix of $\mathbb{Q}_t$ is the $q \times q$ non-negative block matrix

$$\mathbb{B} = \mathbb{B}(\theta) \stackrel{\text{def}}{=} \begin{bmatrix} \mathbb{B}_1 & \mathbb{B}_2 & \cdots & \mathbb{B}_{q-1} & \mathbb{B}_q \\ & \mathbb{I}_{q-1}\otimes\mathbb{I}_m & & & \mathbb{0} \end{bmatrix} = \{\mathbb{B}_{ij}, 1 \le i,j \le q\}, \tag{2.10}$$

where each block is $m \times m$. When the model has an ergodic solution, this matrix will be the driving force for the vectorized form of the forthcoming likelihood process in Section 3.

Theorem 2.3. Let $s_{\mathbb{B}} \stackrel{\text{def}}{=} \sum_{j=1}^{q}\sum_{v}\beta_j(v)$ and $\rho_{\mathbb{B}}$ be the spectral radius of $\mathbb{B}$.

(i) If $\rho_{\mathbb{B}} < 1$, then $s_{\mathbb{B}} = \|\mathbb{B}^q\|$ and $s_{\mathbb{B}} \le \rho_{\mathbb{B}} \le s_{\mathbb{B}}^{1/q}$.

Assume that A3 holds. Then for $\theta = \theta_0$,

(ii) $\rho_{\mathbb{B}} < 1$.
(iii) There is a unique adapted ergodic solution of (2.3).
(iv) If also A2 is satisfied, then $X \in L^{2\delta}$ and $\sigma \in L^{2\delta}$.

Proof. By Bougerol and Picard (1992) (iii) is true and by Berkes *et al.* (2003) we get (iv). For (ii) we see that $\|\prod_{j=0}^{n-1}\mathbb{Q}_{t-j}\| \ge \|\mathbb{B}^n\|$. It remains to prove (i).

Assume that $\rho_{\mathbb{B}} < 1$. Then $\sum_{j=1}^{q}\mathbb{B}_j\mathbf{1}_m = s_{\mathbb{B}}\mathbf{1}_q$ and $\mathbb{B}\mathbf{1}_{mq} \ge (s_{\mathbb{B}} \wedge 1)\mathbf{1}_{mq}$, which shows that $s_{\mathbb{B}} \le \rho_{\mathbb{B}} < 1$. Let $\mathbb{S}_i^{(k)} = \sum_{j=1}^{q}\mathbb{B}_{ij}^{(k)}$ be the ith block row sum of $\mathbb{B}^k$ with $\mathbb{S}_i = \mathbb{S}_i^{(1)}$ and $\mathbb{S}^{(k)} = \text{vecblock}\left[\mathbb{S}_1^{(k)} \ \dots \ \mathbb{S}_q^{(k)}\right]$, a block matrix of dimension $q \times 1$. Now,

$$\mathbb{S}_i^{(k+1)} = \sum_{\ell=1}^{q}\left[\sum_{j=1}^{q}\mathbb{B}_{ij}^{(k)}\mathbb{B}_{j\ell}\right] = \sum_{j=1}^{q}\mathbb{B}_{ij}^{(k)}\mathbb{S}_j,$$

$$\left\|\mathbb{S}_i^{(k+1)}\right\| = \left\|\sum_{j=1}^{q}\mathbb{B}_{ij}^{(k)}\mathbb{S}_j\right\| \le \left\|\mathbb{S}_i^{(k)}\right\|\|\mathbb{S}\| = \left\|\mathbb{S}_i^{(k)}\right\|,$$

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

since

$$\Big\| \sum_{j=1}^{q} \mathbb{B}_{ij}^{(k)} \mathbb{S}_j \Big\| = \left\| \begin{bmatrix} \mathbb{B}_{i1}^{(k)} & \cdots & \mathbb{B}_{iq}^{(k)} \\ \mathbb{0} & \mathbb{I}_{q-2} \otimes \mathbb{0} & \mathbb{0} \end{bmatrix} \begin{bmatrix} \mathbb{S}_1 & & \mathbb{0} \\ \vdots & \mathbb{I}_{q-2} \otimes \mathbb{0} & \vdots \\ \mathbb{S}_q & & \mathbb{0} \end{bmatrix} \right\|.$$

From the definition of $\mathbb{B}$, we see that

$$\begin{aligned} \mathbb{B}_{ij} = \begin{cases} \mathbb{B}_j, & \text{for } i = 1, \\ \mathbb{I}_m \delta_{i-1,j}, & \text{for } i \in [2, q], \end{cases} \quad \Longrightarrow \quad \mathbb{B}_{ij}^{(2)} = \begin{cases} \mathbb{B}_{1j}^{(2)}, & \text{for } i = 1, \\ \mathbb{B}_j, & \text{for } i = 2, \\ \mathbb{I}_m \delta_{i-2,j}, & \text{for } i \in [3, q], \end{cases} \\ \Longrightarrow \mathbb{B}_{ij}^{(k)} = \begin{cases} \mathbb{B}_{ij}^{(k)}, & \text{for } i \in [1, k-1], \\ \mathbb{B}_j, & \text{for } i = k, \\ \mathbb{I}_m \delta_{i-2,j}, & \text{for } i \in [k+1, q], \end{cases} \quad \text{and} \quad \|\mathbb{S}_i^{(k)}\| \le s_{\mathbb{B}},\ i \in [1, k], \end{aligned} \tag{2.11}$$

for $k \in [2, q-1]$, where the implication is revealed by tracking what happens with the rows going from $\mathbb{B}^{k-1}$ to $\mathbb{B}^k$ when multiplying by $\mathbb{B}$. From (2.11), a last matrix multiplication gives $\mathbb{B}_{qj}^{(k)} \equiv \mathbb{B}_j$ for all $j \in [1, q]$ and

$$\begin{aligned} \big\|\mathbb{S}_i^{(q)}\big\| \le s_{\mathbb{B}}, \quad i \in [1, q-1] \quad \text{and} \quad \big\|\mathbb{S}_q^{(q)}\big\| = s_{\mathbb{B}}, \\ \|\mathbb{B}^q\| = \big\|\mathbb{S}^{(q)}\big\| = \max_i \big\|\mathbb{S}_i^{(q)}\big\| = s_{\mathbb{B}}. \end{aligned}$$

Since $\|\mathbb{B}^{nq}\|^{1/nq} \le \|\mathbb{B}^q\|^{1/q} = s_{\mathbb{B}}^{1/q}$, we must have $\rho_{\mathbb{B}} \le s_{\mathbb{B}}^{1/q}$. □

Remark 2.5. Theorem 2.3 will also hold locally for all θ in a sufficiently small compact neighbourhood of the non-zero part of θ_0. We also see that $\rho_{\mathbb{B}} < 1$ is a necessary, but not sufficient condition for the existence of a stationary solution of $\{\mathbf{X}_t\}$.

As long as A3 is fulfilled, $\{X_t(u)\}$ is a strictly stationary process in the sense that, for all $n \ge 1$ and for any $(k, v) \in \mathbb{Z} \times \mathcal{R}$,

$$\{X_t(u),\ (t,u) \in [1,n] \times \mathcal{R}\} \overset{d}{=} \{X_{t+k}(u+v),\ (t,u) \in [1,n] \times \mathcal{R}\},$$

by Theorem 2.3(iii). Like for univariate GARCH(p, q) models, a CSTGARCH(p, q) is weakly stationary if and only if $\mathbb{E}\, Z^2$ is finite and

$$s_{\mathrm{AB}} \overset{\text{def}}{=} \sum \alpha_i(v) + \sum \beta_j(v) < 1. \tag{2.12}$$

This condition is much easier verified and does not depend on $\mathcal{R}$, but it is somewhat more restrictive than A2–A3 with $\delta = 1$. We will prove this fact next.

Theorem 2.4. Suppose that A2 holds for a $\delta \in (0, 1]$. Let $\mathbb{Q}_t = \{q_t(v, v'),\ (v, v') \in \mathcal{R}^2\}$ and $\mathbb{Q}_t^{\circ\delta} = \{q_t^\delta(v, v')\}$. Then

$$\mathbf{1}' \mathbb{E}^k \mathbb{Q}_0^{\circ\delta} \mathbf{1} = o(1) \quad \text{w.r.t. } k, \tag{2.13}$$

is sufficient for the Lyapunov part of A3.

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

Proof. Let $\mathbb{F}_k = \prod_{j=0}^{k-1} \mathbb{Q}_{t-j}$ and $\mathbb{F}_k^{\circ\delta} = \prod_{j=0}^{k-1} \mathbb{Q}_{t-j}^{\circ\delta}$. By the row-sum norm,

$$\|\mathbb{F}_k\| = \Big\| \prod_{j=0}^{k-1} \mathbb{Q}_{t-j} \Big\| = \text{max row sum } \mathbb{F}_k,$$
$$\|\mathbb{F}_k\|^\delta = [\text{max row sum } \mathbb{F}_k]^\delta \le \text{max row sum } \mathbb{F}_k^{\circ\delta} = \|\mathbb{F}_k^{\circ\delta}\|.$$

Thus

$$\mathbb{E}\|\mathbb{F}_k\|^\delta \le \mathbb{E}\|\mathbb{F}_k^{\circ\delta}\| \le \mathbb{E} \sum_{v_0,\dots,v_k} \prod_{j=0}^{k-1} q_{t-j}^\delta(v_j, v_{j+1}) = \sum_{v_0,\dots,v_k} \prod_{j=0}^{k-1} \mathbb{E} q_{t-j}^\delta(v_j, v_{j+1})$$
$$= \sum_{v_0,\dots,v_k} \prod_{j=0}^{k-1} \mathbb{E} q_0^\delta(v_j, v_{j+1}) = \mathbf{1}' \mathbb{E}^k \mathbb{Q}_0^{\circ\delta} \mathbf{1} < 1,$$

for k large enough. □

Remark 2.6. This is a replacement inequality. We can replace all stochastic terms in $\mathbb{Q}_{t-j}^{\circ\delta}$ by their respective expected value.

Corollary 2.1. Let $a = q \vee (p-1)$, $r = p + q - 1$ and $\mathbb{G} = \mathbb{E}\mathbb{Q}_0$. Suppose that $s_{\text{AB}} < 1$. Then $s_{\text{AB}} = \|\mathbb{G}^a\|$ and $\|\mathbb{F}_k\| < 1$ for $k = ha$ with $h > -\log(rm)/\log s_{\text{AB}}$.

Proof. Denote the first row of $\mathbb{G}$ as the ρ-row. We have that

$$\mathbb{G}_{ij} = \begin{cases} \mathbb{I}_m \delta_{1,i} \mathbb{G}_{1j} & \text{for } i = 1,\ j \ge 1, \\ \mathbb{I}_m \delta_{i-1,j}, & \text{for } i \in [2, q], \\ \mathbb{I}_m \delta_{1,j}, & \text{for } i = q+1, \\ \mathbb{I}_m \delta_{i-1,j}, & \text{for } i \in [q+2, r]. \end{cases}$$

The interesting rows consist of the two sets $[2, q]$ and $[q+2, r]$ which have only one non-zero block equal to the unit matrix $\mathbb{I}_m$. We consider $\mathbb{G}^k$ for $k \ge 1$. The row $q+1$ changes to the first row after one iteration. For each iteration the two sets of interesting rows are reduced by one row as their respective top row is removed and the non-zero block in each of the remaining rows is shifted one step to the left. So for $k = 2$ these sets are $[3, q]$ and $[q+3, r]$. This process continues until both sets are empty and that happens exactly after $k = a$ steps. In this situation either row q or row r in $\mathbb{G}^k$ equals the ρ-row. During this sequence of iterations all rows have been a ρ-row. Since $\|\sum_j \mathbb{G}_{ij}\| \le \|\mathbb{I}\|$ for all i, any block row sum is decreasing by each iteration. Now, let $k = ha$,

$$\mathbf{1}' \mathbb{E}^k \mathbb{Q}_0 \mathbf{1} \le rm \, \|\mathbb{G}^k\| \le rm \, \|\mathbb{G}^a\|^h = rm \, s_{\text{AB}}^h < 1,$$

for $h > -\log(rm)/\log s_{\text{AB}}$. □

Corollary 2.2. Let $s_{\text{AB}}^{(\delta)} = \sum \mathbb{E}|\alpha_j(v) Z_j^2(v) + \beta_j(v)|^\delta$. Suppose that $s_{\text{AB}}^{(\delta)} < 1$. Then $s_{\text{AB}}^{(\delta)} = \|\mathbb{E}^{q\vee(p-1)} \mathbb{Q}_0^{\circ\delta}\|$ and $\mathbb{E}\|\mathbb{F}_k\|^\delta < 1$ for $k = ha$ with $h > -\log(rm)/\log s_{\text{AB}}^{(\delta)}$.

Proof. Invoke the previous proof. □

Remark 2.7. The sufficient condition in Corollary 2.2 does not depend on $\mathbf{m}$.

Remark 2.8. For $p = q = 1$, we get from (2.13) with $k = 1$,

$$s_{\mathrm{AB}}^{(\delta)} = \sum_{v} \mathbb{E}|\alpha(v)Z_1^2(v) + \beta(v)|^{\delta} < 1,$$

which corresponds to the stationarity requirement of Nelson (1990) for univariate GARCH(1, 1).

2.5. Spatio-temporal Dependency

Since CSTGARCH is a spatio-temporal model, the spatial dependence structure is particularly interesting. It is well known that a GARCH(p, q) process has an ARMA$(p \vee q, q)$ representation. Here we use such an ARMA representation for a CSTGARCH process to study the spatio-temporal dependence structure implied by the model. We derive the ARMA representation, find an expression for the autocovariance function for the squared process and illustrate the dependence structure for the CSTGARCH(1,1).

Theorem 2.5. Suppose that (2.12) holds and let $r = p \vee q$. Then an extended spatio-temporal autocorrelation function exists and is given by

$$\rho(h, v) = \mathbb{R}(h)[u + v, u], \quad (h, v) \in \mathbb{Z} \times \mathcal{R}, \tag{2.14}$$

for any $u \in \mathcal{R}$, where

$$\begin{aligned} \mathbb{R}(h) &= \operatorname{diag}^{-1}\left(\sum_{j=0}^{\infty} \Psi_j \Psi_j'\right) \sum_{k=0}^{\infty} \Psi_{k+h} \Psi_k', \\ \Psi_k &= \begin{cases} \mathbb{0}, & \text{for } k < 0, \\ \mathbb{I}_m, & \text{for } k = 0, \\ \sum_{j=1}^{r\wedge k} (\mathbb{A}_j + \mathbb{B}_j)\Psi_{k-j} - \mathbb{B}_k, & \text{for } k > 0. \end{cases} \end{aligned} \tag{2.15}$$

Proof. It is well known that we can rewrite a CCC-GARCH model as a VARMA(r, q) model,

$$\begin{aligned} \mathbf{X}_t^2 &= \boldsymbol{\sigma}_t^2 \circ \mathbf{Z}_t^2 = \boldsymbol{\sigma}_t^2 + \mathbf{U}_t, \quad \mathbf{U}_t \overset{\text{def}}{=} \boldsymbol{\sigma}_t^2 \circ (\mathbf{Z}_t^2 - \mathbf{1}_m), \text{ so that } \boldsymbol{\sigma}_t^2 = \mathbf{X}_t^2 - \mathbf{U}_t, \\ &= \omega \mathbf{1}_m + \sum_{i=1}^{r} (\mathbb{A}_i + \mathbb{B}_i)\mathbf{X}_{t-i}^2 + \sum_{i=1}^{q} (-\mathbb{B}_i)\mathbf{U}_{t-i} + \mathbf{U}_t \\ &= \omega \mathbf{1}_m + \sum_{i=1}^{r} \Phi_i \mathbf{X}_{t-i}^2 + \sum_{i=0}^{q} \Theta_i \mathbf{U}_{t-i}, \text{ say.} \end{aligned}$$

This is a second order stationary causal VARMA(r, q) model if the roots of the corresponding determinant of the characteristic matrix polynomial are strictly outside the unit circle and the residual process $\{\mathbf{U}_t\}$ has finite second-order moment (Brockwell and Davis, 1991, Theorem 11.3.1, p. 418). Moreover, by the same theorem, $\mathbf{X}_t^2 = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Psi_j \mathbf{U}_{t-j}$, where the filter is given by (2.15) and $\boldsymbol{\mu}$ is the expectation of $\mathbf{X}_t^2$. The root condition is implied by (2.12) and for the moment one we proceed by assuming it holds. Later on we will relax this assumption.

Now, the covariance matrix of $\mathbf{U}_t$ is proportional to the unit matrix and therefore the multivariate autocovariance, $\mathbb{\Gamma}$, function for $\{\mathbf{X}_t^2\}$ is given by

$$\mathbb{\Gamma}(h) = \operatorname{Cov}(\mathbf{X}_{t+h}^2, \mathbf{X}_t^2) \propto \sum_{k} \Psi_{k+h} \Psi_k. \tag{2.16}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

By definition,

$$\mathbb{R}(h) = [\mathrm{diag}^{-1/2}\,\Gamma(0)]\,\Gamma(h)\,[\mathrm{diag}^{-1/2}\,\Gamma(0)],$$

and that gives the first part of (2.15). We see that (2.15) itself does not rely on the assumption of finite variance of $\mathbf{X}_t^2$. It just requires (2.12) to make sense and for that purpose we use $\mathbb{I}_m$ for the covariance matrix of $\mathbf{U}_t$.

It remains to show that $\mathbb{R}(h)$ is a stationary matrix for any h in the sense of Definition 2.2. All matrices in (2.4) are stationary (Theorem 2.2), which in turn implies that all the coefficients in the VARMA(r, q) formulation are stationary (Proposition 2.1). Hence, $\{\Psi_j\}$ is stationary. By the same argument now used on (2.16), we see that $\Gamma(h)$ is also stationary and therefore (2.14) holds. The stationarity implies that $\Psi_k(u, u) = \Psi_k(u + v, u + v) = \Psi_k(v, v)$, which also means that $\mathrm{diag}\,\Gamma(0)$ is proportional to the unit matrix. □

Remark 2.9. If $\mathbf{X}_t^2$ has finite variance, then $\rho(h, v) = \mathrm{Cor}(X_{t+h}^2(u + v), X_t^2(u))$.

Remark 2.10. Alternatively to (2.15), we can use the multivariate Yule–Walker equations (Brockwell and Davis, 1991, 11.3.12, 11.3.15, pp. 419–20) to describe and compute $\mathbb{R}$,

$$\Gamma(h) = \sum_{j=1}^{r} \Phi_j \Gamma(h - j) + \sum_{j=1}^{q} \Theta_j \Psi'_{j-h}, \text{ for } h \geq 0.$$

Example 2. For $p = q = 1$ with $\mathbb{S} = \Gamma(0)$,

$$\mathbb{S} = \Gamma(0) = \Phi\,\Gamma'(1) + \mathbb{I}_m + \Theta(\Phi' + \Theta'), \quad \Gamma(1) = \Phi\,\mathbb{S} + \Theta, \quad \Gamma(h) = \Phi\,\Gamma(h - 1), \quad h \geq 2,$$

which gives $\mathbb{C} = \mathbb{I}_m + (\Phi\,\Theta' + \Theta\,\Phi') + \Theta\,\Theta'$ and $\mathbb{S} = \Phi\,\mathbb{S}\,\Phi' + \mathbb{C}$. For $\mathbb{R}$ this means

$$\mathbb{R}(h) = \begin{cases} \mathrm{diag}^{-1}(\mathbb{S})\,\mathbb{S}, & \text{for } h = 0, \\ \mathrm{diag}^{-1}(\mathbb{S})\,\Phi^{h-1}\,[\Phi\,\mathbb{S} + \Theta], & \text{for } h \geq 1, \end{cases} \qquad \mathbb{S} = \sum_{k=0}^{\infty} \Phi^k \mathbb{C} \Phi^{k\prime}.$$

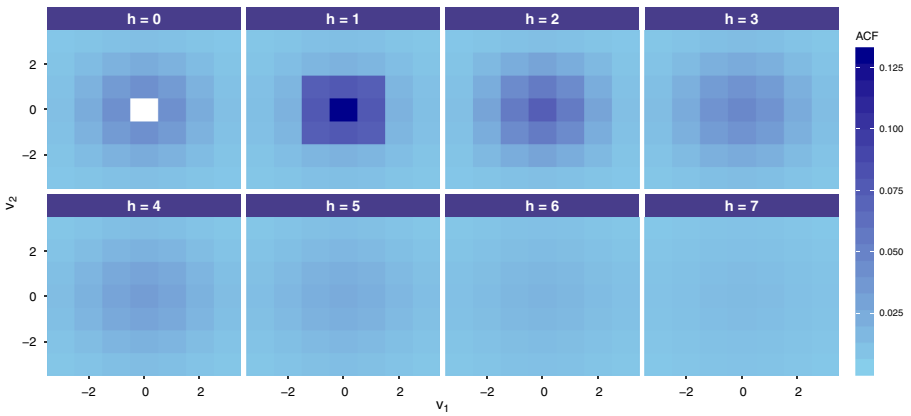

Figure 3. Correlation as function of spatio-temporal lag (h, v) with $h \in [0, 7]$ and $v \in [-3, 3] \times [-3, 3]$. At $h = 0$, the square at $(0, 0)$ is white, because this correlation is 1 and would inflate the scale if included. [Color figure can be viewed at wileyonlinelibrary.com]

Example 3. For $\mathcal{R} = [0,6] \times [0,6]$. we visualize the spatio-temporal correlation $\rho(h,v) = \text{Cor}(X^2_{t+h}(u+v|\mathbf{m}), X^2_t(u))$, with time lag $h \in [0,7]$ and spatial lag $v = (v_1, v_2) \in [-3,3] \times [-3,3]$, in Figure 3. The ACF is given by (2.14) and the model specifications are the same as Example 1, with $a_0 = 0.1$, $a_1 = 0.4/8$, $b_0 = 0.06$ and $b_1 = b_2 = 0.3/8$, giving $\sum_v \alpha(v) + \sum_v \beta(v) = 0.86 < 1$. The scale parameter ω does not influence the correlation, thus it is not specified. For $h = 1$, the nine dark blue squares are the nine neighbours. As the temporal lag increase, the correlation spreads out and fades. An interesting element is the low magnitude of the correlations, peaking around 0.13. This is most likely due to the high number of neighbours and the chosen parameters.

3. CONDITIONAL MAXIMUM QUASI-LIKELIHOOD ESTIMATION

In the first part of this section, we introduce a version of the volatility process that is defined by a free parameter. With this framework established, we derive the conditional quasi-likelihood.

Let θ_0 be the true parameter vector which has generated the $\{\mathbf{X}_t\}$ process and the not directly observable squared volatility process $\{\sigma^2_t(u)\}$. Related to the last one is $\{h_t(u,\theta),\ \theta \in \Theta\}$ which we call the likelihood process. It extends the squared volatility process $\{\sigma^2_t(u)\}$ to a function-valued process. Define equivalently to (2.4),

$$\mathbf{h}_t \overset{\text{def}}{=} \omega\mathbf{1} + \sum_{i=1}^{p} \mathbb{A}_i \mathbf{X}^2_{t-i} + \sum_{j=1}^{q} \mathbb{B}_j \mathbf{h}_{t-j}, \quad t \in \mathbb{Z}. \tag{3.1}$$

with $\mathbf{h}_t = \mathbf{h}_t(\theta) = \{h_t(u,\theta),\ u \in \mathcal{R}\} \in \mathbb{R}^m$. Note that $\theta = (\omega, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Theta$, while $\{\mathbf{X}_t\}$ is generated by θ_0. Thus, we have that $\mathbf{h}_t(\theta_0) = \boldsymbol{\sigma}^2_t$ a.s. By the forthcoming assumptions A4 and A5, it will become clear that $\{h_t(u,\cdot)\}$ is a spatio-temporal process with values in $\mathbb{C}[\Theta, \mathbb{R}]$, the space of real continuous functions defined on Θ, uniquely defined by (3.1).

We vectorize (3.1) with $\mathbb{B}$ defined in (2.10) and the $mq \times 1$ vector processes

$$\mathbf{H}_t \overset{\text{def}}{=} \begin{bmatrix} \mathbf{h}_t & \mathbf{h}_{t-1} & \cdots & \mathbf{h}_{t-q+1} \end{bmatrix}, \qquad \mathbf{C}_t \overset{\text{def}}{=} \begin{bmatrix} \mathbf{D}_t & \mathbf{0} \end{bmatrix}, \tag{3.2}$$

where

$$\mathbf{D}_t \overset{\text{def}}{=} \omega\mathbf{1}_m + \sum_{i=1}^{p} \mathbb{A}_i \mathbf{X}^2_{t-i} \quad \text{and} \quad \mathbf{h}_t = \sum_{j=1}^{q} \mathbb{B}_j \mathbf{h}_{t-j} + \mathbf{D}_t \tag{3.3}$$

from (3.1). If $q \in \{0,1\}$, let $\mathbb{B} = \mathbb{B}_{11} = \mathbb{B}_1$, $\mathbf{C}_t = \mathbf{D}_t$ and $\mathbf{H}_t = \boldsymbol{h}_t$ with $\mathbb{B}_1 = \mathbb{0}$ for $q = 0$. Note that we have suppressed the dependency of θ, but $\mathbb{B}$, $\mathbf{H}_t$ and $\mathbf{C}_t$ depend on θ. From (2.10) and (3.1)–(3.3), we get a first order SRE for the likelihood process given by

$$\mathbf{H}_t = \mathbb{B}\,\mathbf{H}_{t-1} + \mathbf{C}_t, \qquad t \in \mathbb{Z}. \tag{3.4}$$

Note that $\mathbf{H}_t(\theta_0) = \mathbf{S}_{t-1}$ in the context of (2.8). The special case when $q = 0$ gives $\mathbf{h}_t = \mathbf{D}_t$ in (3.4) and $\mathbf{h}_t$ is then fully observable from the observations. Most of what we discuss here is only relevant when $q > 0$, namely when the model is not a pure ARCH.

An important concept for SRE's is convergence with exponential rate.

Definition 3.1. For a sequence $\{X_n\}$, we write $X_n \xrightarrow{\text{e.a.s.}} 0$ or $X_n = o(1)$ e.a.s. if $X_n = o(a^n)$ a.s. for some fixed $a \in (0,1)$.

Lemma 3.1. Suppose that $X_n = o(1)$ e.a.s and $\{Y_n\}$ is bounded in L^δ for a $\delta > 0$. Then $X_n Y_n = o(1)$ e.a.s.

Proof. The proof is straightforward (Straumann and Mikosch, 2006). □

Remark 3.1. If $X_n = \mathcal{O}(a^n)$ a.s. for a fixed $a \in (0,1)$, then $X_n = o(1)$ e.a.s.

By carrying out the recurrence in (3.4), we get

$$\mathbf{H}_t = \sum_{k=0}^{\infty} \mathbb{B}^k \mathbf{C}_{t-k}. \tag{3.5}$$

The infinite sum in (3.5) converges e.a.s. uniformly with respect to θ contained in any compact subset of the neighbourhood described in Remark 2.5. To make sure that this holds globally on Θ, we need the following assumptions:
A4: On Θ, $\|\sum_{j=1}^{q} \mathbb{B}_j\| < 1$.
A5: The parameter space Θ is compact.

We extend Definition 2.3.

Definition 3.2. For $g : \Theta \rightsquigarrow \mathbb{R}^{r\times s}$, $r, s \geq 1$, $S \subseteq \Theta$, let $\|g\|_S \stackrel{\text{def}}{=} \sup_{\theta\in S} \|g(\theta)\|$.

Since we do not observe the infinite past of $\{\mathbf{C}_t\}$, let

$$\widehat{\mathbf{X}}_t \stackrel{\text{def}}{=} \begin{cases} 0, & \text{for } t < 1-p, \\ \widetilde{\mathbf{x}}_t, & \text{for } t = 1-p, \dots, 0, \\ \mathbf{X}_t, & \text{for } t = 1, \dots, n, \end{cases} \qquad \widehat{\mathbf{h}}_t \stackrel{\text{def}}{=} \begin{cases} \mathbf{0}, & \text{for } t < 1-q, \\ \widetilde{\mathbf{h}}_t, & \text{for } t = 1-q, \dots, 0, \\ \widehat{\mathbf{D}}_t + \sum_{i=1}^{q} \mathbb{B}_i \widehat{\mathbf{h}}_{t-i}, & \text{for } t = 1, \dots, n, \end{cases} \tag{3.6}$$

with $\widehat{\mathbf{D}}_t \stackrel{\text{def}}{=} \omega \mathbf{1}_m + \sum_{i=1}^{p} \mathbb{A}_i \widehat{\mathbf{X}}_{t-i}^2$ and where $\{\widetilde{\mathbf{x}}_t\}$ and $\{\widetilde{\mathbf{h}}_t\}$ are initial values. The theoretical counterpart of this definition is (3.3). Note that $\widehat{\mathbf{D}}_t \equiv \mathbf{D}_t$ for $t \geq p+1$.

Remark 3.2. The initial values represent fixed values that are different from the observations. However, our subsequently derived results allow $\widetilde{\mathbf{h}}_t$ to be stochastic as long as $\mathbb{E}\|\widetilde{\mathbf{h}}_t\|_\Theta^\delta < \infty$ for some $\delta > 0$. The same applies to $\widetilde{\mathbf{x}}_t^2$.

In Proposition 3.1 we prove that the difference between the empirical $\widehat{\mathbf{h}}_t$ and the stationary process $\mathbf{h}_t$ will converge e.a.s. to zero uniformly on Θ. This implies that the effect of the initial values is asymptotically negligible.

Proposition 3.1. If A2–A5 are satisfied, then $\|\mathbf{h}_t - \widehat{\mathbf{h}}_t\|_\Theta = o(1)$ e.a.s.

Proof. Let $\widehat{\mathbf{C}}_t \stackrel{\text{def}}{=} (\widehat{\mathbf{D}}_t, \mathbf{0}, \dots, \mathbf{0}) \in \mathbb{R}^{mq}$ and $\widehat{\mathbf{H}}_t \stackrel{\text{def}}{=} (\widehat{\mathbf{h}}_t, \dots, \widehat{\mathbf{h}}_{t-q+1})$. Then for $t \in [p+1, n]$, we have that $\widehat{\mathbf{C}}_t = \mathbf{C}_t$, and

$$\widehat{\mathbf{H}}_t = \mathbb{B}\widehat{\mathbf{H}}_{t-1} + \widehat{\mathbf{C}}_t = \mathbb{B}^{t-p-1}\widehat{\mathbf{H}}_{p+1} + \sum_{j=0}^{t-p-2} \mathbb{B}^j \mathbf{C}_{t-j}.$$

Thus

$$\mathbf{H}_t - \widehat{\mathbf{H}}_t = \mathbb{B}^{t-p-1}(\mathbf{H}_{p+1} - \widehat{\mathbf{H}}_{p+1}). \tag{3.7}$$

Using the triangle inequality on (3.7), we have

$$\|\mathbf{H}_t - \widehat{\mathbf{H}}_t\|_\Theta \leq \|\mathbb{B}^{t-p-1}\|_\Theta \big(\|\mathbf{H}_{p+1}\|_\Theta + \|\widehat{\mathbf{H}}_{p+1}\|_\Theta\big), \tag{3.8}$$

and by A4 and A5 together with Theorem 2.3(i),

$$\|\mathbb{B}^t\|_\Theta \leq \|\mathbb{B}^q\|_\Theta^{\lfloor t/q \rfloor} \max_{0\leq r<q} \|\mathbb{B}^r\|_\Theta = o(1) \text{ e.a.s.}$$

The other terms are finite with probability one, so (3.8) must go to zero with an exponential rate as $t \to \infty$. □

3.1. Conditional Quasi-likelihood

Assuming that Z is a spatio-temporal sequence of i.i.d. standard normally distributed variables opens the door to Gaussian quasi-likelihood estimation. If the distribution of Z is truly standard normal, what we derive here will be the true conditional likelihood. If not, we call it quasi-likelihood.

Let $\mathbf{W}_k \stackrel{\text{def}}{=} (\mathbf{X}_k, \dots, \mathbf{X}_{-p+1}, \boldsymbol{\sigma}_0^2, \dots, \boldsymbol{\sigma}_{-q+1}^2)$, $k = 0, \dots, n$. The density of $\mathbb{X}_n \stackrel{\text{def}}{=} [\mathbf{X}_1 \dots \mathbf{X}_n]$ conditional on $\mathbf{W}_0$, can be written as

$$f_{\mathbb{X}_n|\mathbf{W}_0}(\mathbb{x}_n|\mathbf{w}_0) = f_{\mathbf{X}_1|\mathbf{W}_0}(\mathbf{x}_1|\mathbf{w}_0) f_{\mathbf{X}_2|\mathbf{W}_1}(\mathbf{x}_2|\mathbf{w}_1) \dots f_{\mathbf{X}_n|\mathbf{W}_{n-1}}(\mathbf{x}_n|\mathbf{w}_{n-1}).$$

By using the first part of (2.4), we see that

$$f_{\mathbf{X}_k|\mathbf{W}_{k-1}}(\mathbf{x}_k|\mathbf{w}_{k-1}) = \prod_{u\in\mathcal{R}} \frac{1}{\sigma_k(u)}\, f_{\mathbf{Z}}\Big(\frac{\mathbf{x}_k}{\boldsymbol{\sigma}_k}\Big),$$

where the vector division is a Hadamard one. Given $\mathbf{W}_{k-1}$, $\boldsymbol{\sigma}_k$ is successively given from the second part of (2.4).

We get the quasi-likelihood from the conditional simultaneous density with $f_{\mathbf{Z}}$ as the standard normal density and where the empirical likelihood process $\widehat{h}$ replaces σ^2. Taking the logarithm of this likelihood gives

$$\widehat{L}_n(\theta) \stackrel{\text{def}}{=} \sum_{t=1}^{n} \sum_{u\in\mathcal{R}} \widehat{\ell}_t(u,\theta), \quad \widehat{\ell} \stackrel{\text{def}}{=} -\frac{1}{2}\Big\{ \log\widehat{h} + \frac{\widehat{X}^2}{\widehat{h}} \Big\}, \tag{3.9}$$

with the processes $\widehat{X}$ and $\widehat{h}$ given by (3.6).

Remark 3.3. In (3.9) we have denoted $\widehat{\ell}_t(u,\theta)$, $\widehat{h}_t(u,\theta)$ and $\widehat{X}_t(u)$, without their respective space-time locations (t,u) and parameter inputs. This works since the involved processes are simultaneously strictly stationary and the actual expression does not depend on a particular point (t,u). We will use this convention when considering an arbitrary variable, but also when we discuss the processes as one unit, for example, $\widehat{\ell}$, $\widehat{h}$ and $\widehat{X}$.

We use the previously specified $\widehat{h}$ as the empirical counterpart of h. Substituting h for $\widehat{h}$ and X for $\widehat{X}$, defines the theoretical likelihood L_n and its terms ℓ_t. From Proposition 3.1 we have that $\|\widehat{h}_t(u) - h_t(u)\|_\Theta$ converges e.a.s. to zero. Therefore, $\widehat{L}_n$ is usable as an approximation of the theoretical log likelihood function as the notation indicates. The two likelihoods have the respective maximum likelihood estimators

$$\widetilde{\theta}_n \stackrel{\text{def}}{=} \operatorname*{argmax}_\Theta L_n \quad \text{and} \quad \widehat{\theta}_n \stackrel{\text{def}}{=} \operatorname*{argmax}_\Theta \widehat{L}_n.$$

The distinction between L_n and $\widehat{L}_n$ is important. The difference lies in the h-functions and X's used.

Remark 3.4. Note that the likelihood function in (3.9) is closer to the univariate GARCH likelihood than to the multivariate CCC-GARCH likelihood.

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

4. LARGE SAMPLE PROPERTIES

Here we present asymptotic results for the maximum quasi-likelihood estimator, both consistency and asymptotic normality, under certain regularity conditions. Proofs are found in Section 8.

Let $\mathcal{A}(\theta, z) \overset{\text{def}}{=} \sum_{i=1}^{p} \mathbb{A}_i z^i$ and $\mathcal{B}(\theta, z) \overset{\text{def}}{=} \mathbb{I} - \sum_{j=1}^{q} \mathbb{B}_j z^j$ be the matrix polynomials associated with the model. Then $\mathcal{A}$ and $\mathcal{B}$ are left coprime if any matrix polynomial factorizations $\mathcal{A}(\theta_1, \cdot) = \mathcal{U}(\cdot)\mathcal{A}(\theta_2, \cdot)$ and $\mathcal{B}(\theta_1, \cdot) = \mathcal{U}(\cdot)\mathcal{B}(\theta_2, \cdot)$ with $\theta_1, \theta_2 \in \Theta$, imply that $\mathcal{U}$ is unimodular, that is, that the determinant of $\mathcal{U}$ is a non-zero constant.

The following list of assumptions will be used.

A6: The projections $\mathbb{\Delta}_{ki} \rightsquigarrow \mathbb{\Lambda}_{ki}$ into $\mathcal{R}$ preserve cardinality.
A7: At θ_0, $\mathcal{A}$ and $\mathcal{B}$ are left coprime and either $[\mathbb{A}_p \mid \mathbb{B}_q]$ or $[\mathbb{A}'_p \mid \mathbb{B}'_q]$ has full rank.
A8: The squared residual process Z^2 is non-degenerate with expectation 1.
A9: The interior of Θ as a subset of the Euclidean space contains θ_0.
A10: The variance of the squared residual process, $\tau_Z \overset{\text{def}}{=} \operatorname{Var} Z^2$, is finite.

For consistency we need A2–A8, while in addition A9 and A10 are needed for the central limit theorem.

These assumptions are similar to those of the univariate GARCH literature, especially Berkes *et al.* (2003), Francq and Zakoïan (2004) and Straumann and Mikosch (2006). Our assumptions are also related to Francq and Zakoïan (2011, Ch. 11) for the CCC-GARCH, but the comparison to univariate theory is more appropriate (see Remark 3.4) perhaps with the exception of identifiability.

The conditions A2 and A3 ensure stationarity and ergodicity of the process, while A4 and A5 ensure the global existence of the likelihood process on Θ. The compactness of Θ in A5 is also needed for the maximization of the likelihood. The identifiability of the model follows from A6 and A7. As discussed in Remark 2.2, the dimension of the parameter space will necessarily be reduced unless A6 is satisfied. The coprime part of assumption A7 is not enough to guarantee identifiability. Identifiability means here that $\mathcal{B}^{-1}(\theta, z)\mathcal{A}(\theta, z) = \mathcal{B}^{-1}(\theta_0, z)\mathcal{A}(\theta_0, z)$ does not have any solution in Θ except $\theta \equiv \theta_0$. With $\mathcal{A}$ and $\mathcal{B}$ coprime, additional conditions must ensure that the only possible unimodular matrix $\mathcal{U}$, as a common factor, is the identity matrix. The full rank assumption is easily confirmed after estimation, but can in some cases also be confirmed a priori. A sufficient condition is that for any pair of corresponding eigenvalues of the two circulant matrices $\mathbb{A}_p$ and $\mathbb{B}_q$, at least one of them is nonzero. The condition for identifiability of CCC-GARCH models corresponds to A7 (Francq and Zakoïan, 2011, p. 295), but it is quite possible that A7 can be simplified since our model is not really multivariate. The non-degeneracy in A8 is clearly necessary. Otherwise, X^2 will be constant and the model degenerates.

Condition A9 is also necessary, because if for instance $\alpha_1(v) = 0$ for some $v \in \mathbb{\Delta}_{11}$, $N^{1/2}(\widehat{\alpha}_1(v) - \alpha_1(v))$ with $N \overset{\text{def}}{=} N(n) = nm$, cannot be normal with zero expectation because it can only take non-negative values. Berkes *et al.* (2003) assumed A9 for consistency as well, but like Francq and Zakoïan (2004, 2011) and Straumann and Mikosch (2006), we avoid that. A10 is necessary for the finiteness of the asymptotic covariance matrix.

Remark 4.1. The condition A6 means that each $\mathbb{\Delta}_{ki}$ can be contained in $[\mathbf{0}, \mathbf{m} - \mathbf{1}]$ by a translation.

With the assumptions above, we can present the asymptotic results. The first being consistency and the second asymptotic normality of the QMLE. Convergence in distribution of X_n to X is written as $X_n \Rightarrow X$.

Theorem 4.1. Under the assumptions A2–A8, $\widehat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$.

Theorem 4.2. Under the assumptions A2–A10, $\widehat{\theta}_n$ is strongly consistent and asymptotic normally distributed: $N^{1/2}(\widehat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(\mathbf{0}, \kappa\, \mathbb{I}_0^{-1})$, where $\kappa = 2^{-1}\tau_Z$ and $\mathbb{I}_0$ is the information matrix given by

$$\mathbb{I}_0 = 2^{-1}\mathbb{E}(\nabla \log h_0)(\nabla \log h_0)', \qquad h_0 = h(\theta_0). \tag{4.1}$$

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

Remark 4.2. When the residual process is Gaussian, the constant $\kappa = 1$ and $\mathbb{I}_0$ is the Fisher information matrix. Note that $\mathbb{I}_0$ depends only on the marginal distribution of $(h_0, \nabla h_0)$ and thus it is influenced by the squared residual distribution beyond its first two moments.

Although it is not quite clear from the expression, the matrix $\mathbb{I}_0$ depends on $\mathcal{R}$, but the model parameters do not, though the model does.

Remark 4.3. Under A2–A10, $\widehat{\theta}_n$ will eventually satisfy the likelihood equations $\nabla \widehat{L}_n = \mathbf{0}$, since $\widehat{\theta}_n$ is consistent for the interior point θ_0 and $\widehat{L}_n$ is smooth and takes its maximum in an open ball around θ_0. The same relation is of course true for $\widetilde{\theta}_n$ and ∇L_n.

5. SIMULATION EXPERIMENTS

We conduct a simulation experiment to see how well the estimation procedures described in previous sections perform on finite samples. We will both consider circular and non-circular data.

Based on circular data, we should get consistent and asymptotic normally distributed estimates, $\widehat{\theta}_n$, according to Theorems 4.1 and 4.2. On non-circular data however, the circular estimator will be biased due to the projection of the neighbourhoods, but this bias may be compensated by a parametric bootstrap bias correction (PBBC). For this procedure to be successful, we need that $\mathbb{E}_{\theta_0}(\widehat{\theta}_n - \theta_0) \approx \mathbb{E}_{\widehat{\theta}_n}(\overline{\theta}_* - \widehat{\theta}_n)$, where $\overline{\theta}_*$ denotes the mean estimate of the parametric bootstrap samples.

The parameters are estimated by maximizing the likelihood $\widehat{L}_n$. The case $d = 1$ is not so interesting since we know that the circular approximation will perform well. We therefore focus on $d = 2$. The model we consider is the same as in Example 1, with $a_0 = a_1 = \alpha$ and $b_0 = b_1 = b_2 = \beta$. Thus, we only have three parameters, ω, α and β, giving equal weight to all nine members of the neighbourhood. The true parameter vector that generates all the datasets is $\theta_0 = (0.31, 0.024, 0.070)$ and we can write the models in the vector notation of (2.4) with $\mathbb{A}_1 = \alpha\mathbb{W}$ and $\mathbb{B}_1 = \beta\mathbb{W}$, as

$$\mathbf{X}_t = \boldsymbol{\sigma}_t \circ \mathbf{Z}_t, \qquad \boldsymbol{\sigma}_t^2 = \omega + \alpha\mathbb{W}\mathbf{X}_{t-1}^2 + \beta\mathbb{W}\boldsymbol{\sigma}_{t-1}^2,$$

where $\mathbb{W}[u, v] = 1(v \in \mathcal{N}(u))$ is a neighbourhood matrix. The independent innovations are sampled from a standard normal distribution. We use 200 bootstrap replicates in the bias correction, 500 Monte Carlo repetitions and three different sample sizes. The performance is primarily evaluated by the mean square error (MSE) of the estimates, but we also report the bias, standard deviation, the squared bias divided by the variance, marginal and simultaneous coverage. MSE can be decomposed into the sum of squared bias and variance. The squared bias to variance ratio is therefore useful to better understand what drives the development in MSE. Coverage means the proportion of the 500 Monte Carlo simulations where the 95% confidence interval based on the parameter estimate contains the true parameter. The simultaneous coverage is found by checking the condition

$$(\widehat{\theta}_n - \theta_0)' \, \widehat{\mathbb{\Sigma}}^{-1} \, (\widehat{\theta}_n - \theta_0) \leq \chi^2_{0.95}(3) \approx 7.81,$$

where $\widehat{\mathbb{\Sigma}}$ is the Monte Carlo approximated covariance matrix of $\widehat{\theta}_n$, while the marginal checks whether $|\widehat{\theta}_{n,i} - \theta_{0,i}| / \widehat{\mathrm{SD}}(\widehat{\theta}_{n,i}) \leq z_{0.975} \approx 1.96$ for every parameter estimator $\widehat{\theta}_{n,i}$, where $\widehat{\mathrm{SD}}(\widehat{\theta}_{n,i})$ is the square root of the diagonal elements of $\widehat{\mathbb{\Sigma}}$. For comparison purposes, we use Monte Carlo estimates for the covariance matrices. We could use an approximation of the information matrix in (4.1), but this will only be correct in the circular case where the Monte Carlo estimate and the information matrix approximation give similar results. We will refer to circular estimates of circular data by CC, circular estimates of non-circular data by CNC and the parametric bootstrap bias corrected CNC by PBBC.

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

Table I. Estimation results based on 500 Monte Carlo simulations of both circular and non-circular processes of different spatial dimension. [Color figure can be viewed at wileyonlinelibrary.com]

		Circular data			Non-circular data					
		Circular estimates			Circular estimates			PBBC		
Dimension	Parameter	ω	α	β	ω	α	β	ω	α	β
	$100\,\theta_0$	31.00	2.400	7.000	31.00	2.400	7.000	31.00	2.400	7.000
	(Scale MSE)	(10^{-3})	(10^{-5})	(10^{-5})	(10^{-3})	(10^{-5})	(10^{-5})	(10^{-3})	(10^{-5})	(10^{-5})
$5\times5\times3000$	$100\,\text{BIAS}(\widehat{\theta})$	1.064	0.012	−0.070	12.63	−0.489	−0.209	−0.270	−0.075	0.089
	$100\,\text{SD}(\widehat{\theta})$	4.289	0.121	0.299	6.254	0.119	0.392	7.070	0.147	0.463
	MSE	1.949	0.147	0.944	19.84	2.529	1.970	4.995	0.271	2.219
	$\text{BIAS}^2/\text{SD}^2$	0.062	0.010	0.056	4.084	16.98	0.283	0.001	0.260	0.037
	Marg. Cov	0.942	0.952	0.942	0.488	0.012	0.916	0.942	0.932	0.946
	Simult. Cov.		0.942			0.000			0.936	
	Uncond. SD	1.419 (9.318×10^{-3})			1.419 (6.350×10^{-3})			1.419 (6.441×10^{-3})		
$10\times10\times3000$	$100\,\text{BIAS}(\widehat{\theta})$	0.871	0.007	−0.055	2.947	−0.260	0.096	−0.432	−0.029	0.052
	$100\,\text{SD}(\widehat{\theta})$	2.909	0.060	0.190	3.097	0.060	0.199	3.152	0.069	0.211
	MSE	0.920	0.036	0.391	1.826	0.710	0.488	1.010	0.055	0.470
	$\text{BIAS}^2/\text{SD}^2$	0.090	0.013	0.085	0.907	18.44	0.235	0.019	0.178	0.062
	Marg. Cov	0.938	0.950	0.948	0.852	0.008	0.926	0.954	0.940	0.958
	Simult. Cov		0.960			0.006			0.932	
	Uncond. SD	1.419 (4.326×10^{-3})			1.419 (3.780×10^{-3})			1.419 (3.831×10^{-3})		
$15\times15\times3000$	$100\,\text{BIAS}(\widehat{\theta})$	0.775	0.005	−0.048	2.125	−0.173	0.057	−0.109	−0.014	0.020
	$100\,\text{SD}(\widehat{\theta})$	1.971	0.038	0.127	2.083	0.040	0.131	2.080	0.043	0.135
	MSE	0.448	0.015	0.184	0.885	0.315	0.203	0.433	0.021	0.185
	$\text{BIAS}^2/\text{SD}^2$	0.155	0.016	0.141	1.042	19.16	0.187	0.003	0.101	0.023
	Marg. Cov.	0.928	0.940	0.936	0.822	0.012	0.934	0.956	0.932	0.954
	Simult. Cov		0.940			0.002			0.946	
	Uncond. SD	1.419 (2.896×10^{-3})			1.419 (2.648×10^{-3})			1.419 (2.661×10^{-3})		

The scales in the fifth row are for the mean square error (MSE). The estimated unconditional standard deviation (uncond. SD) is given with its standard deviation in parenthesis.

In Table I, the blue and red numbers are the most and least optimal values respectively, for each parameter (ω, α, β) in each row. This colour coding makes it quite clear that the circular model is best on circular data, which should be no surprise to anyone. As expected the absolute bias, the standard deviation and the MSE goes down as the spatial sample size increases in almost all cases. The exception is ω in the PBBC, where the bias increases slightly from spatial dimension 5×5 to 10×10. The standard deviations of the circular estimates are about the same based on circular and non-circular data, and higher for the PBBC estimates. For non-circular data it is interesting to compare the pre- and post PBBC estimates. In all cases, the PBBC successfully reduces the bias of the original estimate extensively. With the exception of β in the smallest sample size, this leads to a smaller MSE. For CC, the squared bias goes down slower than the variance for ω and β, while for α the bias is practically zero in all cases. For the CNC estimates, the squared bias goes down faster than the variance for ω, while for α the small standard deviation and the large bias is inflating the ratio evidently. For β the relationship between squared bias and variance is quite stable. In the PBBC case, we find all the blue numbers for ω and β with values close to zero. For α we see a reduction in the squared bias to variance relation as the sample size increase. For the coverage, we are quite pleased with getting results around 95% for the CC and PBBC estimates.

The kernel density estimates in Figure 4 are green for the CC estimator, blue for CNC and orange for the bias corrected estimator. The red dashed vertical lines indicate the correct value of each parameter. By visual inspection, the circular estimator on circular data is centred around θ_0 with decreasing spread as spatial sample size increases. For ω, it seems that the CNC is moving towards the correct value. One may think that the constant term should not be influenced by the circular assumption, but due to strong correlation with the other estimators it is. In fact, the unconditional standard deviation exists, since (2.12) is fulfilled under θ_0, that is, $9(\alpha+\beta) = 0.846 < 1$, and $\mathbb{E}(\sigma_t^2)^{1/2} = (\omega/(1-9\alpha-9\beta))^{1/2}$ is 1.419 under θ_0. The table shows that this quantity is, to some surprise, preserved in all different simulations with high accuracy. For the α column in Figure 4, we see why the coverage in Table I

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

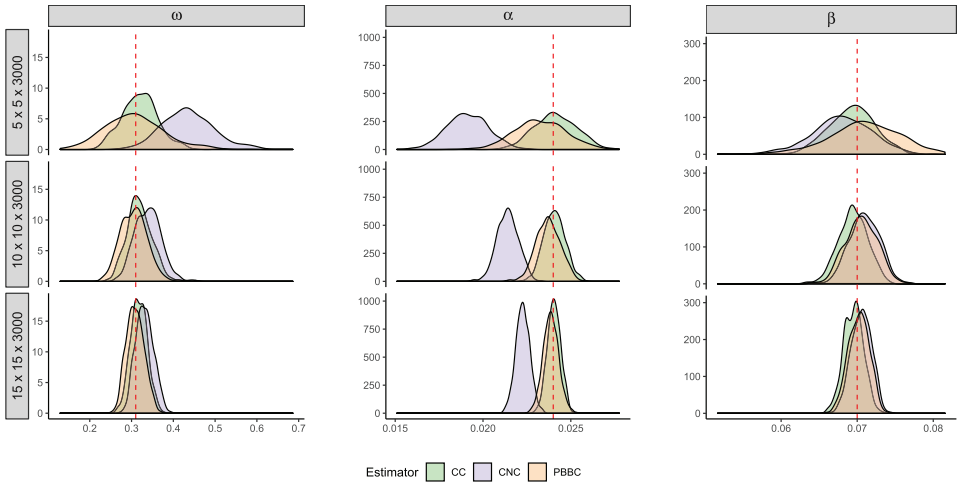

Figure 4. Kernel density estimates based on 500 Monte Carlo simulations with increasing spatial dimension on the vertical and $\widehat{\theta}_n$ on the horizontal axes. In each frame there are densities of three variables: the circular estimator on circular data (CC), circular estimator on non-circular data (CNC) and the parametric bootstrap bias corrected CNC (PBBC). The red dashed vertical lines are the true values. [Color figure can be viewed at wileyonlinelibrary.com]

is so low for the CNC. The bell curve does not even touch the red line, although it seems to be moving towards it with increasing sample sizes. The effect of the bias correction is largest in this column, where the improvement in location is substantial and the cost in terms of higher variance is definitely worth its price. The β column is not as influenced by the circular assumption. Especially for the largest sample, there is very little difference pre- and post PBBC.

This experiment shows that the estimation of circular models is meaningful for circular and non-circular data, but in the latter case it is important to include a parametric bootstrap bias correction step if you want to reproduce the non-circular parameters.

6. REAL DATA EXAMPLE: SEA SURFACE TEMPERATURE ANOMALIES

The data we consider has been studied in great detail by Berliner *et al.* (2000) and used as an example by Wikle and Hooten (2010), Cressie and Wikle (2011), Wikle and Holan (2011) amongst others. These are monthly averaged sea surface temperature (SST) anomalies dating from January 1970 to March 2003, measured on a 2°×2° resolution grid in the tropical Pacific Ocean. This is not intended to be a comprehensive example, but merely an illustration of how one can implement CSTGARCH on real non-circular data. In this regard, the model is mainly descriptive here. However, the GARCH part could be used to improve prediction intervals by volatility forecasting. The data is available for download as supplementary material to the book by Cressie and Wikle (2011) (link at the end).

We choose a rectangle area without observations over land (see Figure 5) and reduce the data to a $4° \times 4°$ grid by mean aggregation. This reduction in sample size is to make the computations less demanding. Finally, we spatially difference the data to get a stationary series. That is, let $\{Y_t(u_1, u_2)\}$ denote the observations, then

$$\nabla_1\nabla_2 Y_t(u_1, u_2) = Y_t(u_1, u_2) - Y_t(u_1, u_2 - 1) - Y_t(u_1 - 1, u_2) + Y_t(u_1 - 1, u_2 - 1),$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

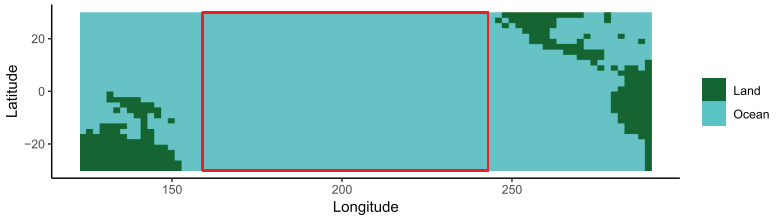

Figure 5. Area of interest: Rectangle without land observations. Longitudes should normally be between −180 and 180, but we have extended the scale to make it easier to visualize. [Color figure can be viewed at wileyonlinelibrary.com]

Table II. Parameter estimates of the CSTARMA model in (6.1) and of the CSTGARCH model for the SST anomalies dataset.

	CSTARMA			CSTGARCH	
	Estimates	SD		Estimates	SD
$\phi_{1,0}$	0.361	2.87×10^{-3}	ω	5.06×10^{-4}	8.02×10^{-5}
$\phi_{1,2}$	0.225	1.28×10^{-2}	α_0	6.25×10^{-2}	3.91×10^{-3}
$\phi_{5,0}$	0.026	2.81×10^{-3}	α_1	6.57×10^{-2}	3.82×10^{-3}
$\phi_{11,0}$	0.360	9.89×10^{-3}	α_2	3.68×10^{-2}	3.62×10^{-3}
$\phi_{12,1}$	−0.248	2.04×10^{-2}	α_3	4.05×10^{-2}	4.23×10^{-3}
$\theta_{2,1}$	0.163	9.28×10^{-3}	β_0	2.66×10^{-1}	7.4×10^{-2}
$\theta_{3,1}$	0.143	9.29×10^{-3}	β_1	5.09×10^{-1}	6.94×10^{-2}
$\theta_{11,0}$	−0.378	1.06×10^{-2}			
$\theta_{12,1}$	0.205	2.29×10^{-2}			

where ∇_1 and ∇_2 denote the spatial difference operators in the two spatial dimensions respectively. The data we are left with is of dimension $20 \times 14 \times 399$ (longitude × latitude × time) or 111720 data points, with longitudes from 165°E to 241°E (119°W) and latitudes from 24°S to 28°N.

The differenced data is correlated and shows signs of a 12 month season (see Figure 7b), so we fit a circular spatio-temporal ARMA (CSTARMA) model using the `starma` R-package (Cheysson, 2016). The CSTARMA model is, by adapting the definition of Pfeifer and Deutsch (1980),

$$\mathbf{Y}_t = \sum_{j=1}^{12} \sum_{k=0}^{2} \phi_{jk} \mathbb{W}^{(k)} \mathbf{Y}_{t-j} + \sum_{j=1}^{12} \sum_{k=0}^{2} \theta_{jk} \mathbb{W}^{(k)} \mathbf{X}_{t-j} + \mathbf{X}_t, \tag{6.1}$$

where $\mathbb{W}^{(k)}$ is a circular queen-contiguity neighbourhood matrix, characterized by the movement patterns of a chess queen, with k as the spatial lag. The construction of such a circular neighbourhood matrix for a regular grid is implemented in the `spdep` package (Bivand and Wong, 2018). Here, the spatial lag refers to how many tiles the chess queen moves over and is illustrated in Figure 6a. We use backward-stepwise model selection, that is, we start with temporal order 12 and spatial order 2 and fit the model (6.1) with $(12 \times 2 \times 3 =)72$ parameters. Then, gradually remove the least significant one and estimate the model again, until only one parameter remains. We then have 72 model candidates and choose the one that minimize Akaike's information criterion (AIC). This model, which also minimize BIC, has only nine parameters and the largest p-value is of order 10^{-19}. The nine parameter estimates with their corresponding standard deviations are given in the left column of Table II.

The residuals of the CSTARMA model is modelled using a CSTGARCH$(1, 1)$ model with one spatial lag, corresponding to the one used in Example 1. We allow for different parameters in the different directions and it turns out that for the GARCH part, a rook contiguity neighbourhood only in the longitudinal direction is sufficient,

DOI: 10.1111/jtsa.12498

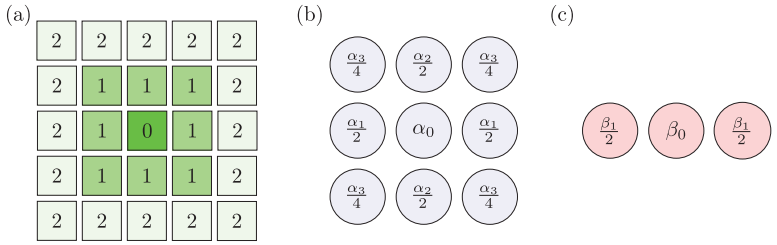

Figure 6. (a) This illustrates how the spatial lags are defined for a queen contiguity neighbourhood, used for the CSTARMA model and autocorrelation plot in Figure 7b. Parameter specification and dependency structure at one time lag for the (b) ARCH- and (c) GARCH parts of the CSTGARCH model applied to the SST anomalies data. The horizontal and vertical directions correspond to longitudinal and latitudinal respectively. [Color figure can be viewed at wileyonlinelibrary.com]

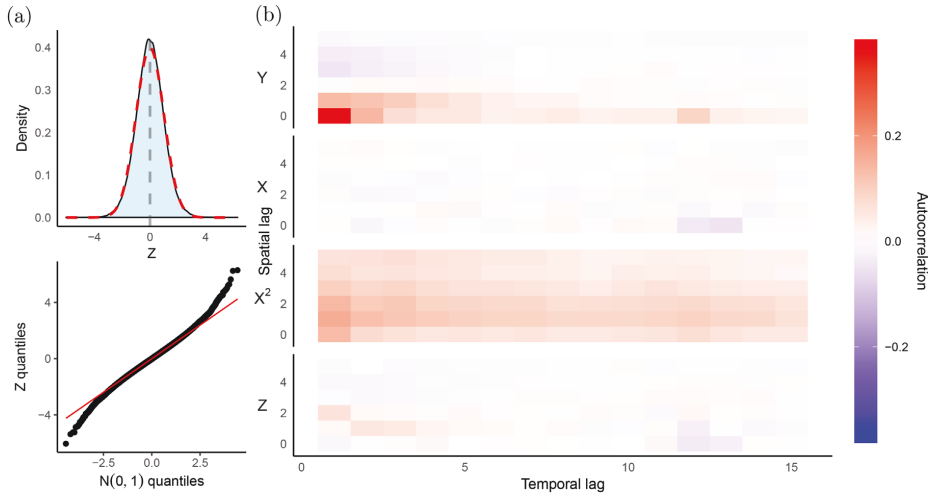

Figure 7. (a) Marginal kernel density estimate for the standardized residuals, Z, along with a standard normal density in red. Below it is a QQ-normality plot of Z. (b) Sample autocorrelation functions of the processes Y, X, X^2 and Z. The spatial lag is given in Figure 6a. The 12-month seasonal effect is seen as the stronger colours around temporal lag 12. [Color figure can be viewed at wileyonlinelibrary.com]

while for the ARCH part a queen neighbourhood fits best. The model specification we end up choosing is visualized in Figure 6(b,c), corresponding to the illustration used in Figure 2.

The CSTGARCH parameter estimates are presented in the right column of Table II. The empirical variance of the data is ≈ 0.025, which fits well with $\widehat{\omega}/\{1 - \sum_j(\widehat{\alpha}_j + \widehat{\beta}_j)\} \approx 0.027$. Plotting four-dimensional processes is difficult, and animations of all the fitted spatio-temporal series Y, $X = Y - \widehat{Y}$, $\widehat{h}$ and $Z = X/\widehat{h}$ (cf. Remark 3.3), are therefore given as supplementary material (details at the end). In these animations, notice how the fitted conditional volatility clusters in space and time. The spatio-temporal autocorrelation of Y, X, X^2 and Z are plotted in Figure 7b. Here we see that the original data and the squared fitted GARCH process are correlated with points close in space and time, while the GARCH process and the standardized residuals are not. This relates to one of the

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

stylized facts about GARCH processes: Little autocorrelation in the process itself, yet profound correlation in the squared process. When comparing autocorrelation of X^2 to that of Z, it seems the model fits well. The empirical distribution of the standardized residuals is close to the standard normal (see Figure 7a). We have not performed any correlation tests here, but since the sample size is so large, we do not expect to reject a hypothesis of correlated residuals.

7. CONCLUDING REMARKS

Finite area models are common in spatial statistics and wrapping a spatial model onto a toroidal surface is a known strategy for forming spatially stationary models in this field. The construction of a circular model is not limited to GARCH models and can be applied to other spatial- and spatio-temporal models. The circular assumption is fruitful for calculations, simulation and asymptotic theory. It is also crucial for having an explicit efficient likelihood with a fixed temporal boundary not depending on the sample size.

We consider CSTGARCH as a model in its own right, but it can definitely be used as an approximation of a non-circular situation as well. Estimating an STGARCH model will lead to a boundary issue due to unobserved sites outside the area of interest, and our simulation experiment indicates that a bias corrected circular approximation is a viable alternative. The upside to the circular method is a utilization of all data points with a complete disappearance of spatial boundaries. The circular processes are Markov, which is not the case for other ways of approximating an STGARCH. The downside is its misspecification, mostly accentuated near the boundaries, if the true model is not circular.

The boundary problem of spatial and spatio-temporal processes is well known and different approaches have been proposed. Guyon (1982) showed that the edge effect goes to zero like $(nm)^{-1/(d+1)}$ in our notation. It is therefore necessary to handle the edge effect in a proper way. For dealing with the boundary issue in spatial processes, Cressie (2015, p. 422) mentioned integrating out the unobserved data from the conditioning event, but warned that this might lead to a complicated likelihood. We clearly find his warning justified for the STGARCH model. Another suggestion of his is to form a *guard area* inside the perimeter of the area of interest, where observations contribute to the likelihood only through their neighbourhood relations with internal sites. The volatility process must be estimated within the guard area, which means the guard area has to be quite wide and the practitioner must set boundary values for the volatility at every time point. For a pure STARCH model, the guard area approach is an alternative and a possible next step is to approximate an STGARCH with an STARCH. However, this procedure will inevitably lead to biased estimates and a sacrifice of a significant proportion of the observations, depending on the sample size. Estimation of GARCH models is infamous for requiring large samples and we cannot afford losing too much data. The circular model leaves none behind.

8. PROOFS

The main structure of the proofs for Theorems 4.1 and 4.2 is an adaptation of established theory. In particular we are influenced by Straumann and Mikosch (2006) and Francq and Zakoïan (2004). There are also some new elements.

The circular projection, as we have seen, makes an STGARCH model into a CCC-GARCH with a specific circular neighbourhood structure (Theorem 2.1). Asymptotic theory exists for CCC-GARCH models (Ling and McAleer (2003), Francq and Zakoïan (2011, pp. 289–307)), but the stationary spatio-temporal context here is quite far from that framework. The following proofs of consistency and asymptotic normality are adapted to the current context and therefore considerably simplified in comparison to the multivariate case. The proofs are relevant for further work on GARCH models in the space-time domain, with and without the circular assumption. In particular, in a working paper by Karlsen and Hølleland (2019), these proofs are used as a framework for the likelihood estimation.

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

8.1. Proof of Consistency

Throughout this subsection, we assume that A2–A8 hold.

The subscript 0 refers to θ_0 for quantities depending on θ, for example, $h_0 = h(\theta_0)$. An arbitrary, but fixed point (t, u) is suppressed whenever that is convenient to do (cf. Remark 3.3). We use $\varrho_\omega = \min_\Theta \omega(\theta)$ so that $h \geq \varrho_\omega$. The total number of observations is $N = nm$ with n, m as the number of time- and spatial points respectively.

8.1.1. Identifiability

Proposition 8.1. The model is identifiable; $h \equiv h_0$ a.s. if and only if $\theta = \theta_0$.

Proof. Due to A6, the VARMA form of (3.1) is fully parametrized and therefore possibly identifiable. The conditions are therefore in terms of the associated matrix polynomials $\mathcal{A}$ and $\mathcal{B}$. By A4, $\mathcal{B}$ is causal and A7 states the coprime property. Since $\mathbb{A}_p$ and $\mathbb{B}_q$ are circulants of level d (Theorem 2.2) and simultaneously diagonalizable (Davis, 1994, Thm. 5.8.4), the two alternative conditions in A7 are equivalent, and the full rank requirement on $[\mathbb{A}_p, \mathbb{B}_q]$ is satisfied. These three properties guaranties identifiability (Reinsel, 2003, p. 37). □

8.1.2. The Asymptotic Likelihood has a Unique Global Maximum

Proposition 8.2. The asymptotic likelihood,

$$L \stackrel{\text{def}}{=} \mathbb{E}\ell = -2^{-1}\mathbb{E}\Big(\log h + \frac{X^2}{h}\Big),$$

has a unique global maximum at θ_0, that is, $L(\theta) < L(\theta_0)$ for $\theta \in \Theta\backslash\{\theta_0\}$.

Proof. The proof is an adapted version of an argument by Straumann and Mikosch (2006). Let

$$Q \stackrel{\text{def}}{=} 2L + \mathbb{E}\log h_0 = \mathbb{E}(\log h_0 + 2\ell) = \mathbb{E}\bigg(\log h_0 - \log h - \frac{X^2}{h}\bigg) = \mathbb{E}\bigg(\log\frac{h_0}{h} - \frac{h_0}{h}\bigg),$$

since $X^2 = h_0 Z^2$. Now, $\log x - x < -1$ unless $x = 1$. This means that $Q(\theta) < Q(\theta_0)$ with equality if and only if $h = h_0$ a.s. □

8.1.3. The Observable and Non-observable Likelihood are Equivalent

Proposition 8.3.

$$\|\widehat{L}_n - L_n\|_\Theta = o(1) \text{ e.a.s.} \tag{8.1}$$

Proof. By the mean value theorem $|\log y - \log x| \leq \min^{-1}(y, x)$ and

$$\begin{aligned} |\widehat{\ell} - \ell| &\leq |\log\widehat{h} - \log h| + X^2(\widehat{h}h)^{-1}|\widehat{h} - h|, \\ \|\widehat{\ell} - \ell\|_\Theta &\leq (\varrho_{\boldsymbol{\omega}} \vee 1)^{-2}(1 + X^2)\|\widehat{h} - h\|_\Theta, \end{aligned}$$

so that

$$\|\widehat{L}_n - L_n\|_\Theta \leq \sum_{t=1}^{n}\sum_{u\in\mathcal{R}}(\varrho_{\boldsymbol{\omega}} \vee 1)^{-1}(1 + X_t^2(u))\|\widehat{h}_t(u) - h_t(u)\|_\Theta.$$

By Proposition 3.1 and Theorem 2.3 the conditions in Lemma 3.1 hold and (8.1) follows. □

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

8.1.4. Consistency by the Upper Semicontinuous Framework
The following important result is a device to face the possibility of non-uniform convergence.

Theorem 8.1 (Pfanzagl, 1969, Lemma 3.11)**.** Let $\Theta \subset \mathbb{R}^k$ be compact and let Θ' denote an arbitrary compact subset of Θ, T be an ergodic transformation on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\xi_0(\theta)$ be a stochastic variable in $[-\infty, \infty]$ for all $\theta \in \Theta$. Define $\xi_t(\theta) = \xi_t(\omega, \theta) = \xi_0(T^t\omega, \theta)$ for $t \geq 1$ and $\omega \in \Omega$.

Assume that

(i) $\xi_0(\cdot)$ is upper semicontinuous (usc) with probability one.
(ii) For all $\Theta' \subseteq \Theta$: $\|\xi_0\|_{\Theta'}$ is a stochastic variable.
(iii) For all $\Theta' \subseteq \Theta$: $\mathbb{E}\|\xi_0^+\|_{\Theta'}$ is finite.

Let $\mu = \mathbb{E}\xi_0$ and $\overline{\xi}_n = n^{-1}\sum_{t=1}^n \xi_t$. Then

(iv) $\overline{\xi}_n \xrightarrow[\text{n}]{\text{a.s.}} \mu$.

(v) For all $\Theta' \subseteq \Theta$: $\overline{\lim}\|\overline{\xi}_n\|_{\Theta'} \leq \|\mu\|_{\Theta'}$.

Proof of 4.1. Let $\xi_t = m^{-1}\sum_{u\in\mathcal{R}} \ell_t(u)$. Then $\{\xi_t\}$ is ergodic and (i) and (ii) of Lemma 8.1 is satisfied, since ℓ is continuous. For (iii) $\ell^+ \leq -\log h \leq -\log \varrho_\omega < \infty$. From (iv), we get $\overline{L}_n = \overline{\xi}_n = \mu + o(1) = L + o(1)$ a.s. and (v) $\overline{\lim}\|\overline{L}_n\|_{\Theta'} \leq \|L\|_{\Theta'}$. Since $\widehat{L}_n = L_n + (\widehat{L}_n - L_n)$ we get from Lemma 8.3 that $\{\widehat{L}_n\}$ also satisfies (iv) and (v). With (iv) and (v) and Lemmas 8.1–8.3 at hand, the arguments presented by for instance Ferguson (1996, pp. 114–5) do the rest for us. □

8.2. Proof of the CLT

Throughout this subsection, we assume A2–A10.

Let $S_\theta(\kappa)$ be an open ball with centre in $\theta \in \Theta$ and radius $\kappa > 0$. When A9 holds, let $S_0 = S_0(\kappa) = S_{\theta_0}(\kappa)$ be contained in Θ for some $\kappa > 0$ and let $\overline{S}_0$ be the closure of this open ball. We use γ both as a component of the parameter vector θ and its reference index. As an index, we write $\gamma \in \theta$ and for multiple indexes $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k) \in \theta^k$. The following properties of the model inherent in the assumptions are important input for the CLT proof.

(i) There exist a closed ball with center θ_0 in the parameter space Θ. Without loss of generality we can assume that the radius is one; $\overline{S}_0(1) \subseteq \Theta$.
(ii) When we consider a fixed outcome outside a fixed null set, then both $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ converges to θ_0. Hence for any $\kappa > 0$, both these estimators are inside $S_0(\kappa)$ and satisfies their respective log likelihood equation for all n large enough.
(iii) The components of Θ are uniformly bounded in both directions; $\varrho \stackrel{\text{def}}{=} \|\theta\|_{S_0} \vee \|1/\theta\|_{S_0}$ so that $0 < \varrho^{-1} \leq \gamma \leq \varrho$ for any component $\gamma \in \theta$ on S_0.
(iv) The elements of $\mathbb{B}^k$, $b^{(k)}$, decrease with an exponential rate to zero as k goes to infinity; $b^{(k)} \leq \rho^{k-q}$ with $\rho = \rho_{\mathbb{B}}^{1/q}$, where $\rho_{\mathbb{B}}$ is the spectral radius of $\mathbb{B}$.

The proof goes along relatively standard lines. The starting point is a Taylor expansion of the theoretical log likelihood.

8.2.1. Taylor Expansion
Proposition 8.4.

$$-(\overline{\mathbb{J}}_n + \overline{\mathbb{R}}_n)N^{1/2}(\widetilde{\theta}_n - \theta_0) = N^{-1/2}\nabla L_n(\theta_0), \tag{8.2}$$

where

$$\mathbb{R}_n = \mathbb{R}_n(\widetilde{\theta}_n, \theta_0) \stackrel{\text{def}}{=} \int_0^1 \big(\nabla^2 L_n(\theta_0 + s(\widetilde{\theta}_n - \theta_0)) - \nabla^2 L_n(\theta_0)\big)\mathrm{d}s,$$
$$\mathbb{J}_n = \mathbb{J}_n(\theta_0) \stackrel{\text{def}}{=} -\nabla^2 L_n(\theta_0), \quad \overline{\mathbb{J}}_n = N^{-1}\mathbb{J}_n, \quad \overline{\mathbb{R}}_n = N^{-1}\mathbb{R}_n.$$

Proof. Let $\delta_n = \widetilde{\theta}_n - \theta_0$. We use an integrated mean value theorem for the vector-valued multivariate ∇L_n.

$$\begin{aligned}\nabla L_n(\widetilde{\theta}_n) &= \nabla L_n(\theta_0) + \left[\int_0^1 \nabla^2 L_n(\theta_0 + s\,\delta_n)\mathrm{d}s\right]\delta_n \\ &= \nabla L_n(\theta_0) + \left[\nabla^2 L_n(\theta_0) + \int_0^1 \Big(\nabla^2 L_n(\theta_0 + s\,\delta_n) - \nabla^2 L_n(\theta_0)\Big)\mathrm{d}s\right]\delta_n \\ &= \nabla L_n(\theta_0) + (\mathbb{J}_n + \mathbb{R}_n)\delta_n.\end{aligned}$$

Since $\widetilde{\theta}_n$ satisfies the log likelihood equations, that is, $\nabla L_n(\widetilde{\theta}_n) = 0$, the left-hand side is zero and (8.2) follows by a simple rearrangement and dividing by $N^{1/2}$. □

The main points to show are:

(i) The remainder term $\overline{\mathbb{R}}_n$ can be neglected.
(ii) The observed information matrix, $\overline{\mathbb{J}}_n$, converges to a positive definite matrix.
(iii) The observable estimator $\widehat{\theta}_n$ and its theoretical companion $\widetilde{\theta}_n$ are square root N equivalent.

8.2.2. The Remainder Term
Recall that $h = h_t(u)$ and $\mathbb{B}_{11}^{(k)}$ is the first block of $\mathbb{B}^k$.

Lemma 8.1.

$$h = \sum_{k=0}^{\infty} f_k, \qquad f_k(t,u,\theta) \stackrel{\text{def}}{=} \mathbb{B}_{11}^{(k)}\mathbf{D}_{t-k}(u) \tag{8.3}$$

$$\varrho^{-1} g_k \le f_k \le \varrho\, g_k, \qquad g_k(t,u,\theta) \stackrel{\text{def}}{=} \mathbb{B}_{11}^{(k)}\mathbf{D}_{t-k}(u,\theta_0). \tag{8.4}$$

Proof. From (3.2)–(3.5), we can write

$$h_t(u) = \mathbf{h}_t(u) = \sum_{k=0}^{\infty} \mathbb{B}_{11}^{(k)}\mathbf{D}_{t-k}(u) = \sum_{k=0}^{\infty} f_k(t,u).$$

For the second part,

$$\varrho^{-1}\mathbf{D}_t(\theta_0) \le \mathbf{D}_t(\theta) \le \varrho\,\mathbf{D}_t(\theta_0).$$ □

The main influence on h from the parameter θ goes through $\mathbb{B}$. It is therefore advantageous to neutralize the impacts from the $\mathbf{D}_t$'s. This is the content of the second point in the lemma above.

Lemma 8.2. Let $\mathcal{D}^\gamma = \partial/\partial\gamma$ for $\gamma \in \theta$ and successive first order partial derivatives is denoted by $\mathcal{D}^{\boldsymbol{\gamma}} = \prod_i \mathcal{D}^{\gamma_i}$ with $\boldsymbol{\gamma} = \{\gamma_i\}$. Let $r = \dim(\boldsymbol{\gamma})$ and $k_{(r)} = \prod_{j=0}^{r-1}(k-j) \vee 1$. Then

$$\mathbb{0} \le \mathcal{D}^{\boldsymbol{\gamma}}\mathbb{B}^k \le \varrho^r k_{(r)}\mathbb{B}^k, \qquad 0 \le \mathcal{D}^{\boldsymbol{\gamma}}\mathbf{D}_t(u) \le \varrho^r \mathbf{D}_t(u). \tag{8.5}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

Proof. Each element of $\mathbb{B}^k$ is a multivariate polynomial in the variables of $\{\beta_j(v)\}$ and has at most degree k. In addition, each non-zero term in the polynomial has coefficient 1. Due to A4 the actual values of any of the variables cannot exceed 1. Doing r successive differentiations of an element will at most give the factor $k_{(r)}$ and, at the same time, a size increase due to a reduction of total power of the factors that constitute the term. The size increase is therefore at most ϱ^r. The power reduction is r if and only if the result is greater than zero. The left-hand side is trivial, since each factor in each element is non-negative. The second part of (8.5) is straightforward. □

Lemma 8.3. For any $\boldsymbol{\gamma} \in \theta^r$ with $r \in \mathbb{N}$,

$$0 \le h^{(\gamma)} \le \varrho^{r+1} \sum_{k=0}^{\infty} k_{(r)} g_k.$$

Proof. Let $\boldsymbol{\gamma}$ be fixed. We split this vector so that $\boldsymbol{\gamma}^{(1)}$ is the $\mathbb{B}$-part and $\boldsymbol{\gamma}^{(2)}$ is the remaining part. Corresponding to this structure, we write $D^{(j)}$ for $j = 1, 2$. By (8.3),

$$D^{\gamma} h = \sum_{k=0}^{\infty} D^{\gamma} f_k,$$

and by Lemma 8.2

$$\begin{aligned} D^{\gamma} f_k(t, \cdot) &= D^{\gamma} \mathbb{B}_{11}^{(k)} \mathbf{D}_{t-k} = D^{(1)} \mathbb{B}_{11}^{(k)} D^{(2)} \mathbf{D}_{t-k} \\ &\le \varrho^r k_{(r)} \mathbb{B}_{11}^{(k)} \mathbf{D}_{t-k} \le \varrho^{r+1} k_{(r)} \mathbb{B}_{11}^{(k)} \mathbf{D}_{t-k}(\theta_0) \\ &= \varrho^{r+1} k_{(r)} g_k(t, \cdot). \end{aligned}$$

The lower bound also follows from Lemma 8.2. □

The following definition extends Definition 3.2.

Definition 8.1. Let $V = V(\theta)$ be a stochastic matrix depending on $\theta \in \mathcal{S}$. Then $\|V\|_{\mathcal{S},p} \stackrel{\text{def}}{=} \mathbb{E}^{1/(p \vee 1)} \|V\|_{\mathcal{S}}^p$ for any $p > 0$. If V is independent of θ, we may drop the subscript $\mathcal{S}$.

Remark 8.1. $\|V\|_{\mathcal{S},p}$ fulfils the triangle inequality for $p \in (0, \infty)$.

Lemma 8.4. For any $\boldsymbol{\gamma} \in \theta^r$, with $r \in \mathbb{N}$ and any moment $p \in \mathbb{N}_+$, there exists a $\kappa > 0$ such that, with $\mathcal{S}_0 = \mathcal{S}_0(\kappa)$,

$$\text{i)} \quad \left\| \frac{h_0}{h} \right\|_{\mathcal{S}_0,p} < \infty.$$

$$\text{ii)} \quad \left\| \frac{h^{(\gamma)}}{h} \right\|_{\mathcal{S}_0,p} < \infty.$$

Proof. The proof of the two statements is quite similar. Let p be fixed and choose $\delta > 0$ so that $p\delta > 0$ satisfy Theorem 2.3(iv). Let $\rho = \rho(\mathbb{B}_0)$.

Step 1: Expansion and simplifications of the two fractions.

Let $g_{0k} = g_k(\theta_0) = f_k(\theta_0)$. By (8.4),

$$f_\ell \ge \varrho^{-1} g_\ell \quad \text{and} \quad g_0 = g_{00} \ge \varrho^{-1},$$

and from (8.3), we can write

$$\begin{aligned}\frac{h_0}{h} &= \frac{\sum_{k=0}^{\infty} g_{0k}}{\sum_{\ell=0}^{\infty} f_\ell} \le \frac{g_{00}}{f_0} + \sum_{k=1}^{\infty} \frac{g_{0k}}{f_0 + \sum_{\ell=1}^{\infty} f_\ell} \\ &\le \varrho^2 \left(1 + \sum_{k=1}^{\infty} \frac{g_{0k}}{1 + \sum_{\ell=1}^{\infty} g_\ell}\right) \le \varrho^2 \left(1 + \sum_{k=1}^{\infty} \frac{g_{0k}}{1 + g_k}\right).\end{aligned} \tag{8.6}$$

For (ii) we get from Lemma 8.3,

$$\begin{aligned}\frac{h^\gamma}{h} &\le \frac{\varrho^{r+1} \sum_{k=0}^{\infty} k_{(r)} g_k}{\sum_{\ell=0}^{\infty} f_\ell} \le \varrho^{r+2} \left(1 + \sum_{k=1}^{\infty} \frac{k_{(r)} g_k}{g_0 + \sum_{\ell=1}^{\infty} g_\ell}\right) \\ &\le \varrho^{r+2} \left(1 + \sum_{k=1}^{\infty} \frac{k_{(r)} g_k}{1 + \sum_{\ell=1}^{\infty} g_\ell}\right) \le \varrho^{r+2} \left(1 + \sum_{k=1}^{\infty} \frac{k_{(r)} g_k}{1 + g_k}\right).\end{aligned}$$

Step 2: Part (i)

We start with the first part. Let (t, u) be fixed. Let $g_k(\theta) \stackrel{\text{def}}{=} g_k(t, u, \theta)$, $\mathbb{B}_{11}^{(k)} = \{b_{11}^{(k)}(u, v, \theta)\}$, $V_k(v) \stackrel{\text{def}}{=} \mathbf{D}_{t-k}(v, \theta_0)$ and

$$b_k(v, \theta) \stackrel{\text{def}}{=} b_{11}^{(k)}(u, v, \theta), \quad a_k(v) \stackrel{\text{def}}{=} b_k(v, \theta_0). \tag{8.7}$$

By (8.4) and (8.7),

$$g_k(t, u) = \mathbb{B}_{11}^{(k)} \mathbf{D}_{t-k}(u, \theta_0) = \sum_{v \in \mathcal{R}} b_{11}^{(k)}(u, v) V_k(v) = \sum_{v \in \mathcal{R}} b_k(v) V_k(v). \tag{8.8}$$

Inserting (8.7) and (8.8) into (8.6) gives

$$\sum_{k=1}^{\infty} \frac{g_{0k}}{1 + g_k} = \sum_{k=1}^{\infty} \sum_{v \in \mathcal{R}} \frac{a_k V_k}{1 + b_k V_k}.$$

Let $\kappa > 0$, $\mathcal{T}_0(\kappa) \stackrel{\text{def}}{=} \mathcal{S}_0(\kappa \|\theta_0\|)$ and

$$\mathcal{A}_k(\kappa) \stackrel{\text{def}}{=} \{\theta \in \Theta : (1 - \kappa)^k \mathbb{B}_0^k \le \mathbb{B}^k \le (1 + \kappa)^k \mathbb{B}_0^k\}, \quad k \ge 1.$$

Then by looking at $\mathbb{B}^{k+1} = \mathbb{B}^k \mathbb{B}$, we see that

$$\mathcal{T}_0 \subseteq \cup_k \mathcal{A}_k. \tag{8.9}$$

Choose κ_0 so that $\mathcal{T}_0(\kappa_0) \subset \Theta$. By (8.9), we have for any $\kappa \in (0, \kappa_0]$,

$$(1 - \kappa)^k a_k \le b_k \le (1 + \kappa)^k a_k, \quad a_k \le \rho^{k-q} \quad \text{on } \mathcal{T}_0(\kappa). \tag{8.10}$$

On $\mathcal{T}_0(\kappa)$ and $a_k > 0$, this gives

$$\begin{aligned}\frac{a_k V_k}{1 + b_k V_k} &= \left(\frac{a_k}{b_k}\right)\left(\frac{b_k V_k}{1 + b_k V_k}\right) \le \left(\frac{a_k}{b_k}\right) b_k^\delta V_k^\delta = a_k b_k^{\delta - 1} V_k^\delta \\ &\le \rho^{-q\delta} \left(\frac{\rho^\delta}{(1 - \kappa)^{1-\delta}}\right)^k V_k^\delta = \rho^{-q\delta} \tau^k V_k^\delta, \text{ say,}\end{aligned}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

since $x(1+x)^{-1} \le x^{\delta}$ for $x \ge 0$ and by (8.10). Choose $\kappa < (1-\xi) \wedge \kappa_0$ with $\xi = \rho^{\delta/(1-\delta)}$. Then $\tau < 1$. It is clear that our choice of κ is independent of u and v and $\{b_k > 0\} = \{a_k > 0\}$ on $\mathcal{T}_0(\kappa)$. Thus,

$$\begin{aligned}\varrho^{-2}\left\|\frac{h_0}{h}\right\|_{S_0,p} &\le \sum_{k=1}^{\infty}\sum_{v\in\mathcal{R}}\left\|\frac{a_k V_k}{1+b_k V_k}\right\|_{S_0,p} \le \sum_{k=1}^{\infty}\sum_{v\in\mathcal{R}}\rho^{-q\delta}\tau^k\|\mathbf{D}_{t-k}(v,\theta_0)\|_{S_0,p\delta}\\ &= \left[\rho^{-q\delta}\tau(1-\tau)^{-1}m\right]\|\mathbf{D}_t(v,\theta_0)\|_{p\delta} < \infty,\end{aligned}$$

which ends the proof of part one.

Step 3: The second part.

For (ii), we use that $b_k \le (1+\kappa)a_k$,

$$\frac{k_{(r)}b_k V_k}{1+b_k V_k} \le k_{(r)}b_k^{\delta}V_k^{\delta} \le \rho^{-q\delta}k_{(r)}[(1+\kappa)^{\delta}\rho^{\delta}]^k V_k^{\delta}.$$

This is convergent for $\kappa < (\rho^{-1}-1) \wedge \kappa_0$ and in that case ii) holds. □

Lemma 8.5. For any $\boldsymbol{\gamma}$ and for $\kappa > 0$ sufficiently small, $\|\ell^{(\boldsymbol{\gamma})}\|_{S_0(\kappa),1} < \infty$.

Proof. It is easy to verify that

$$\ell^{(\boldsymbol{\gamma})} = \sum_{k=1}^{M} a_k U_k + \sum_{k=1}^{L} b_k V_k,$$

for appropriate constants where the structure of the U_k's and V_k's can be written as

$$U = \prod_{j=1}^{r}\frac{h^{(\boldsymbol{\xi}_j)}}{h}, \quad V = U'\frac{X^2}{h}, \quad \sum \dim(\boldsymbol{\xi}_j) \le \dim(\boldsymbol{\gamma}),$$

with U and U' in general different, but of the same type.

Now, with $S_0 = S_0(\kappa)$,

$$\begin{aligned}\|U\|_{S_0,p} &\le \max_j\left\|\frac{h^{(\boldsymbol{\xi}_j)}}{h}\right\|_{S_0,rp}, \quad p = 1,2,\\ \|V\|_{S_0,1} &\le \|U\|_{S_0,2}\left\|\frac{h_0}{h}\right\|_{S_0,2}\|Z^2\|_2,\end{aligned}$$

which is finite by Lemma 8.4 and A10. □

Proposition 8.5. $\overline{\mathbb{R}}_n = o(1)$ a.s.

Proof. Let $\delta_n = \widetilde{\theta}_n - \theta_0$. We start by applying an integrated version of the multivariate mean value theorem, this time on vec $\nabla^2 L_n$,

$$\operatorname{vec} \nabla^2 L_n(\theta_0 + s\,\delta_n) = \operatorname{vec} \nabla^2 L_n(\theta_0) + \int_0^1 \Big[\nabla \operatorname{vec} \nabla^2 L_n(\theta_0 + rs\,\delta_n) \mathrm{d}r\Big] s\,\delta_n,$$

$$\operatorname{vec} \overline{\mathbb{R}}_n = \int_0^1 \int_0^1 \nabla \operatorname{vec} \nabla^2 \overline{L}_n(\theta_0 + rs\,\delta_n) s\,\delta_n \mathrm{d}s\mathrm{d}r.$$

Hence,

$$\|\operatorname{vec} \overline{\mathbb{R}}_n\| \le \int_0^1 \int_0^1 \|\nabla \operatorname{vec} \nabla^2 \overline{L}_n(\theta_0 + rs\,\delta_n)\| \mathrm{d}s\mathrm{d}r\, \|\delta_n\|,$$

$$\|\operatorname{vec} \overline{\mathbb{R}}_n\|_{S_0} \le \Big(N^{-1} \sum_{t=1}^{n} \sum_{u \in \mathcal{R}} \sum_{\gamma \in \theta^3} \|\ell_t^{\gamma}(u)\|_{S_0}\Big) \|\delta_n\| = \overline{V}_n \|\delta_n\|, \text{ say,}$$

where $\{V_t\}$ is stationary. By Lemma 8.5,

$$\|V\|_1 \le \sum_{\gamma \in \theta^3} \|\ell^{\gamma}\|_{S_0,1} < \infty. \tag{8.11}$$

Therefore, $\overline{V}_n = \mathcal{O}(1)$ a.s. by the ergodic theorem and $\delta_n = o(1)$ a.s. since $\widetilde{\theta}_n$ is strongly consistent. This gives

$$\|\operatorname{vec} \overline{\mathbb{R}}_n\|_{S_0} = \mathcal{O}(1) o(1) \text{ a.s.} = o(1) \text{ a.s.},$$

and the assertion holds. □

8.2.3. The Asymptotic Information Matrix

Proposition 8.6.

$$\overline{\mathbb{J}}_n = \tau_Z \mathbb{I}_0 + o(1) \text{ a.s.},$$

$$\mathbb{I}_0 = 2^{-1} \mathbb{E} \nabla \log h_0 \nabla' \log h_0 > 0.$$

Proof. The asymptotic covariance matrix takes the standard form $\mathbb{E}^{-1} \nabla^2 \ell_0\ \mathbb{E} \nabla \ell_0 \nabla' \ell_0\ \mathbb{E}^{-1} \nabla^2 \ell_0$ which can be seen from (8.2). By some calculations, this reduces to $\tau_Z \mathbb{I}_0$. The first statement is a consequence of the ergodic theorem with Lemma 8.5 guaranteeing that the first moments are finite.

From its definition the matrix $\mathbb{I}_0$ is positive semidefinite. Assume $\xi' \mathbb{I}_0 \xi = 0$ for some non-zero vector $\xi = \{\xi_\gamma, \gamma \in \theta\}$, indexed by the elements of θ. We choose ξ such that $\theta_0 - \xi$ is an interior point of Θ. This leads to $\mathbb{E}|\xi' \nabla \log h_0|^2 = 0$ which implies that $\xi' \nabla h_0 = 0$ a.s. and by stationarity of the process $\xi' \nabla h_t(u, \theta_0) \equiv 0$ a.s. for all (t, u). From the component form of (3.1), we see that h is linear in θ, that is, $h_t(u, \theta) = y' \theta$ for some vector y.

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

Hence, for $(t,u) \in \mathbb{Z} \times \mathcal{R}$, the following holds almost surely:

$$\begin{aligned} 0 = \xi' \nabla h_t(u,\theta_0) &= \sum_{\gamma\in\theta} \xi_\gamma \Big\{ \delta(\gamma,\omega) + \sum_{i=1}^{p} \sum_{v\in\mathbb{N}_{1i}} \delta(\gamma,\alpha_i(v)) X^2_{t-i}(u-v) \\ &\quad + \sum_{j=1}^{q} \sum_{v\in\mathbb{N}_{2i}} \delta(\gamma,\beta_j(v)) h_{t-j}(u-v,\theta_0) \Big\} + \sum_{j=1}^{q} \sum_{v\in\mathbb{N}_{2i}} \beta_j(v) \xi' \nabla h_{t-j}(u-v,\theta_0) \\ &= y'\xi. \end{aligned}$$

By adding zero, we have almost surely $\sigma_t^2(u) = y'\theta_0 - y'\xi = y'(\theta_0 - \xi) = h_t(u,\theta_0 - \xi)$. This holds if and only if $\theta_0 - \xi = \theta_0$ by Lemma 8.1, which is only possible if $\xi_\gamma \equiv 0$. □

8.2.4. Square Root n Equivalence of the Two Estimators

Lemma 8.6.

$$\|\nabla L_n - \nabla \widehat{L}_n\|_\Theta = o(1) \text{ e.a.s.}$$

Proof. Let $A = \nabla \log h$ and $U = 1 - X^2/h$ with $\widehat{A}$ and $\widehat{U}$ for the corresponding hatted versions. Now,

$$-2\ell = \log h + \frac{X^2}{h} \Longrightarrow -2\nabla\ell = AU. \tag{8.12}$$

Most of the simple computations below are carried out in terms of differences of the kind $\mathcal{D}\xi = \xi - \widehat{\xi}$. We proceed with this notation. By combining the hatted and the unhatted (8.12), we get

$$\begin{aligned} -2\nabla\mathcal{D}\ell &= AU - \widehat{A}\widehat{U} = A\,\mathcal{D}U + \mathcal{D}A\,\widehat{U}, \\ \|\nabla\mathcal{D}\ell\|_\Theta &\le \|A\|_\Theta \|\mathcal{D}U\|_\Theta + \|\mathcal{D}A\|_\Theta \|\widehat{U}\|_\Theta. \end{aligned} \tag{8.13}$$

Direct computation gives

$$\|A\|_\Theta \le \varrho \|\nabla h\|_\Theta, \quad \|\widehat{U}\|_\Theta \le \varrho X^2, \quad \mathcal{D}U = \frac{X^2}{h\widehat{h}}\mathcal{D}h, \quad \|\mathcal{D}U\|_\Theta \le \varrho^2 X^2 \|\mathcal{D}h\|_\Theta,$$

and

$$\mathcal{D}A = \frac{\nabla \mathcal{D}h}{h} + \frac{\nabla\widehat{h}\mathcal{D}h}{\widehat{h}h}.$$

We know that all the denominators are bounded below by ϱ^{-1}. By Lemma 8.4 and Proposition 3.1, both terms on the right-hand side of (8.13) is $o(1)$ e.a.s. □

Proposition 8.7. $N^{1/2}(\widetilde{\theta}_n - \widehat{\theta}_n) = o(1)$ a.s.

Proof. This time, we use $\delta_n = \widetilde{\theta}_n - \widehat{\theta}_n$. As in Proposition 8.4, we use a version of the integrated vector-valued multivariate mean value theorem,

$$\nabla L_n(\widetilde{\theta}_n) - \nabla L_n(\widehat{\theta}_n) = \Big[\int_0^1 \nabla^2 L_n(\widehat{\theta}_n + s\,\delta_n) \mathrm{d}s \Big] \delta_n. \tag{8.14}$$

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

Now, $\nabla L_n(\widetilde{\theta}_n) = 0 = \nabla \widehat{L}_n(\widehat{\theta}_n)$, so the left-hand side of (8.14) equals $\nabla \widehat{L}_n(\widehat{\theta}_n) - \nabla L_n(\widehat{\theta}_n)$. For the right-hand side, we have

$$\Big[\int_0^1 \nabla^2 L_n(\widehat{\theta}_n + s\,\delta_n)\mathrm{d}s\Big]\,\delta_n = \big[\mathbb{J}_n(\widehat{\theta}_n) + \mathbb{R}_n(\widetilde{\theta}_n, \widehat{\theta}_n)\big]\,\delta_n.$$

Normalizing both sides of the modified (8.14) with $N^{-1/2}$ gives

$$N^{-1/2}[\nabla \widehat{L}_n(\widehat{\theta}_n) - \nabla L_n(\widehat{\theta}_n)] = \Big[\overline{\mathbb{J}}_n(\widehat{\theta}_n) + \overline{\mathbb{R}}_n(\widetilde{\theta}_n, \widehat{\theta}_n)\Big]\,N^{1/2}\,\delta_n. \tag{8.15}$$

By the same arguments used in Lemma 8.5, we find that $\overline{\mathbb{R}}_n = o(1)$ a.s., $\overline{\mathbb{J}}_n = \mathbb{I}_0 + o(1)$ a.s. This is working since for any $\kappa > 0$ both $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ will stay in $S_0(\kappa)$ for all n large enough with probability one. This means that (8.11) holds here and the first factor of the right-hand side converges with probability one to a positive definite matrix. Since the left-hand side of (8.15) is $o(1)$ with probability one by Lemma 8.6 with necessity the same must be true for the right-hand side. This means that the rightmost factor is $o(1)$ with probability one. □

8.2.5. Proof of Theorem 4.2
By Proposition 8.4–8.7,

$$N^{1/2}(\widehat{\theta}_n - \theta_0) = N^{1/2}(\widetilde{\theta}_n - \theta_0) + o_{\mathbb{P}}(1) = \mathbb{I}_0^{-1} N^{1/2} \nabla L_n(\theta_0) + o_{\mathbb{P}}(1).$$

Now,

$$\nabla L_n(\theta_0) = 2^{-1} \sum_{t=1}^{n} \sum_{u \in \mathcal{R}} \nabla \log h_t(u, \theta_0)(Z_t^2(u) - 1) = \sum_{t=1}^{n} \sum_{u \in \mathcal{R}} W_t(u), \text{ say,} \tag{8.16}$$

with $W = 2^{-1} \nabla \log h(Z^2 - 1)$. Let

$$\mathbf{W}_t = \mathrm{vec}\,\{W_t(u),\, u \in \mathcal{R}\}. \tag{8.17}$$

From Lemma 8.4 and A10, it follows that $\{\mathbf{W}_t\}$ is a multivariate square integrable ergodic martingale which satisfies a martingale CLT (Hall and Heyde, 1980, Thm. 3.2, p. 58),

$$n^{-1/2} \sum_{t=1}^{n} \mathbf{W}_t \Rightarrow \mathrm{N}(0, \kappa\, \mathbb{I}_m \otimes \mathbb{I}_0). \tag{8.18}$$

Combining (8.16)–(8.18) completes the proof.

ACKNOWLEDGEMENTS

Thanks to our three anonymous reviewers for valuable input and comments.

DATA AVAILABILITY STATEMENT

The SST anomalies data used in Section 6 is publicly available from the website of Cressie and Wikle (2011): ftp://ftp.wiley.com/public/sci_tech_med/spatio_temporal_data.

Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498

SUPPLEMENTARY MATERIAL

Two animations from the real data example in Section 6: (i) The fitted processes, X, $\widehat{h}$, Z: `Animation_fitted_processes.avi`. (ii) The mean aggregated and differenced SST anomalies Y: `Animation_Y.avi`.

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Ali MM. 1979. Analysis of stationary spatial-temporal processes: estimation and prediction. *Biometrika* **66**(3): 513–518.
Berkes I, Horváth L, Kokoszka P. 2003. GARCH processes: structure and estimation. *Bernoulli* **9**(2): 201–227.
Berliner LM, Wikle CK, Cressie N. 2000. Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *Journal of Climate* **13**(22): 3953–3968.
Bivand R, Wong DWS. 2018. Comparing implementations of global and local indicators of spatial association. *TEST* **27**(3): 716–748.
Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3): 307–327.
Bollerslev T. 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics* **72**(3): 498–505.
Borovkova S, Lopuhaä R. 2012. *Spatial GARCH: a spatial approach to multivariate volatility modeling*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2176781.
Bougerol P, Picard N. 1992. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52**(1-2): 115–127.
Brockwell PJ, Davis RA. 1991. *Time series: theory and methods*, second ed., Springer Series in Statistics. New York: Springer-Verlag.
Cheysson F. 2016. *STARMA: modelling space time autoregressive moving average (STARMA) processes*. R package version 1.3, https://CRAN.R-project.org/package=starma.
Cressie N, Wikle CK. 2011. *Statistics for Spatio-temporal Data*. New York: John Wiley & Sons.
Cressie NAC. 2015. *Statistics for spatial data*, revised ed., Wiley Classics Library. New York: John Wiley & Sons.
Davis PJ. 1994. *Circulant Matrices*, 2nd ed. Providence, RI: AMS Chelsea Publishing.
Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4): 987–1007.
Ferguson TS. 1996. *A course in large sample theory*, Texts in Statistical Science Series. London: Chapman & Hall.
Francq C, Zakoïan J-M. 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**(4): 605–637.
Francq C, Zakoïan J-M. 2011. *GARCH Models: Structure, Statistical Inference And Financial Applications*. New York: John Wiley & Sons.
Guyon X. 1982. Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* **69**(1): 95–105.
Hall PG, Heyde CC. 1980. *Martingale Limit Theory and Its Applications*. New York: Academic Press.
Jeantheau T. 1998. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**(1): 70–86.
Karlsen HA, Hølleland S. 2019. Spatial- and spatio-temporal GARCH models. Working paper.
Ling S, McAleer M. 2003. Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19**(2): 280–310.
Moran PA. 1973a. A Gaussian Markovian process on a square lattice. *Journal of Applied Probability* **10**: 54–62.
Moran PA. 1973b. Necessary conditions for Markovian processes on a lattice. *Journal of Applied Probability* **10**: 605–612.
Nelson DB. 1990. Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory* **6**(3): 318–334.
Otto P, Schmid W, Garthoff R. 2018. Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity. *Spatial Statistics* **26**: 125–145.
Pfanzagl J. 1969. On the measurability and consistency of minimum contrast estimates. *Metrika* **14**(1): 249–272.
Pfeifer PE, Deutsch SJ. 1980. A three-stage iterative procedure for space-time modeling. *Technometrics* **22**(1): 35–47.
Reinsel GC. 2003. *Elements of Multivariate Time Series Analysis*. Berlin: Springer Science & Business Media.
Robinson PM. 2009. Large-sample inference on spatial dependence. *The Econometrics Journal* **12**: S68–S82.
Sato T, Matsuda Y. 2017. Spatial autoregressive conditional heteroskedasticity models. *Journal of the Japan Statistical Society* **47**(2): 221–236.
Straumann D, Mikosch T. 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Annals of Statistics* **34**(5): 2449–2495.
Wikle CK, Holan SH. 2011. Polynomial nonlinear spatio-temporal integro-difference equation models. *Journal of Time Series Analysis* **32**(4): 339–350.
Wikle CK, Hooten MB. 2010. A general science-based framework for dynamical spatio-temporal models. *Test* **19**(3): 417–451.

DOI: 10.1111/jtsa.12498 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd

Errata for proof section of Paper A

Sondre Hølleland and Hans A. Karlsen

July 2020

p. 201 In the proof of Theorem 4.1, the references to Lemma 8.1 and Lemma 8.3 should be to Theorem 8.1 and Proposition 8.3, respectively.
"Lemmas 8.1-8.3" should be "Propositions 8.1-8.3".

p. 202 The matrix $\mathbb{J}_n = -\nabla^2 L_n(\theta_0)$ is defined with the minus sign, while in the calculations below do not account for this. The definition with minus is used elsewhere in the paper and Proposition 8.4 should be corrected accordingly.

p. 206 Proposition 8.6: The first line should be $\bar{\mathbb{J}}_n = \mathbb{I}_0 + o(1)$ a.s.
In the proof below: "By some calculations this reduces to $2^{-1}\tau_Z \mathbb{I}_0^{-1}$".

p. 207 In the proof of Proposition 8.6, the reference to Lemma 8.1 should be to Proposition 8.1. Although, the final argument does not hold, since $y = y(\theta_0)$ and $y'(\theta_0 - \xi) \neq h_t(u, \theta_0 - \xi)$. This should be replaced by the argument used in the proof of Proposition 7 in Paper D.

p. 207 Lemma 8.6 should be local, i.e. $\|\nabla L_n - \nabla \widehat{L}_n\|_{\mathcal{S}_0} = o(1)$ e.a.s.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230856451 (print)
9788230867822 (PDF)