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...
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 Lopuhai (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 invariance in spatial dimension. 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
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 \( \alpha : \mathbb{Z} \times \mathbb{Z}^d \rightarrow \mathbb{R}_0 \) be a function with finite support. For fixed \( i, \alpha_i : \mathbb{Z}^d \rightarrow \mathbb{R}_0 \). The function \( \beta \) is defined in the same way. We refer to \( \alpha, \beta \) as the parameter functions. The STGARCH model is given by

\[
X_t(u) = \sigma_t(u)Z_t(u), \quad u \in \mathbb{Z}^d
\]

\[
\sigma_t^2(u) = \omega + \sum_i \sum_v \alpha_i(v)X_{t-i}^2(u - v) + \sum_i \sum_v \beta_i(v)\sigma_{t-i}^2(u - v), \quad u \in \mathbb{Z}^d,
\]

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 \( \Delta_1 = \{v \in \mathbb{Z}^d : \alpha_i(v) > 0 \} \) and \( \Delta_2 = \{v \in \mathbb{Z}^d : \beta_i(v) > 0 \} \) for \( i \geq 1 \). For \( i < p \), the model allows for some zero-valued \( \alpha_i(v) \) for \( v \in \Delta_1 \) and likewise for the \( \beta_i(v) \). The order of the model is defined as the largest \( (p, q) \) so that \( \Delta_1 \) and \( \Delta_2 \) are non-empty. With the order defined, the second part of (2.1) is expressed more specifically as

\[
\sigma_t^2(u) = \omega + \sum_i \sum_{v \in \Delta_1} \alpha_i(v)X_{t-i}^2(u - v) + \sum_i \sum_{v \in \Delta_2} \beta_i(v)\sigma_{t-i}^2(u - v).
\]

Let \( \alpha = \{\alpha_i(v), v \in \Delta_1, i = 1, \ldots, p\} \), \( \beta = \{\beta_i(v), v \in \Delta_2, i = 1, \ldots, q\} \) and write \( \theta \equiv (\omega, \alpha, \beta) \) for the parameter vector contained in the parameter space \( \Theta \), 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 \( [0, \mathbf{m} - 1] \) with \( \mathcal{R} \). For \( v \in \mathbb{Z}^d \), we use the notation \( \langle v \rangle_{\mathbf{m}} = v - \lfloor v / \mathbf{m} \rfloor \mathbf{m} \in \mathcal{R} \), where \( \lfloor v / \mathbf{m} \rfloor \) is the integer closest to \( v / \mathbf{m} \).

We write \( \mathbf{v}' \sim v \) if \( \langle \mathbf{v}' \mathbf{m} \rangle = \langle v \mathbf{m} \rangle \). In the circular model, we replace \( X_{t-i}(u - v) \)
with \( X_{t-1}^2 (u-v|m) \), and likewise for the last part of (2.2). Thus, the circular model is given by

\[
X_t(u) = \sigma_t(u) Z_t(u), \quad u \in \mathcal{R},
\]

\[
\sigma_t^2(u) = \omega + \sum_{i=1}^p \sum_{v \in \Delta_i} a_i(v) X_{t-i}^2 (u-v|m) + \sum_{j=1}^q \sum_{v \in \Delta_j} \beta_j(v) \sigma_{t-j}^2 (u-v|m), \quad u \in \mathcal{R},
\]

for \( t \in \mathbb{Z} \). Note that in general, \( \Delta_{ui} \not\subseteq \mathcal{R} \). 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 \( \sigma \) and \( \mathbb{Z} \).

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 \( m \) 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 \( m = |\mathcal{R}| = \prod_{i=1}^q m_i \) is the number of spatial locations, and likewise for \( \mathbf{Z}_t \) and \( \sigma_t^2 \). We also need \( \mathbf{X}_t^2 = \mathbf{X}_t \circ \mathbf{X}_t \). 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 \( \Delta_{ui} \) to ensure that our neighbourhood systems fulfil this property.

**A1:** Each parameter domain \( \Delta_{ui} \) is symmetric in the sense that \( -\Delta_{ui} = \Delta_{ui} \).

The neighbourhood systems for STGARCH are defined by \( \mathcal{M}_1(u) = u \cup \Delta_{ui} \). The spatial part of \( X \) that explicitly influences \( \sigma_t(u) \) is located at \( \cup_i \mathcal{M}_1(u) \) and, in parallel, the direct effect from previous spatial values of the volatility have sites \( \cup_i \mathcal{M}_2(u) \). We see that these systems are translation invariant, that is, \( \mathcal{M}_1(u+h) = \mathcal{M}_1(u) \oplus h \).

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

\[
X_t = \sigma_t \circ Z_t, \quad \sigma_t^2 = \omega 1_m + \sum_{i=1}^p A_i X_{t-i}^2 + \sum_{i=1}^q B_i \sigma_{t-i}^2, \quad t \in \mathbb{Z}.
\]
with components

\[ \sigma_i^2(u) = \omega + \sum_{j=1}^{p} \sum_{v \in N_i(u)} a_j(u,v) X_{t-i}^2(v) + \sum_{j=1}^{q} \sum_{v \in N_i(u)} b_j(u,v) \sigma_j^2(v), \]

\[ a_i(u,v) \triangleq \sum_{v' \sim v} \alpha_i(u - v'), \quad \Delta_i = \{ a_i(u,v), (u,v) \in \mathcal{R}^2 \}, \]

and likewise for the \( b_j \)'s. For the spatial neighbourhood systems we have

(i) Same structure as \( \mathcal{M}_s(u); \mathcal{N}_s(u) = u \oplus \Delta_s, u \in \mathcal{R} \) with \( \Delta_s = (\Delta_s|m) \subseteq \mathcal{R} \).

(ii) Translation invariance: \( \mathcal{N}_s(u + h) = \mathcal{N}_s(u) \oplus h \) for \( u, h \in \mathcal{R} \).

(iii) Mutual neighbourhood relations under \( A1 \): \( v \in \mathcal{N}_s(u) \Leftrightarrow u \in \mathcal{N}_s(v) \).

(iv) Relation to STGARCH: \( \mathcal{N}_s(u) = (\mathcal{M}_s(u))|m \).

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

\[ \sum_{v \in \mathcal{Z}} a_i(v) X_{t-i}^2(u - v) = \sum_{v \in \mathcal{Z}} a_i(u - v) X_{t-i}^2(v) = \sum_{r \in \mathcal{R}} \sum_{v' \sim v} a_i(u - v') X_{t-i}^2(v) \]

\[ = \sum_{r \in \mathcal{R}} \left( \sum_{v' \sim v} a_i(u - v') \right) X_{t-i}^2(v) = \sum_{r \in \mathcal{R}} a_i(u,v) X_{t-i}^2(v). \]

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

\[ \{ a_i(u, \cdot) > 0 \} = \left\{ v \in \mathcal{R} : \sum_{v' \sim v} a_i(u - v') > 0 \right\} \]

\[ = \left\{ v \in \mathcal{R}, \exists h \in \mathbb{Z}^d : u - v + h \oplus m \in \Delta_{1i} \right\} \]

\[ = \left\{ v \in \mathcal{R}, \forall h \in \mathbb{Z}^d : v \in u \oplus \Delta_{1i} \oplus h \oplus m \right\} \]

\[ = \left\{ v \in \mathcal{R} : v \in (u \oplus \Delta_{1i}|m) \right\} = u \oplus \Delta_{1i} = \mathcal{N}_{1i}(u). \]

The translation invariance holds since

\[ \mathcal{N}_{1i}(u + h) = (u \oplus h \oplus \Delta_{1i}|m) = (\mathcal{N}_{1i}(u) \oplus h|m) = \mathcal{N}_{1i}(u) \oplus h, \quad u, h \in \mathcal{R}, \]

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 \( v \in \mathcal{N}_{1i}(u) = (u \oplus \Delta_{1i}|m) \). Then, there exists \( h \in \Delta_{1i} \) such that \( (u - h|m) = v \). Thus \( u = (v + h|m) \in (v \oplus \Delta_{1i}|m) = \mathcal{N}_{1i}(v) \) due to \( A1 \). \( \square \)

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

**Remark 2.2.** If the projection of \( \Delta_{1i} \Rightarrow \Delta_{1j} \) 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 \( \Delta_{2j} \).

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 \( 4 \times 4 \) grid with a circular neighbourhood structure. This means that \( m = (4, 4), m = 16 \) 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 $R = [0, 6]^2$, 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 $R$ 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 $\alpha$ and one for $\beta$. 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 $\Delta = \Delta_{11} = \Delta_{21} = \{(v_1, v_2) \in \mathbb{Z}^2 : |v_i| \leq 1, i = 1, 2\}$, or more explicitly

$$
\Delta = \{-1, 1, 0, 0, 1, 1, -1, 0, 0, 0, 1, 1, -1, -1, 0, -1, 1, -1\}.
$$

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

$$
\alpha(v) = \begin{cases} 
    a_0, & \text{if } v = (0, 0), \\
    a_1, & \text{if } \max_i |v_i| = 1, \\
    0, & \text{otherwise,}
\end{cases}
$$

and

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

which means that the parameter vector, $\theta = (a_0, a_1, a_0, a_1, b_1, b_1)$, consists of in total 6 parameters. The parametrization of $\alpha$ gives equal weight to each neighbour ($a_1$) and another weight to the site of the observation ($a_0$). For $\beta$ 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 \( \Delta_{11} = \Delta_{21} \). In this case,
\[
\mathcal{N}(u) = (u \ominus \Delta|m) = \{ u \} \cup \mathcal{N}^{(1)}(u) \cup \mathcal{N}^{(2)}(u), \quad u \in \mathcal{R},
\]
where \( \mathcal{N}^{(1)}(u) = \{(u + v|m) : \sum_i |v_i| = 1 \} \) and \( \mathcal{N}^{(2)}(u) = \{(u + v|m) : |v_1| = 1, |v_2| = 1 \}. \) Then,
\[
a(u, v) = \begin{cases}
a_0, & \text{if } v = u, \\
a_1, & \text{if } v \not\in \mathcal{N}(u) \setminus \{u\}, \\
0, & \text{otherwise},
\end{cases}
\]
and \( b(u, v) = \begin{cases}
b_0, & \text{if } v = u, \\
b_1, & \text{if } v \in \mathcal{N}^{(1)}(u), \\
b_2, & \text{if } v \in \mathcal{N}^{(2)}(u), \\
0, & \text{otherwise}.
\end{cases} \)

In the vector notation of (2.4), \( \sigma_i^2 = \omega \mathbf{I}_m + \mathbf{A} \mathbf{X}_{i-1}^2 + \mathbf{E} \sigma_{i-1}^2 \), where \( \mathbf{E} \) is given by
\[
\mathbf{E} = \begin{bmatrix}
\mathbf{S} & \mathbf{T} & 0 & \mathbf{T} \\
\mathbf{T} & \mathbf{S} & \mathbf{T} & 0 \\
0 & \mathbf{T} & \mathbf{S} & \mathbf{T} \\
0 & 0 & \mathbf{T} & \mathbf{S}
\end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix}
b_0 & b_1 & 0 & b_1 \\
b_1 & b_0 & b_1 & 0 \\
0 & b_1 & b_0 & b_1 \\
b_1 & 0 & b_1 & b_0
\end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix}
b_1 & b_2 & 0 & b_2 \\
b_2 & b_1 & b_2 & 0 \\
0 & b_2 & b_1 & b_2 \\
b_2 & 0 & b_2 & b_1
\end{bmatrix}.
\]

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

**Remark 2.3.** If \( m \) becomes larger, the order of \( \mathbf{A} \) and \( \mathbf{E} \), which is \( m \times m \), increases. However, the row and column sums will remain the same, \( a_0 + 8a_4 \) 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
\[
\mathbf{C} = \text{circ}(c_0, \ldots, c_{n-1}) = \begin{bmatrix}
c_0 & c_1 & \cdots & c_{n-1} \\
c_{n-1} & c_0 & \cdots & c_{n-2} \\
\vdots & \vdots & \ddots & \vdots \\
c_1 & c_2 & \cdots & c_0
\end{bmatrix} = \{c_{ij}\} = \{c(i - j|n)\}.
\]

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 \( \mathbf{A} \) and \( \mathbf{E}_i \) 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 \( \mathbf{A} = \{a(u, v), (u, v) \in \mathcal{R}^2\} \) is a generalized circulant on \( \mathcal{R} \) if \( a(u, v) = a(u - v|m) \).

**Definition 2.2.** A matrix \( \mathbf{A} \) 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.
where each block is of dimension $m \times m$ by Proposition 2.2. For circulants of level $2$, this can be seen in (2.6) of Example 1, where $C$ is a block circulant whose blocks are circulant blocks of level $2$. This can be seen in (2.6) of Example 1, where $S$ and $T$ are circulant blocks of $E$, that is, $E = circ(S, T, 0, T)$. For $d \geq 2$, $A$ 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 $Q_t$ be a square block matrix of order $(p + q - 1)$ defined as

$$Q_t = \begin{bmatrix}
    A_1 Z_t^2 & E_1 & B_2 & \cdots & E_{q-1} & E_q & A_2 & \cdots & A_{p-1} & A_p \\
    I_{q-1} \otimes I_m & 0 & 0 & \cdots & 0 & 0 \\
    Z_t & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
    0 & 0 & \cdots & 0 & 0 & I_{p-2} \otimes I_m & 0
\end{bmatrix}, \tag{2.7}
$$

where each block is of dimension $m \times m$, $Z_t^2 = \text{diag}\{Z_t^2\}$, $I_k$ is a $k \times k$ identity matrix and $0$ is a null matrix. Apart from the first row, the matrix $Q_t$ is defined by $Q_t[q+1, 1] = Z_t^2$, $Q_t[i, i-1] = I_m$ for $i \in [2, q] \cup [q + 2, p + q - 1]$ and $0$ otherwise. Let $V_t$ and $c$ both be vectors of dimension $(p + q - 1)m$, defined by

$$V_t = \begin{bmatrix} S_t \\ Y_t \end{bmatrix}, \quad c = \begin{bmatrix} \omega 1_m \\ 0 \end{bmatrix}, \tag{2.8}
$$

where $S_t = (\sigma_{r+1}^2, \ldots, \sigma_{r+q+2}^2)$ and $Y_t = (X_t^2, \ldots, X_{p+2}^2)$. If $p = 1$ and $q = 0$, $Q_t = [A_1 Z_t^2]$, $V_t = S_t = \sigma_{r+1}^2$ and $c = \omega 1_m$. The same is true for $p = q = 1$, except that $Q_t = [A_1 Z_t^2 + E_1]$. Combining (2.7) and (2.8), we have that the CSTGARCH model can be expressed as a stochastic recurrence equation (SRE),

$$V_t = Q_t V_{t-1} + c, \quad t \in \mathbb{Z}. \tag{2.9}$$
Under an i.i.d. assumption on the residuals, this is the state space representation of the CSTGARCH model since the sequence \{(Q_i, \epsilon_i)\} is i.i.d. and any reasonable solution \{V_j\} is Markov. From the SRE in (2.9), we formally get

\[ V_0 = \sum_{k=0}^{\infty} \prod_{r=0}^{k-1} Q_{t-r} \epsilon. \]

When \( \log^+ \|Q_0\| < \infty \), the Lyapunov exponent is defined and satisfy

\[ \gamma_Q = \inf \left\{ E[n^{-1} \log \|Q_0Q_{n-1} \cdots Q_{n+1}\|], n \in \mathbb{N} \right\} = \lim_{n \to \infty} n^{-1} \log \|Q_0Q_{n-1} \cdots Q_{n+1}\|. \]

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

**Definition 2.3.** Let \( \|y\|_\infty \defeq \|y\|_\infty \) be the maximum absolute row sum, defined below, is a convenient choice in this context.

We will refer to the following assumptions, where \( \theta_0 \) 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 \( \theta_0 \), the Lyapunov exponent is strictly negative, \( \omega > 0 \) and \( \log^+ \|Q_0\| < \infty \).

**Remark 2.4.** The condition \( \log^+ \|Q_0\| < \infty \) is implied by A2 together with \( \omega > 0 \).

An important submatrix of \( Q \) is the \( q \times q \) non-negative block matrix

\[ E = E(\theta) \defeq \begin{bmatrix} E_1 & E_2 & \cdots & E_{q-1} & E_q \\ I_{q-1} \otimes I_m & 0 \end{bmatrix} = \{ E_{ij}, 1 \leq i, j \leq q \}, \]  

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 \( s_E \defeq \sum_{j=1}^q \sum_{i=1}^j \beta_j(\theta) \) and \( \rho_E \) be the spectral radius of \( E \).

(i) If \( \rho_E < 1 \), then \( s_E = \|E^q\| \) and \( s_E \leq \rho_E \leq s_E^{1/q} \).

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

(ii) \( \rho_E < 1 \).

(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 \( \|E_{ij}\| \geq \|E^q\| \). It remains to prove (i).

Assume that \( \rho_E < 1 \). Then \( \sum_{j=1}^q E_{ij} I_m = s_E I_q \) and \( E_{mq} \geq (s_E \wedge 1) I_{mq} \), which shows that \( s_E \leq \rho_E < 1 \). Let \( S_i^{(k)} = \sum_{j=1}^q E_{ij}^{(k)} \) be the \( i \)th block row sum of \( E^k \) with \( S_i = S_i^{(1)} \) and \( S_i^{(k)} = \text{vecblock} \left[ S_i^{(k)} \cdots S_i^{(k)} \right] \), a block matrix of dimension \( q \times 1 \). Now,

\[ S_i^{(k+1)} = \sum_{j=1}^q \left[ \sum_{j=1}^q E_{ij}^{(k)} E_{ij} \right] S_{j}, \]

\[ \|S_i^{(k+1)}\| = \left\| \sum_{j=1}^q E_{ij}^{(k)} S_{j} \right\| \leq \|S_i^{(k)}\| \|S_i\| = \|S_i^{(k)}\|. \]
Then this condition is much easier verified and does not depend on the non-zero part of $s_1$. Suppose that $A_2$ holds for $a$ by Theorem 2.3(iii). Like for univariate GARCH models, a CSTGARCH($p, q$) is weakly stationary if and only if $\mathbb{E} Z^2$ is finite and

$$s_{AB} \overset{\text{def}}{=} \sum a_i(v) + \sum b_j(v) < 1. \quad (2.12)$$

This condition is much easier verified and does not depend on $\mathcal{R}$, but it is somewhat more restrictive than $A_2$–$A_3$ with $\delta = 1$. We will prove this fact next.

**Theorem 2.4.** Suppose that $A_2$ holds for a $\delta \in (0, 1]$. Let $Q_i = \{q_i(v, v'), (v, v') \in \mathbb{R}^2\}$ and $Q_i^{ad} = \{q_i(v, v')\}$. Then

$$\text{Tr}^k Q_i^{ad} \mathbf{1} = o(1) \quad \text{w.r.t. } k, \quad (2.13)$$

is sufficient for the Lyapunov part of $A_3$. 

---

Since

$$\left\| \sum_{j=1}^q E^{(k)}_{ij} S_j \right\| = \left\| \begin{bmatrix} E^{(k)}_{i1} & \cdots & E^{(k)}_{iq} \\ 0 & I_{q-2} \otimes 0 & 0 \end{bmatrix} \begin{bmatrix} S_j \\ I_{q-2} \otimes 0 \\ 0 \end{bmatrix} \right\|. \quad (2.11)$$

From the definition of $E$, we see that

$$E^{(k)}_{ij} = \begin{cases} E_j, & \text{for } i = 1, \\ I_n \delta_{i-1,j}, & \text{for } i \in [2, q], \end{cases} \quad \Rightarrow \quad E^{(k)}_{ij} = \begin{cases} E_j, & \text{for } i = 1, \\ I_n \delta_{i-2,j}, & \text{for } i \in [3, q], \end{cases} \quad \Rightarrow \quad E^{(k)}_{ij} = \begin{cases} E_j, & \text{for } i = 1, \\ I_n \delta_{i-2,j}, & \text{for } i \in [k+1, q]. \end{cases} \quad (2.11)$$

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

$$\left\| S_{ij} \right\| \leq s_B, \quad i \in [1, q - 1] \quad \text{and} \quad \left\| S_{ij} \right\| = s_B,$$

$$\left\| \mathbb{E}^q \right\| = \left\| S_q \right\| = \max_i \left\| S_q^{(i)} \right\| = s_B.$$

Since $\left\| \mathbb{E}^q \right\|^{1/2} \leq \left\| S_q \right\|^{1/2} = s_B^{1/2}$, we must have $\rho_B \leq s_B^{1/2}.$

**Remark 2.5.** Theorem 2.3 will also hold locally for all $\theta$ in a sufficiently small compact neighbourhood of the non-zero part of $\theta_0$. We also see that $\rho_B < 1$ is a necessary, but not sufficient condition for the existence of a stationary solution of $\{X_t\}$. 

As long as $A_3$ is fulfilled, $\{X_t(u)\}$ is a strictly stationary process in the sense that, for all $n \geq 1$ and for any $(k, v) \in \mathbb{Z} \times \mathcal{R}$,

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

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

$$s_{AB} \overset{\text{def}}{=} \sum a_i(v) + \sum b_j(v) < 1. \quad (2.12)$$

This condition is much easier verified and does not depend on $\mathcal{R}$, but it is somewhat more restrictive than $A_2$–$A_3$ with $\delta = 1$. We will prove this fact next.

**Theorem 2.4.** Suppose that $A_2$ holds for a $\delta \in (0, 1]$. Let $Q_i = \{q_i(v, v'), (v, v') \in \mathbb{R}^2\}$ and $Q_i^{ad} = \{q_i(v, v')\}$. Then

$$1^T E^k Q_i^{ad} \mathbf{1} = o(1) \quad \text{w.r.t. } k, \quad (2.13)$$

is sufficient for the Lyapunov part of $A_3$. 

---

wileyonlinelibrary.com/journal/jtsa © 2019 The Authors. *J. Time Ser. Anal.* 41: 177–209 (2020) *Journal of Time Series Analysis* published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498
Proof. Let $F_k = \prod_{j=0}^{k-1} Q_{r-j}$ and $F_k^{\delta} = \prod_{j=0}^{k-1} Q_{r-j}^{\delta}$. By the row-sum norm,

$$\|F_k\| = \|\prod_{j=0}^{k-1} Q_{r-j}\| = \text{max row sum } F_k,$$

$$\|F_k\|^\delta = \left[\text{max row sum } F_k\right]^\delta \leq \text{max row sum } F_k^{\delta} = \|F_k^{\delta}\|.$$

Thus

$$E\|F_k\|^\delta \leq E\|F_k^{\delta}\| \leq E \sum_{v_0, \ldots, v_q} \prod_{j=0}^{k-1} Q_{r-j}^\delta(v_j, v_{j+1}) = \sum_{v_0, \ldots, v_q} \prod_{j=0}^{k-1} E Q_{r-j}^\delta(v_j, v_{j+1})$$

$$= \sum_{v_0, \ldots, v_q} \prod_{j=0}^{k-1} E Q_0^\delta(v_j, v_{j+1}) = 1'E_0 Q_0^{\delta} \mathbf{1} < 1,$$

for $k$ large enough.

**Remark 2.6.** This is a replacement inequality. We can replace all stochastic terms in $Q_{r-j}^{\delta}$ by their respective expected value.

**Corollary 2.1.** Let $a = q \lor (p-1)$, $r = p + q - 1$ and $C = E Q_0$. Suppose that $s_{AB} < 1$. Then $s_{AB} = \|G\|$ and $\|F_k\| < 1$ for $k = ha$ with $h > -\log(rm)/\log s_{AB}$.

**Proof.** Denote the first row of $C$ as the $\rho$-row. We have that

$$G_{ij} = \begin{cases} 
\|m \delta_{1,j} C_{ij} & \text{for } i = 1, j \geq 1, \\
\|m \delta_{i-1,j} & \text{for } i \in [2, q], \\
\|m \delta_{i+1,j} & \text{for } i = q + 1, \\
\|m \delta_{i-1,j} & \text{for } i \in [q + 2, r].
\end{cases}$$

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 $I_m$. We consider $G^\delta$ for $k \geq 1$. 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 $G^\delta$ equals the $\rho$-row. During this sequence of iterations all rows have been a $\rho$-row. Since $\|\sum_i G_{ij}\| \leq \|\|\|$ for all $i$, any block row sum is decreasing by each iteration. Now, let $k = ha$,

$$1'E_0 Q_0 \mathbf{1} \leq rm \|G\| \leq \|G^\delta\|^h = rm s_{AB}^h < 1,$$

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

**Corollary 2.2.** Let $s_{AB}^{(\delta)} = \sum \|a(v)Z_1^\delta(v) + \beta'(v)\|^\delta$. Suppose that $s_{AB}^{(\delta)} < 1$. Then $s_{AB}^{(\delta)} = \|E^{(\delta)}(\sum_{i=0}^{k-1} Q_{r-i}^{\delta})\|$ and $E\|F_k\|^\delta < 1$ for $k = ha$ with $h > -\log(rm)/\log s_{AB}^{(\delta)}$.

**Proof.** Invoke the previous proof.
Remark 2.7. The sufficient condition in Corollary 2.2 does not depend on \( m \).

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

\[
\gamma^{(d)}_{AB} = \sum_v \mathbb{E} [\alpha(v) Z_t^2(v) + \beta(v)]^d < 1,
\]

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 model has an ARMA\( (p \lor q, q) \) 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 \( r = p \lor q \). Then an extended spatio-temporal autocorrelation function exists and is given by

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

for any \( u \in \mathbb{R} \), where

\[
\mathbb{R}(h) = \text{diag}^{-1} \left( \sum_{j=0}^{\infty} \Psi_j \right) \sum_{k=0}^{\infty} \Psi_{k+h} \Psi_k',
\]

\[
\Psi_k = \begin{cases} 0, & \text{for } k < 0, \\ 1, & \text{for } k = 0, \\ \sum_{r=1}^{\infty} (\hat{\theta}_j + \hat{\theta}_j) \Psi_{k+r} - \Psi_k, & \text{for } k > 0. \end{cases}
\]

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

\[
X_t^2 = \sigma_t^2 o Z_t^2 = \sigma_t^2 + U_t, \quad U_t \overset{\text{d}}{=} \sigma_t^2 o (Z_t^2 - 1_m), \quad \text{so that } \sigma_t^2 = X_t^2 - U_t.
\]

\[
= \omega 1_m + \sum_{i=1}^{\infty} (\hat{\theta}_i + \hat{\theta}_i) X_{t-i}^2 + \sum_{i=1}^{\infty} (-\hat{\theta}_i) U_{t-i} + U_t,
\]

\[
= \omega 1_m + \sum_{i=1}^{\infty} \phi_i X_{t-i}^2 + \sum_{i=0}^{q} \theta_i U_{t-i}, \quad \text{say.}
\]

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 \( \{U_t\} \) has finite second-order moment (Brockwell and Davis, 1991, Theorem 11.3.1, p. 418). Moreover, by the same theorem, \( X_t^2 = \mu + \sum_{i=0}^{\infty} \psi_i U_{t-i} \), where the filter is given by (2.15) and \( \mu \) is the expectation of \( X_t^2 \). 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 \( U_t \) is proportional to the unit matrix and therefore the multivariate autocovariance, \( \Gamma \), function for \( \{X_t^2\} \) is given by

\[
\Gamma(h) = \text{Cov}(X_{t+h}^2, X_t^2) \propto \sum_k \Psi_{k+h} \Psi_k.
\]
By definition,

\[ R(h) = \left[ \text{diag}^{-1/2} \Gamma(0) \right] \Gamma(h) \left[ \text{diag}^{-1/2} \Gamma(0) \right], \]

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 \( X_t^2 \). It just requires (2.12) to make sense and for that purpose we use \( I_m \) for the covariance matrix of \( U_t \).

It remains to show that \( R(h) \) 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 \( \Psi_k(u, v) = \Psi_k(u + v, u + v) = \Psi_k(v, v) \), which also means that \( \text{diag} \Gamma(0) \) is proportional to the unit matrix.

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

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 \( R \),

\[ \Gamma(h) = \sum_{j=1}^r \Phi_j \Gamma(h-j) + \sum_{j=1}^q \Theta_j \Psi_j'(h-j), \quad \text{for } h \geq 0. \]

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

\[ S = \Gamma(0) = \Phi \Gamma'(1) + I_m + \Theta(\Phi' + \Theta'), \quad \Gamma(1) = \Phi S + \Theta, \quad \Gamma(h) = \Phi \Gamma(h-1), \quad h \geq 2, \]

which gives \( C = I_m + (\Phi \Theta' + \Theta \Phi') + \Theta \Theta' \) and \( S = \Phi S \Phi' + C \). For \( R \) this means

\[ R(h) = \begin{cases} \text{diag}^{-1}(S) S, & \text{for } h = 0, \\ \text{diag}^{-1}(S) \Phi^{h-1} [\Phi S + \Theta], & \text{for } h \geq 1, \end{cases} \]

\[ S = \sum_{k=0}^{\infty} \Phi^k C \Phi'^k \]

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 \( R = [0, 6] \times [0, 6] \), we visualize the spatio-temporal correlation \( \rho(h, v) = \text{Cor}(X_{t+h}^2(u + v|m), X_t^2(u)) \), 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 \( \sum a(v) + \sum b(v) = 0.86 < 1 \). The scale parameter \( \omega \) 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 increases, 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 \( \theta_0 \) be the true parameter vector which has generated the \( \{X_i\} \) process and the not directly observable squared volatility process \( \{\sigma_t^2(u)\} \). Related to the last one is \( \{h_i(u, \theta), \theta \in \Theta\} \) which we call the likelihood process. It extends the squared volatility process \( \{\sigma_t^2(u)\} \) to a function-valued process. Define equivalently to (2.4),

\[
\mathbf{h}_t \overset{a.s.}{=} \omega \mathbf{I} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{X}_{t-i}^2 + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{h}_{t-j}, \quad t \in \mathbb{Z}.
\]

(3.1)

with \( \mathbf{h}_t = \mathbf{h}_t(\theta) = \{h_i(u, \theta), u \in R \} \in \mathbb{R}^m \). Note that \( \theta = (\omega, \mathbf{a}, \mathbf{b}) \in \Theta \), while \( \{X_i\} \) is generated by \( \theta_0 \). Thus, we have that \( \mathbf{h}_t(\theta_0) = \sigma_t^2 \) a.s. By the forthcoming assumptions A4 and A5, it will become clear that \( \{h_i(u, \theta)\} \) is a spatio-temporal process with values in \( C(\Theta, \mathbb{R}) \), the space of real continuous functions defined on \( \Theta \), uniquely defined by (3.1).

We vectorize (3.1) with \( \mathbb{E} \) defined in (2.10) and the \( m \times 1 \) vector processes

\[
\mathbf{H}_t \overset{a.s.}{=} [\mathbf{h}_t, \mathbf{h}_{t-1}, \cdots, \mathbf{h}_{t-q+1}], \quad \mathbf{C}_t \overset{a.s.}{=} [\mathbf{D}_t, \mathbf{0}],
\]

(3.2)

where

\[
\mathbf{D}_t \overset{a.s.}{=} \omega \mathbf{I} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{X}_{t-i}^2 \quad \text{and} \quad \mathbf{h}_t = \sum_{j=1}^{q} \mathbf{B}_j \mathbf{h}_{t-j} + \mathbf{D}_t
\]

(3.3)

from (3.1). If \( q \in \{0, 1\} \), let \( \mathbb{E} = \mathbb{E}_11 \) \( = \mathbb{E}_1 \), \( \mathbf{C}_t = \mathbf{D}_t \) and \( \mathbf{H}_t = \mathbf{h}_t \) with \( \mathbb{E}_1 = \mathbf{0} \) for \( q = 0 \). Note that we have suppressed the dependency of \( \theta \), but \( \mathbb{E}, \mathbf{H}_t \) and \( \mathbf{C}_t \) depend on \( \theta \). From (2.10) and (3.1)–(3.3), we get a first order SRE for the likelihood process given by

\[
\mathbf{H}_t = \mathbb{E} \mathbf{H}_{t-1} + \mathbf{C}_t, \quad t \in \mathbb{Z}.
\]

(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 \( X_n \overset{\text{e.a.s.}}{\rightarrow} 0 \) 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 \( L^\delta \) for a \( \delta > 0 \). Then \( X_nY_n = o(1) \) 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

$$H_t = \sum_{k=0}^{\infty} B_k C_{t-k}. \tag{3.5}$$

The infinite sum in (3.5) converges e.a.s. uniformly with respect to $\Theta$ contained in any compact subset of the neighbourhood described in Remark 2.5. To make sure that this holds globally on $\Theta$, we need the following assumptions:

A4: On $\Theta$, $\|\sum_{j=1}^{q} E_j \| < 1$.

A5: The parameter space $\Theta$ is compact.

We extend Definition 2.3.

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

Since we do not observe the infinite past of $\{C_t\}$, let

$$\hat{X}_t \overset{\text{def}}{=} \begin{cases} 0, & \text{for } t < 1-p, \\ \hat{x}_t, & \text{for } t = 1-p, \ldots, 0, \\ X_t, & \text{for } t = 1, \ldots, n, \end{cases} \quad \hat{h}_t \overset{\text{def}}{=} \begin{cases} 0, & \text{for } t < 1-q, \\ \hat{h}_t, & \text{for } t = 1-q, \ldots, 0, \\ \hat{D}_t + \sum_{j=1}^{q} E_j \hat{h}_{t-j}, & \text{for } t = 1, \ldots, n, \end{cases} \tag{3.6}$$

with $\hat{D}_t \overset{\text{def}}{=} \omega_1 \hat{a}_t + \sum_{j=1}^{p} \hat{a}_j \hat{X}_{t-j}$ and where $\{\hat{x}_t\}$ and $\{\hat{h}_t\}$ are initial values. The theoretical counterpart of this definition is (3.3). Note that $\hat{D}_t \equiv D_t$ 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 $\hat{h}_t$ to be stochastic as long as $E\|\hat{h}_t\|_\Theta^\delta < \infty$ for some $\delta > 0$. The same applies to $\hat{X}_t$.

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

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

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

$$\hat{H}_t = E\hat{H}_{t-1} + \hat{C}_t = E^{t-p} \hat{H}_{p+1} + \sum_{j=0}^{t-p-2} E/C_{t-j}. \quad (3.7)$$

Thus

$$H_t - \hat{H}_t = E^{t-p} (H_{p+1} - \hat{H}_{p+1}). \tag{3.7}$$

Using the triangle inequality on (3.7), we have

$$\|H_t - \hat{H}_t\|_\Theta \leq \|E^{t-p}\|_\Theta^\delta (\|H_{p+1}\|_\Theta + \|\hat{H}_{p+1}\|_\Theta). \tag{3.8}$$
and by A4 and A5 together with Theorem 2.3(i),
\[ \| E'_{\theta} \|_{\Theta} \leq \| E'_{\theta} \|^{1/q} \max_{0 \leq q < q} \| E'_{\Theta} \|_{\Theta} = o(1) \text{ e.a.s.} \]

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 \( W_k \overset{\text{d}}{=} (X_k, \ldots, X_{k+p+1}, \sigma_0^2, \ldots, \sigma_{q+1}^2) \), \( k = 0, \ldots, n \). The density of \( X_n \overset{\text{d}}{=} [X_1 \ldots X_n] \) conditional on \( W_0 \), can be written as
\[ f_{X_n|W_0}(x_n|w_0) = f_{X_1|w_0}(x_1|w_0)f_{X_2|w_1}(x_2|w_1) \ldots f_{X_n|w_{n-1}}(x_n|w_{n-1}). \]

By using the first part of (2.4), we see that
\[ f_{X_k|w_{k-1}}(x_k|w_{k-1}) = \prod_{k \in R} \frac{1}{\sigma_k(u)} f_{x_k}(x_k/\sigma_k), \]
where the vector division is a Hadamard one. Given \( W_{k-1}, \sigma_k \) is successively given from the second part of (2.4).

We get the quasi-likelihood from the conditional simultaneous density with \( f_Z \) as the standard normal density and where the empirical likelihood process \( \hat{h} \) replaces \( \sigma^2 \). Taking the logarithm of this likelihood gives
\[ \hat{L}_n(\theta) \overset{\text{d}}{=} \sum_{t=1}^n \sum_{u \in R} \hat{\ell}(u, \theta), \quad \hat{\ell} \overset{\text{d}}{=} -\frac{1}{2} \left\{ \log \hat{h} + \frac{\hat{X}^2}{\hat{h}} \right\}, \tag{3.9} \]
with the processes \( \hat{X} \) and \( \hat{h} \) given by (3.6).

Remark 3.3. In (3.9) we have denoted \( \hat{\ell}(u, \theta) \), \( \hat{h}(u, \theta) \), and \( \hat{X}(u) \), 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, \( \hat{\ell}, \hat{h} \) and \( \hat{X} \).

We use the previously specified \( \hat{h} \) as the empirical counterpart of \( h \). Substituting \( h \) for \( \hat{h} \) and \( X \) for \( \hat{X} \), defines the theoretical likelihood \( L_n \) and its terms \( \ell_n \). From Proposition 3.1 we have that \( \| \hat{h}(u) - h(u) \|_{\Theta} \) converges e.a.s. to zero. Therefore, \( \hat{L}_n \) is usable as an approximation of the theoretical log likelihood function as the notation indicates. The two likelihoods have the respective maximum likelihood estimators
\[ \hat{\theta}_n \overset{\text{d}}{=} \arg\max_{\Theta} L_n \quad \text{and} \quad \hat{\theta}_n \overset{\text{d}}{=} \arg\max_{\Theta} \hat{L}_n. \]

The distinction between \( L_n \) and \( \hat{L}_n \) 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.
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 \( A(\theta, z) = \sum_{i,j} \hat{a}_{i,j} z^i z^j \) and \( B(\theta, z) = 1 - \sum_{i,j} \hat{b}_{i,j} z^i z^j \) be the matrix polynomials associated with the model. Then \( A \) and \( B \) are left coprime if any matrix polynomial factorizations \( A(\theta_1, \cdot) = U(\cdot)A(\theta_2, \cdot) \) and \( B(\theta_1, \cdot) = U(\cdot)B(\theta_2, \cdot) \) with \( \theta_1, \theta_2 \in \Theta \) imply that \( U \) is unimodular, that is, that the determinant of \( U \) is a non-zero constant.

The following list of assumptions will be used.

**A6:** The projections \( A_{i,j} \rightarrow A_{i,i} \) into \( \mathcal{R} \) preserve cardinality.

**A7:** At \( \theta_0 \), \( A \) and \( B \) are left coprime and either \( \hat{a}_{i,j} \mid \mathbb{E} \) or \( \hat{a}_{i,j} \mid \mathbb{E} \) has full rank.

**A8:** The interior of \( \Theta \) as a subset of the Euclidean space contains \( \theta_0 \).

**A9:** The variance of the squared residual process, \( \tau_Z \), is finite.

**A10:** The condition \( A_6 \) means that each \( \hat{a}_{i,j} \) can be contained in \([0, m - 1]\) 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, \( \hat{\theta}_n \) is strongly consistent; \( \hat{\theta}_n \rightarrow^a \theta_0 \).

**Theorem 4.2.** Under the assumptions A2–A10, \( \hat{\theta}_n \) is strongly consistent and asymptotically normally distributed:

\[
N^{1/2}(\hat{\theta}_n - \theta_0) \Rightarrow N(0, \kappa I_0^{-1}),
\]

where \( \kappa = 2^{-1} \tau_Z \) and \( I_0 \) is the information matrix given by

\[
I_0 = 2^{-1} \mathbb{E}(\nabla \log h_0)(\nabla \log h_0)', \quad h_0 = h(\theta_0).
\]
Remark 4.2. When the residual process is Gaussian, the constant $\kappa = 1$ and $I_0$ is the Fisher information matrix. Note that $I_0$ depends only on the marginal distribution of $(h_0, \forall h_0)$ 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 $I_0$ depends on $R$, but the model parameters do not, though the model does.

Remark 4.3. Under A2–A10, $\hat{\theta}_n$ will eventually satisfy the likelihood equations $\nabla \hat{L}_n = 0$, since $\hat{\theta}_n$ is consistent for the interior point $\theta_0$ and $\hat{L}_n$ is smooth and takes its maximum in an open ball around $\theta_0$. The same relation is of course true for $\tilde{\theta}_n$ and $\nabla \hat{L}_n$.

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, $\hat{\theta}_n$, 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

$$\lim_{n \to \infty} \frac{\hat{\theta}_n - \theta_0}{\sqrt{\hat{\Sigma}_n}} \Rightarrow N(0, I_0),$$

where $\hat{\Sigma}_n$ is the Monte Carlo approximated covariance matrix of $\hat{\theta}_n$, while the marginal checks whether $|\hat{\theta}_{nj} - \theta_{nj}|/\hat{SD}(\hat{\theta}_{nj}) \leq z_{0.975} \approx 1.96$ for every parameter estimator $\hat{\theta}_{nj}$, where $\hat{SD}(\hat{\theta}_{nj})$ is the square root of the diagonal elements of $\hat{\Sigma}$. 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.
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]

| Dimension | Parameter  | Circular data | Non-circular data | Circular data | Non-circular data | PBBC |
|-----------|------------|---------------|-------------------|---------------|-------------------|------|
|           |            | $\omega$ | $\alpha$ | $\beta$ | $\omega$ | $\alpha$ | $\beta$ | $\omega$ | $\alpha$ | $\beta$ |
| 5 $\times$ 3000 | 100 $\theta_0$ | 31.00 | 2.400 | 7.000 | 31.00 | 2.400 | 7.000 | 31.00 | 2.400 | 7.000 |
|            | (Scale MSE) | $(10^{-5})$ | $(10^{-5})$ | $(10^{-5})$ | $(10^{-5})$ | $(10^{-5})$ | $(10^{-5})$ | $(10^{-5})$ | $(10^{-5})$ | $(10^{-5})$ |
|            | 100 BIAS($\hat{\theta}$) | 1.064 | -0.070 | 12.63 | -0.489 | -0.209 | -0.270 | -0.075 | 0.089 |
|            | 100 SD($\hat{\theta}$) | 4.289  | 0.121  | 0.299 | 2.554  | 0.119  | 0.392 | 7.070  | 0.147  | 0.463 |
|            | MSE | 1.949  | 0.147  | 0.944 | 1.84  | 2.529  | 1.970 | 4.995  | 0.271  | 2.219 |
|            | BIAS$^2$ 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 | 0.940  | 0.000  | 0.936 | 0.000  | 0.000  | 0.936 | 0.000  | 0.000  | 0.936 |
| 10 $\times$ 10 3000 | Uncond. SD | 1.419 (9.318 $\times$ 10$^{-3}$) | 1.419 (9.350 $\times$ 10$^{-3}$) | 1.419 (9.441 $\times$ 10$^{-3}$) | 1.419 (9.318 $\times$ 10$^{-3}$) | 1.419 (9.350 $\times$ 10$^{-3}$) | 1.419 (9.441 $\times$ 10$^{-3}$) | 1.419 (9.318 $\times$ 10$^{-3}$) | 1.419 (9.350 $\times$ 10$^{-3}$) | 1.419 (9.441 $\times$ 10$^{-3}$) |
| 15 $\times$ 15 3000 | 100 BIAS($\hat{\theta}$) | 0.871 | -0.055 | 2.947 | -0.260 | 0.096 | -0.432 | -0.029 | 0.052 |
|            | 100 SD($\hat{\theta}$) | 2.909  | 0.190  | 3.097  | 0.199  | 3.152  | 0.069  | 0.211 |
|            | MSE | 0.920  | 0.391  | 1.826  | 0.488  | 1.010  | 0.055  | 0.470 |
|            | BIAS$^2$ SD$^2$ | 0.090  | 0.085  | 0.907  | 0.235  | 0.019  | 0.178  | 0.062 |
|            | Marg. Cov | 0.938  | 0.948  | 0.852  | 0.926  | 0.954  | 0.940  | 0.958 |
|            | Simult. Cov | 0.960  | 0.006  | 0.932 | 0.006  | 0.006  | 0.932 | 0.006  | 0.006  | 0.932 |
|            | Uncond. SD | 1.419 (4.326 $\times$ 10$^{-3}$) | 1.419 (3.780 $\times$ 10$^{-3}$) | 1.419 (3.831 $\times$ 10$^{-3}$) | 1.419 (4.326 $\times$ 10$^{-3}$) | 1.419 (3.780 $\times$ 10$^{-3}$) | 1.419 (3.831 $\times$ 10$^{-3}$) | 1.419 (4.326 $\times$ 10$^{-3}$) | 1.419 (3.780 $\times$ 10$^{-3}$) | 1.419 (3.831 $\times$ 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 ($\omega$, $\alpha$, $\beta$) 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 $\omega$ 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 $\beta$ in the smallest sample size, this leads to a smaller MSE. For CC, the squared bias goes down slower than the variance for $\omega$ and $\beta$, while for $\alpha$ the bias is practically zero in all cases. For the CNC estimates, the squared bias goes down faster than the variance for $\omega$, while for $\alpha$ the small standard deviation and the large bias is inflating the ratio evidently. For $\beta$ the relationship between squared bias and variance is quite stable. In the PBBC case, we find all the blue numbers for $\omega$ and $\beta$ with values close to zero. For $\alpha$ 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 $\theta_0$ with decreasing spread as spatial sample size increases. For $\omega$, 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 $\theta_0$, that is, $9(\alpha + \beta) = 0.846 < 1$, and $\mathbb{E}(\sigma^2)^{1/2} = (\omega/(1-9\alpha-9\beta))^{1/2}$ is 1.419 under $\theta_0$. The table shows that this quantity is, to some surprise, preserved in all different simulations with high accuracy. For the $\alpha$ column in Figure 4, we see why the coverage in Table I

---

J. Time Ser. Anal. 41: 177–209 (2020) © 2019 The Authors. wileyonlinelibrary.com/journal/jtsa
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 $\hat{\theta}_n$ 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 $\beta$ 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 $2^\circ \times 2^\circ$ 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 $4^\circ \times 4^\circ$ 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 $\{Y_i(u_1, u_2)\}$ denote the observations, then

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

weileyonlinelibrary.com/journal/jtsa © 2019 The Authors. J. Time Ser. Anal. 41: 177–209 (2020)
Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498
We then have 72 model candidates and choose the one that minimize Akaike’s information criterion (AIC). This is implemented in the \texttt{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.

\[
\sum_{j=1}^{12} \sum_{k=1}^{2} \phi_{jk} \psi^{(k)} Y_{t-j} + \sum_{j=1}^{12} \sum_{k=0}^{2} \theta_{jk} \psi^{(k)} X_{t-j} + X_t, \tag{6.1}
\]

where \(\psi^{(k)}\) 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 \texttt{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.

The differenced data is correlated and shows signs of a 12 month season (see Figure 7b), so we fit 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 \texttt{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.

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

| Parameter | CSTARMA Estimates | CSTGARCH Estimates | SD |
|-----------|-------------------|--------------------|----|
| \(\phi_{1,0}\) | 0.361 | \(\omega\) | 5.06 \times 10^{-4} | 8.02 \times 10^{-5} |
| \(\phi_{2,0}\) | 0.225 | \(\alpha_0\) | 6.25 \times 10^{-2} | 3.91 \times 10^{-3} |
| \(\phi_{2,1}\) | 0.026 | \(\alpha_1\) | 6.57 \times 10^{-2} | 3.82 \times 10^{-3} |
| \(\phi_{1,0}\) | 0.360 | \(\alpha_2\) | 3.68 \times 10^{-2} | 3.62 \times 10^{-3} |
| \(\phi_{1,1}\) | -0.248 | \(\alpha_3\) | 4.05 \times 10^{-2} | 4.23 \times 10^{-3} |
| \(\theta_{2,1}\) | 0.163 | \(\beta_0\) | 2.66 \times 10^{-1} | 7.4 \times 10^{-2} |
| \(\theta_{3,1}\) | 0.143 | \(\beta_1\) | 5.09 \times 10^{-1} | 6.94 \times 10^{-2} |
| \(\theta_{11,0}\) | -0.378 | 1.06 \times 10^{-2} | |
| \(\theta_{12,1}\) | 0.205 | 2.29 \times 10^{-2} | |

where \(\nabla_1\) and \(\nabla_2\) 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 \times latitude \times time) or 111720 data points, with longitudes from 165°E to 241°E (119°W) and latitudes from 24°S to 28°N.

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]
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 \( \approx 0.025 \), which fits well with \( \hat{\omega}/\{1 - \sum_j(\hat{\alpha}_j + \hat{\beta}_j)\} \approx 0.027 \). Plotting four-dimensional processes is difficult, and animations of all the fitted spatio-temporal series \( Y, X, X^2 \) and \( Z \) 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
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 McAuleer (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.
8.1. Proof of Consistency

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

The subscript 0 refers to \( \theta_0 \) for quantities depending on \( \theta \), 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 \( \omega_0 = \min_{\theta} \omega(\theta) \) so that \( h \geq \omega_0 \).

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 \( \theta \equiv \theta_0 \).

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 \( \mathbf{A} \) and \( \mathbf{B} \). By A4, \( \mathbf{B} \) is causal and A7 states the coprime property. Since \( \mathbf{A}_p \) and \( \mathbf{B}_q \) 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 \( [\mathbf{A}_p, \mathbf{B}_q] \) is satisfied. These three properties guarantees identifiability (Reinsel, 2003, p. 37). \( \square \)

8.1.2. The Asymptotic Likelihood has a Unique Global Maximum

Proposition 8.2. The asymptotic likelihood,

\[
L \overset{\text{def}}{=} \mathbb{E} \ell = -2^{-1} \mathbb{E} \left( \log h + \frac{X^2}{h} \right),
\]

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

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

\[
Q \overset{\text{def}}{=} 2L + \mathbb{E} \log h_0 = \mathbb{E}(\log h_0 + 2\ell) = \mathbb{E}\left( \log h_0 - \log h - \frac{X^2}{h} \right) = \mathbb{E}\left( \log \frac{h_0}{h} - \frac{h_0}{h} \right),
\]

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. \( \square \)

8.1.3. The Observable and Non-observable Likelihood are Equivalent

Proposition 8.3.

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

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

\[
| \ell' - \ell'_{\theta} | \leq | \log \hat{h} - \log h| + X^2(\hat{h}h)^{-1}|\hat{h} - h|,
\]

\[
\| \ell' - \ell'_{\theta} \|_{\Theta} \leq (\omega_0 \vee 1)^{-1}(1 + X^2)\| \hat{h} - h \|_{\Theta},
\]

so that

\[
\| \hat{L}_n - L_n \|_{\Theta} \leq \sum_{i=1}^{n} \sum_{u \in \mathcal{K}} (\omega_0 \vee 1)^{-1}(1 + X^2(u))\| \hat{h}(u) - h(u) \|_{\Theta},
\]

By Proposition 3.1 and Theorem 2.3 the conditions in Lemma 3.1 hold and (8.1) follows. \( \square \)
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 $\Theta'$ denote an arbitrary compact subset of $\Theta$, $T$ be an ergodic transformation on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ and $\xi_0(\theta)$ be a stochastic variable in $[-\infty, \infty]$ for all $\theta \in \Theta$. Define $\xi_t(\theta) = \xi_0(\omega, \theta) = \xi_0(T^t \omega, \theta)$ 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 $\bar{\xi}_n = n^{-1}\sum_{i=0}^{n-1} \xi_i$. Then

(iv) $\bar{\xi}_n \xrightarrow{a.s.} \mu$.
(v) For all $\Theta' \subseteq \Theta$: $\lim\sup_{n\to\infty}\|\bar{\xi}_n\|_{\Theta'} \leq \|\mu\|_{\Theta'}$.

**Proof of 4.1.** Let $\xi_i = m^{-1}\sum_{u\in R} \xi_i(u)$. Then $\{\xi_i\}$ is ergodic and (i) and (ii) of Lemma 8.1 is satisfied, since $\epsilon'$ is continuous. For (iii) $\epsilon' \leq -\log h \leq -\log \omega < \infty$. From (iv), we get $\tilde{L}_n = \bar{\xi}_n = \mu + \epsilon(1) = L + \epsilon(1)$ a.s. and (v) $\lim\sup_{n\to\infty}\|\tilde{L}_n\|_{\Theta'} \leq \|L\|_{\Theta'}$. Since $\tilde{L}_n = L_n + (\hat{\tilde{L}}_n - L_n)$ we get from Lemma 8.3 that $\{L_n\}$ 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. $\square$

8.2. Proof of the CLT

Throughout this subsection, we assume A2–A10.

Let $S_0(\kappa)$ be an open ball with centre in $\theta \in \Theta$ and radius $\kappa > 0$. When A9 holds, let $S_0 = S_0(\kappa) = S_0(\kappa)$ be contained in $\Theta$ for some $\kappa > 0$ and let $\overline{S}_0$ be the closure of this open ball. We use $\gamma$ both as a component of the parameter vector $\theta$ and its reference index. As an index, we write $\gamma \in \theta$ and for multiple indexes $\gamma = (\gamma_1, \ldots, \gamma_k) \in \theta^k$. 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 $\theta_0$ in the parameter space $\Theta$. 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 $\hat{\theta}_n$ and $\tilde{\theta}_n$ converges to $\theta_0$. 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 $\Theta$ are uniformly bounded in both directions; $\phi = \|\theta\|_{S_0} \vee 1/\|S_0\|_\theta$ so that $0 < \phi^{-1} \leq \gamma \leq \phi$ for any component $\gamma \in \theta$ on $S_0$.
(iv) The elements of $\mathbb{E}^k, b^{(k)}$, decrease with an exponential rate to zero as $k$ goes to infinity; $b^{(k)} \leq \rho^{k-q}$ with $\rho = \rho_{\mathbb{E}}^{1/q}$, where $\rho_{\mathbb{E}}$ is the spectral radius of $\mathbb{E}$.

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.**

\[-(\overline{J}_n + \overline{P}_n)N^{1/2}(\hat{\theta}_n - \theta_0) = N^{-1/2}N_{\alpha}(\theta_0), \tag{8.2} \]

*J. Time Ser. Anal.* 41: 177–209 (2020) © 2019 The Authors. Wiley Online Library, Journal of Time Series Analysis published by John Wiley & Sons Ltd DOI: 10.1111/jtsa.12498
Proof. Let \( \delta_n = \tilde{\theta}_n - \theta_0 \). We use an integrated mean value theorem for the vector-valued multivariate \( \nabla L_n \).

\[
\nabla L_n(\tilde{\theta}_n) = \nabla L_n(\theta_0) + \left[ \int_0^1 \nabla^2 L_n(\theta_0) + s \delta_n \right] \delta_n
\]

Since \( \tilde{\theta}_n \) satisfies the log likelihood equations, that is, \( \nabla L_n(\tilde{\theta}_n) = 0 \), 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 \( \overline{P}_n \) can be neglected.

(ii) The observed information matrix, \( \overline{J}_n \), converges to a positive definite matrix.

(iii) The observable estimator \( \tilde{\theta}_n \) and its theoretical companion \( \theta_n \) are square root \( N \) equivalent.

8.2.2. The Remainder Term

Recall that \( h = h_i(u) \) and \( E_{11}^{(k)} \) is the first block of \( E^k \).

**Lemma 8.1.**

\[
h = \sum_{k=0}^{\infty} f_k, \quad f_k(t, u, \theta) \overset{\text{def}}{=} E_{11}^{(k)} D_{r-k}(u)
\]  

(8.3)

\[
\phi^{-1} g_k \leq f_k \leq \phi \ g_k, \quad g_k(t, u, \theta) \overset{\text{def}}{=} E_{11}^{(k)} D_{r-k}(u, \theta_0).
\]  

(8.4)

**Proof.** From (3.2)–(3.5), we can write

\[
h_i(u) = h_i(u) = \sum_{k=0}^{\infty} E_{11}^{(k)} D_{r-k}(u) = \sum_{k=0}^{\infty} f_k(t, u).
\]

For the second part,

\[
\phi^{-1} D_i(\theta_0) \leq D_i(\theta) \leq \phi D_i(\theta_0).
\]

The main influence on \( h \) from the parameter \( \theta \) goes through \( B \). It is therefore advantageous to neutralize the impacts from the \( D_i \)'s. This is the content of the second point in the lemma above.

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

\[
0 \leq D^s E^k \leq \phi^r k_{(r)} E^k, \quad 0 \leq D^s D_i(u) \leq \phi D_i(u).
\]  

(8.5)
Proof. Each element of $B$ is a multivariate polynomial in the variables of $\{\beta(v)\}$ 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 $k$, 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 $r$. 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 $y \in \theta'$ with $r \in \mathbb{N}$,

$$0 \leq h^{(r)} \leq \phi^{r+1} \sum_{k=0}^{\infty} k_s k.$$

Proof. Let $y$ be fixed. We split this vector so that $y^{(1)}$ is the $B$-part and $y^{(2)}$ is the remaining part. Corresponding to this structure, we write $\mathbb{E}_1$ for $j=1,2$. By (8.3),

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

and by Lemma 8.2

$$D^r f_k(t, \cdot) = D^r \mathbb{E}_1^{(k)} D_{t-k} = D^r \mathbb{E}_1^{(k)} D_{t-k} \leq \phi^{r+1} k_s k(t, \cdot).$$

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 $\theta \in S$. Then $\|V\|_{S, p} \equiv \mathbb{E}^{1/(p+1)} \|V\|^p_S$ for any $p > 0$. If $V$ is independent of $\theta$, we may drop the subscript $S$.

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

Lemma 8.4. For any $y \in \theta'$, with $r \in \mathbb{N}$ and any moment $p \in \mathbb{N}_+$, there exists a $\kappa > 0$ such that, with $S_0 = S_0(\kappa)$,

i) $\|h_{0,r}\|_{S, p} < \infty.$

ii) $\|h^{(r)}\|_{S, p} < \infty.$

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 $\rho = \rho(\mathbb{E}_0)$.

Step 1: Expansion and simplifications of the two fractions.

Let $g_0 = g_0(\theta_0) = f_k(\theta_0)$. By (8.4),

$$f_0 \geq \phi^{-1} g_0 \quad \text{and} \quad g_0 = g_0(\theta_0) \geq \phi^{-1},$$
and from (8.3), we can write

\[ \frac{h_0}{h} = \frac{\sum_{k=0}^{\infty} g_{0k}}{\sum_{\ell=0}^{\infty} f_{\ell}} \leq \frac{g_{00}}{f_0} + \sum_{k=1}^{\infty} \frac{g_{0k}}{f_0 + \sum_{\ell=1}^{\infty} f_{\ell}} \]

\[ \leq \alpha^3 \left( 1 + \sum_{k=1}^{\infty} \frac{g_{0k}}{1 + \sum_{\ell=1}^{\infty} g_{\ell}} \right) \leq \alpha^2 \left( 1 + \sum_{k=1}^{\infty} \frac{g_{0k}}{1 + g_k} \right). \]  

(8.6)

For (ii) we get from Lemma 8.3,

\[ \frac{h^r}{h} \leq \frac{\sum_{\ell=0}^{\infty} k^{(\ell)} g_{k}}{\sum_{\ell=0}^{\infty} f_{\ell}} \leq \alpha^{r+2} \left( 1 + \sum_{k=1}^{\infty} \frac{k^{(\ell)} g_{k}}{f_0 + \sum_{\ell=1}^{\infty} f_{\ell}} \right) \]

\[ \leq \alpha^{r+2} \left( 1 + \sum_{k=1}^{\infty} \frac{k^{(\ell)} g_{k}}{1 + g_k} \right). \]

Step 2: Part (i)

We start with the first part. Let \( (t, u) \) be fixed. Let \( g_k(\theta) \equiv g_k(t, u, \theta), V_k(\nu) \equiv D_{-k}(v, \theta_0) \) and

\[ b_k(\nu, \theta) \equiv b_k^{(t)}(u, v, \theta), \quad a_k(\nu, \theta) \equiv a_k(v, \theta_0). \]  

(8.7)

By (8.4) and (8.7),

\[ g_k(t, u) = E_{11}^{(k)} D_{-k}(u, \theta_0) = \sum_{v \in \mathbb{R}} b_k^{(t)}(u, v) V_k(v) = \sum_{v \in \mathbb{R}} b_k^{(t)}(v)V_k(v). \]  

(8.8)

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

\[ \sum_{k=1}^{\infty} \frac{g_{0k}}{1 + g_k} = \sum_{k=1}^{\infty} \sum_{u \in \mathbb{R}} \frac{a_k V_k}{1 + b_k V_k}. \]

Let \( \kappa > 0, \mathcal{T}_0(\kappa) \equiv S_0(\kappa \| \theta_0 \|) \) and

\[ \mathcal{A}_k(\kappa) \equiv \{ \theta \in \Theta : (1 - \kappa)^k E_0^k \leq E \leq (1 + \kappa)^k E_0^k \}, \quad k \geq 1. \]

Then by looking at \( E^{k+1} = E^k E \), we see that

\[ \mathcal{T}_0 \subseteq \cup_{k} \mathcal{A}_k. \]  

(8.9)

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

\[ (1 - \kappa)^k a_k \leq b_k \leq (1 + \kappa)^k a_k, \quad a_k \leq \rho^{k-q} \text{ on } \mathcal{T}_0(\kappa). \]  

(8.10)

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

\[ \frac{a_k V_k}{1 + b_k V_k} = \left( \frac{a_k}{b_k} \right) \left( \frac{b_k V_k}{1 + b_k V_k} \right) \leq \left( \frac{a_k}{b_k} \right) b_k^q V_k^q = a_k b_k^{k-1} V_k^q \]

\[ \leq \rho^{-q} \left( \frac{\rho^q}{(1 - \kappa)^{k-q}} \right) V_k^q = \rho^{-q} \rho^q V_k^q, \text{ say.} \]
since \((1 + x)^{-1} \leq x^\beta\) for \(x \geq 0\) and by (8.10). Choose \(\kappa < (1 - \xi) \land \kappa_0\) with \(\xi = \rho^{\delta/(1-\delta)}\). Then \(\tau < 1\). It is clear that our choice of \(\kappa\) is independent of \(u\) and \(v\) and \(\{b_k > 0\} = \{a_k > 0\}\) on \(\mathcal{H}_0(\kappa)\). Thus,

\[
\rho^{-\beta} \left\| \frac{h_0}{h} \right\|_{S_{\delta, p}^\beta} \leq \sum_{k=1}^{\infty} \sum_{v \in \mathcal{R}} \left\| \frac{a_k V_k}{1 + b_k V_k} \right\|_{S_{\delta, p}^\beta} \leq \sum_{k=1}^{\infty} \sum_{v \in \mathcal{R}} \rho^{-\delta \beta} \left\| D_{r-k}(v, \theta_0) \right\|_{S_{\delta, p}^\beta} \leq \left[ \rho^{-\delta \beta} r(1 - \tau)^{-1} m \left\| D_i(v, \theta_0) \right\|_{S_{\delta, p}^\beta} \right] < \infty,
\]

which ends the proof of part one.

**Step 3:** The second part.

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

\[
\frac{k_{(r)} b_k V_k}{1 + b_k V_k} \leq k_{(r)} b_k V_k \leq \rho^{-\delta \beta} k_{(r)} (1 + \kappa) \rho^{-1} V_k^\beta.
\]

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

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

**Proof.** It is easy to verify that

\[
e(\gamma) = \sum_{k=1}^{M} a_k U_k + \sum_{k=1}^{L} b_k V_k,
\]

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

\[
U = \prod_{j=1}^{r} \frac{h(\xi_j)}{h}, \quad V = U' \frac{X_2}{h}, \quad \sum \dim(\xi_j) \leq \dim(\gamma),
\]

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

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

\[
\left\| U \right\|_{S_0, p} \leq \max_j \left\| \frac{h(\xi_j)}{h} \right\|_{S_0, p} \quad p = 1, 2,
\]

\[
\left\| V \right\|_{S_0, 1} \leq \left\| U \right\|_{S_0, 2} \left\| \frac{h_0}{h} \right\|_{S_0, 2} \left\| Z^2 \right\|_2,
\]

which is finite by Lemma 8.4 and A10.

**Proposition 8.5.** \(\overline{\omega}_n = o(1)\) a.s.
Proof. Let $\tilde{\delta}_n = \tilde{\theta}_n - \theta_0$. We start by applying an integrated version of the multivariate mean value theorem, this time on $\text{vec} \nabla^2 L_n$,

$$\text{vec} \nabla^2 L_n(\theta_0 + s \tilde{\delta}_n) = \text{vec} \nabla^2 L_n(\theta_0) + \int_0^1 \left[ s \text{vec} \nabla^2 L_n(\theta_0 + rs \tilde{\delta}_n) \right] dr.$$ 

Hence,

$$\|\text{vec} \tilde{R}_n\| \leq \int_0^1 \int_0^1 \|s \text{vec} \nabla^2 L_n(\theta_0 + rs \tilde{\delta}_n)\| dr \|\tilde{\delta}_n\|,$$

$$\|\text{vec} \tilde{R}_n\|_{S_0} \leq \left( N^{-1} \sum_{t=1}^n \sum_{u \in R} \sum_{r \in R} \|e'_t(u)\|_{S_0} \right) \|\tilde{\delta}_n\| = \tilde{V}_n \|\tilde{\delta}_n\|, \text{ say},$$

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

$$\|V\|_1 \leq \sum_{r \in R} \|e^r\|_{S_0,1} < \infty. \quad (8.11)$$

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

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

and the assertion holds. \hfill \Box

8.2.3. The Asymptotic Information Matrix

Proposition 8.6.

$$\tilde{J}_n = r_Z \|l_0 + o(1) \text{ a.s.},$$

$$\|l_0\| = 2^{-1} \mathbb{E} \nabla \log h_0 \nabla' \log h_0 > 0.$$ 

Proof. The asymptotic covariance matrix takes the standard form $\mathbb{E}^{-1} \nabla^2 \ell_0 \mathbb{E} \nabla' \ell_0 \mathbb{E}^{-1} \nabla^2 \ell_0$ which can be seen from (8.2). By some calculations, this reduces to $r_Z l_0$. 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 $l_0$ is positive semidefinite. Assume $\xi' l_0 \xi = 0$ for some non-zero vector $\xi = \{\xi_\gamma, \gamma \in \Theta\}$, indexed by the elements of $\Theta$. We choose $\xi$ such that $\theta_0 - \xi$ is an interior point of $\Theta$. This leads to $\mathbb{E} \xi' \nabla \log h_0 \xi = 0$ 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 $\theta$, that is, $h_t(u, \theta) = y' \theta$ for some vector $y$. 

wileyonlinelibrary.com/journal/jtsa © 2019 The Authors. J. Time Ser. Anal. 41: 177–209 (2020) 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:

\[
0 = \xi' \nabla h_t(u, \theta_0) = \sum_{\gamma \in \Theta} \xi \left\{ \delta(\gamma, \omega) + \sum_{j=1}^{p} \sum_{v \in \Lambda_j} \delta(\gamma, \sigma_j(v)) X_{j-1}(u - v) \right\} + \sum_{j=1}^{q} \sum_{v \in \Lambda_j} \beta_j(v) \xi \{\delta(\gamma, \omega) + \sum_{i=1}^{p} \sum_{v \in \Lambda_i} \delta(\gamma, \alpha_i(v)) X_{i-1}(u - v, \theta_0) \}
\]

\[
= y' \xi.
\]

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 \(\xi \equiv 0\).

8.2.4. Square Root n Equivalence of the Two Estimators

Lemma 8.6.

\[\| \nabla L_n - \hat{\nabla} L_n \|_\Theta = o(1) \text{ e.a.s.}\]

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

\[
-2\ell' = \log h + \frac{X^2}{h} \implies -2\nabla \ell' = AU.
\]

(8.12)

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

\[
-2\nabla D \ell' = AU - \hat{A} \hat{U} = ADU + DA \hat{U},
\]

(8.13)

Direct computation gives

\[
\| A \|_\Theta \leq \varphi \| \nabla h \|_\Theta, \quad \| \hat{U} \|_\Theta \leq \varphi X^2, \quad DU = \frac{X^2}{hh} Dh, \quad \| DU \|_\Theta \leq \varphi^2 X^2 \| Dh \|_\Theta,
\]

and

\[
DA = \frac{\nabla Dh}{h} + \frac{\hat{\nabla} Dh}{\hat{h}}.
\]

We know that all the denominators are bounded below by \(\varphi^{-1}\). 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.

\(N^{1/2} (\tilde{\theta}_n - \hat{\theta}_n) = o(1) \text{ a.s.}\)

Proof. This time, we use \(\delta_n = \tilde{\delta}_n - \hat{\delta}_n\). As in Proposition 8.4, we use a version of the integrated vector-valued multivariate mean value theorem,

\[
\nabla L_n(\tilde{\theta}_n) - \nabla L_n(\hat{\theta}_n) = \left[ \int_0^1 \nabla^2 L_n(\hat{\theta}_n + s \delta_n) ds \right] \delta_n.
\]

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

\[
\left[ \int_0^1 \nabla^2 L_n(\hat{\theta}_n + s \delta_n) ds \right] \delta_n = \left[ J_n(\hat{\theta}_n) + \nabla \hat{L}_n(\hat{\theta}_n) \right] \delta_n.
\]

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

\[
N^{-1/2}[\nabla \hat{L}_n(\hat{\theta}_n) - \nabla L_n(\hat{\theta}_n)] = \left[ J_n(\hat{\theta}_n) + \nabla \hat{L}_n(\hat{\theta}_n) \right] N^{1/2} \delta_n.
\] (8.15)

By the same arguments used in Lemma 8.5, we find that \( \delta \equiv \sigma(1) \) a.s., \( \gamma \equiv \sigma(1) \) a.s. This is working since for any \( \kappa > 0 \) both \( \hat{\theta}_n \) and \( \tilde{\theta}_n \) will stay in \( S(\kappa) \) 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 \( \sigma(1) \) 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 \( \sigma(1) \) with probability one. \( \square \)

8.2.5. Proof of Theorem 4.2

By Proposition 8.4–8.7,

\[
N^{1/2}(\hat{\theta}_n - \theta_0) = N^{1/2}(\tilde{\theta}_n - \theta_0) + o_p(1) = \|\delta\|^{-1} N^{1/2} \nabla L_n(\theta_0) + o_p(1).
\]

Now,

\[
\nabla L_n(\theta_0) = 2^{-1} \sum_{i=1}^n \sum_{u \in R} \nabla \log h_i(u, \theta_0)(Z_i(u) - 1) = \sum_{i=1}^n \sum_{u \in R} W_i(u), \text{ say}, \quad (8.16)
\]

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

\[
W_i = \text{vec } \{ W_i(u), u \in R \}. \quad (8.17)
\]

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

\[
n^{-1/2} \sum_{t=1}^n W_i \Rightarrow N(0, \kappa \|\delta\| \otimes \|\delta\|). \quad (8.18)
\]

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.
SUPPLEMENTARY MATERIAL

Two animations from the real data example in Section 6: (i) The fitted processes, $X$, $\hat{X}$, $Z$: 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*. Berlin: Springer Science & Business Media.

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.

Pflanzagl 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.