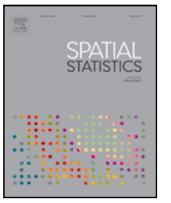
Spatial Statistics 60 (2024) 100823

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



A multivariate spatial and spatiotemporal ARCH Model

Philipp Otto

School of Mathematics and Statistics, University of Glasgow, United Kingdom

ARTICLE INFO

Keywords:
Conditional heteroscedasticity
Multivariate spatiotemporal data
QML estimator
Real-estate prices
Volatility clustering

ABSTRACT

This paper introduces a multivariate spatiotemporal autoregressive conditional heteroscedasticity (ARCH) model based on a vec-representation. The model includes instantaneous spatial autoregressive spill-over effects, as they are usually present in geo-referenced data. Furthermore, spatial and temporal cross-variable effects in the conditional variance are explicitly modelled. We transform the model to a multivariate spatiotemporal autoregressive model using a log-squared transformation and derive a consistent quasi-maximum-likelihood estimator (QMLE). For finite samples and different error distributions, the performance of the QMLE is analysed in a series of Monte-Carlo simulations. In addition, we illustrate the practical usage of the new model with a real-world example. We analyse the monthly real-estate price returns for three different property types in Berlin from 2002 to 2014. We find weak (instantaneous) spatial interactions, while the temporal autoregressive structure in the market risks is of higher importance. Interactions between the different property types only occur in the temporally lagged variables. Thus, we see mainly temporal volatility clusters and weak spatial volatility spillovers.

1. Introduction

In general, spatiotemporal processes can be represented as multivariate time series. However, when analysing spatial and spatiotemporal data, one has to account for one key difference compared to multivariate time series. Due to their spatial nature, geographical proximity between the observations induces instantaneous interactions between them. This is commonly known as Tobler's first law of geography: "Everything is related to everything, but near things are more related than distant things" (Tobler, 1970). This applies not only to the mean behaviour of the data but also to their variance. Thus, spatiotemporal models should always allow for instantaneous spatial interactions.[1] Particularly in the case of high-resolution data, it is commonly noted that smaller spatial units exhibit greater heterogeneity, thereby elevating the local variance, in contrast to larger aggregated spatial units that tend to display greater homogeneity (Espa et al., 1996).

This paper introduces a multivariate spatial and spatiotemporal autoregressive conditional heteroscedasticity (spatial ARCH, briefly spARCH) model. Using a vector representation, the approach follows the same logic as classical time-series vec-ARCH models (cf. Engle and Kroner 1995, see Silvennoinen and Teräsvirta 2009 for an overview on multivariate GARCH models) but has an additional spatial dimension. In that sense, the model aligns with the spatial ARCH models of Otto et al. (2018), Sato and Matsuda (2021), Otto and Schmid (2019). We call the new multivariate, spatiotemporal ARCH process vec-spARCH. All these approaches trace back to the seminal papers of Engle (1982) and Bollerslev (1986). In contrast to previous multivariate spatiotemporal GARCH

E-mail address: philipp.otto@glasgow.ac.uk.

[1] When the contemporaneous spatial dependence is neglected, as often seen in economic models or applications, in favour of the theoretical advantages offered by (multivariate) time series models – they benefit from a clear causal ordering, where only past information can influence the future observation – and only temporally lagged spatial interactions are included, the inherent spatial dependence gets inadvertently projected onto these temporally delayed interactions. This could lead to an incorrect interpretation that economic agents respond to information from neighbouring areas with a temporal delay.

https://doi.org/10.1016/j.spasta.2024.100823
Received 26 November 2023; Received in revised form 11 March 2024; Accepted 19 March 2024
Available online 2 April 2024

models (e.g., Borovkova and Lopuhaa 2012, Caporin and Paruolo 2015), we allow for contemporaneous/instantaneous dependence over space at the same time point, which is important for spatial models. Thus, our multivariate vec-spARCH model can also be applied for purely spatial data when there are spatial volatility clusters (i.e., clustered regions of high/low volatilities).

From an economic perspective, spatiotemporal GARCH models are motivated by the fact that concurrent decisions made by market participants lead to immediate spatial connections. For example, these decisions can range from trading multiple stocks within a financial network to making purchases in the real estate market. In the context of real estate, price and market risk inherently depend on spatial factors due to geographical limitations and the shared reliance on local amenities. This means that buyers often limit their search to specific localities, thereby introducing natural spatial dependencies. Moreover, in finance, the variance is commonly used to evaluate the risk of an investment (over time), with high variances signalling a higher uncertainty. Likewise, risk assessments vary geographically, highlighting the importance of spatial GARCH models for studies focusing on regional dynamics. Here, the evaluation of risk is influenced by the specific location and the associated risks of neighbouring areas. Another promising area for multivariate spatiotemporal ARCH models is the analysis of European electricity production from different sources, which is highly correlated across space/time and the energy sources, esp. renewable energy sources, or the analysis of electricity prices. In the latter case, it comes to pronounced spatial dependence in the price variations/uncertainties due to non-homogenous electricity networks (i.e., some regions are "closer" connected than others). We refer the interested reader to Ziel and Weron (2018), Berrisch and Ziel (2024).

Generally, univariate spatiotemporal models can be considered as multivariate time series models with an unknown covariance matrix, which has to be estimated. This covariance matrix represents the above-mentioned instantaneous interaction across space. Since the number of variables (i.e., the number of cross-sectional or spatial units) is typically large compared to the time points,[2] the full covariance cannot directly be estimated, but a certain structure is implied. The most straightforward structure assumes the same degree of correlation between all spatial units as in the dynamic equicorrelation model of Engle and Kelly (2012). However, this approach violates Tobler's first law and is typically not applicable in spatial settings, where the dependence declines with increasing distance between the locations. Thus, in spatial ARCH and GARCH models, the dependence structure is described with a so-called spatial weights matrix, which specifies the relation between the locations, similar to an adjacency matrix in network models.

In contrast to multivariate GARCH models that have two dimensions (i.e., the temporal dimension and the cross-variable dimension), the novel multivariate, spatiotemporal ARCH process has one additional dimension. To be precise, the process has a temporal, a cross-sectional, and a cross-variable dimension. It is worth noting that the cross-sectional dimension represents the spatial domain and is, therefore, at least a two-dimensional space. Moreover, using the vec-representation, the vec-spARCH is nested in the above-mentioned spatial GARCH proposed by Otto and Schmid (2019), so their results can also be directly applied. Alternative models that allow instantaneous spatial interactions in the variance are spatial stochastic volatility models, as proposed by Taspinar et al. (2021).

The vec-spARCH process distinguishes between three different effects: (1a) instantaneous spatial effects of the same variables, (1b) instantaneous cross-variable spatial effects, (2a) temporal autoregressive effects of the same variables, (2b) cross-variable temporal autoregressive effects, and (3) variable-specific unconditional volatility levels. Each effect is described by a parameter matrix or vector, for which we derive a quasi-maximum-likelihood (QML) estimator. For estimation, a logarithmic transformation of the vec-spARCH is applied (cf. Robinson 2009), such that the model can be represented as a multivariate spatiotemporal autoregressive model of the transformed quantity. The consistency of the QML estimator has been shown by Yang and Lee (2017) for a multivariate spatial autoregressive model (i.e., without temporal dimension) and by Yu et al. (2008) for a univariate spatiotemporal autoregressive process. Under certain regularity conditions, which are commonly used in spatial econometrics, we show the identifiability and consistency of the estimators.

In practice, spatiotemporal ARCH models are particularly important because an ARCH error process can also account for variation due to latent factors. In particular, for small spatial units, it is often difficult to quantify influential factors with the same spatial resolution. For instance, the average income of households in small spatial units, e.g. postal-code areas, does not necessarily reflect these particular units' economic power because people's daily cycles usually span multiple small spatial units. People do not necessarily live where they work or spend most of their time. In such cases, spatial and spatiotemporal ARCH models are important error distributions of any model to account for unobservable factors. This again highlights the importance of instantaneous spatial effects because latent factors would immediately affect the outcome variable, but not via temporally lagged relations.

The remainder of the paper is structured as follows. Firstly, we introduce the multivariate modelling framework and discuss how the model can be transformed into a regular univariate spatiotemporal process. Further, we derive the logarithmic likelihood and show the consistency for the QML estimator under several regularity assumptions that are usually met in spatial statistics. Secondly, we analyse the finite-sample performance of the proposed estimator for several model specifications and two different error distributions, namely standard normal and heavy-tailed errors (t_3-distributed). Thirdly, a real-world application is presented, for which we show that it is important to account for instantaneous spatial interactions and cross-variable correlations. To be precise, Berlin real-estate prices of three different property types are analysed, and we find weak contemporaneous spatial interactions, even though the temporal effects dominate them. These interdependencies are more pronounced when the spatial units and time intervals are small. Finally, Section 5 concludes the paper with a summary and brief outlook on future research and potential fields of application.

[2] Note that the number of unknown parameters is order $\mathcal{O}(n^2)$ when n is the number of spatial locations.

2. Multivariate spatiotemporal ARCH model

In spatial statistics, autoregressive dependence needs to be contemporaneous. From an economic perspective, contemporaneous interactions arise due to the simultaneity of decisions of economic agents, e.g., on a real estate market, property purchase decisions are done also using information from nearby locations (i.e., if the market risks are high in adjacent locations, agents on the real estate market will instantaneously adjust their own risk evaluation). A widespread risk measure is the variance of the process. That is, no time lag is required for shocks to affect neighbouring locations. Instead, we assume that the conditional variance can vary over space depending on the realised variance at adjacent locations. This results in spatial clusters of high and low variances. Considering the real estate market example in the multivariate setting, changes in the risks of different property types (e.g., condominiums, one-family houses, terraced houses, etc.) will typically affect each other.

For previous univariate or multivariate spatiotemporal GARCH models, such as proposed by Borovkova and Lopuhaa (2012), Caporin and Paruolo (2015), Hølleland and Karlsen (2020), spatial spill-overs could only occur after one time instance. In other words, the conditional variance at each location depends on the past squared observations at the same location and its neighbours, but not on their neighbouring locations at the same time point. This is the fundamental difference between multivariate time series models covering spatiotemporal data and approaches from spatial statistics.

Moreover, if the model is applied to the errors of any other spatiotemporal regression model, the autoregressive conditional heteroscedasticity can capture locally varying model uncertainties, opening various fields of applications apart from economics, such as environmental science or epidemiology.

2.1. Model specification

Assume that $\left\{\boldsymbol{Y}_t(\boldsymbol{s}) \in \mathbb{R}^p : \boldsymbol{s} \in D_{\boldsymbol{s}} \subset \mathbb{R}^q, t \in \mathbb{Z}\right\}$ is a p-variate spatiotemporal stochastic process in a q-dimensional space $D_{\boldsymbol{s}}$ with positive volume (cf. Cressie and Wikle 2011) across time. More precisely, the process is observed at n different sites $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n$, i.e., at each location $\boldsymbol{s}_i$ and time point t we observe a vector $\boldsymbol{Y}_t(\boldsymbol{s}_i) = (Y_{1,t}(\boldsymbol{s}_i), \dots, Y_{p,t}(\boldsymbol{s}_i))'$. Moreover, let $\boldsymbol{Y}_{j,t} = (Y_{j,t}(\boldsymbol{s}_1), \dots, Y_{j,t}(\boldsymbol{s}_n))'$ the vector of the jth characteristic at all locations and $\mathbf{Y}_t = (\boldsymbol{Y}_{1,t}, \dots, \boldsymbol{Y}_{p,t})$ be an $n \times p$ matrix of all observations $\left(Y_{j,t}\left(\boldsymbol{s}_i\right)\right)_{i=1,\dots,n, j=1,\dots,p}$ at time point t. Suppose that the process is observed for $t = 1, \dots, T$. It is worth mentioning that a multivariate spatial log-ARCH model is present if $T = 1$, and classical time-series log-ARCH models are also nested if $D_{\boldsymbol{s}}$ is a singleton (i.e., $n = 1$).

Univariate spatial and spatiotemporal ARCH and log-ARCH models have been introduced by Otto et al. (2018) and Sato and Matsuda (2017), respectively. Moreover, Otto and Schmid (2019) generalised the model in a unified framework nesting spatial GARCH, E-GARCH, and log-GARCH models. In this paper, we follow the idea of the symmetric spatial log-GARCH model of Sato and Matsuda (2021), which includes elements of GARCH and E-GARCH models but does not coincide with one or the other even if $D_{\boldsymbol{s}}$ consist of only a single location (i.e., the classical time series case). More precisely, the link function between the spatial volatility term is logarithmic, like for E-GARCH models, while the volatility term depends on some transformation of the squared observed process (similar to GARCH models). In contrast to time-series models, in which the temporal lag is clearly defined by the past observations and future observations are not allowed to influence the current observation, there are complex interdependencies in spatial settings and there is no causal relation between the observations anymore. For instance, with two locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ (i.e., $n = 2$), location $\boldsymbol{s}_1$ would influence $\boldsymbol{s}_2$ at each time point and vice versa. This would correspond to a simultaneous influence from future and past values in a time series context. Thus, for direct generalisation of GARCH or E-GARCH models like in Otto et al. (2018), Otto and Schmid (2019), difficult assumptions for the existence or invertibility of the process are required in the general case. In addition, existing software could be used directly with some adaptations for the spatiotemporal case (see Otto 2019).

The multivariate spatiotemporal ARCH model (vec-spARCH) is given by

$$\left\{\begin{array}{rcl} \mathbf{Y}_t & = & \mathbf{H}_t^{(1/2)} \circ \boldsymbol{\Xi}_t \quad \text{with} \\ \mathbf{H}_t^{(\ln)} & = & \mathbf{A} + \mathbf{W}\,\mathbf{Y}_t^{(\ln,2)}\,\boldsymbol{\Psi} + \mathbf{Y}_{t-1}^{(\ln,2)}\,\boldsymbol{\Pi}, \end{array}\right. \tag{1}$$

where $\mathbf{H}_t^{(1/2)} = \left(h_{j,t}^{1/2}(\boldsymbol{s}_i)\right)_{i=1,\dots,n, j=1,\dots,p}$ is the $n \times p$-dimensional matrix of the element-wise square roots of $h_{j,t}$, $\mathbf{H}_t^{(\ln)} = \left(\ln h_{j,t}(\boldsymbol{s}_i)\right)_{i=1,\dots,n, j=1,\dots,p}$ is the matrix of all natural logarithms of $h_{j,t}(\boldsymbol{s}_i)$, and $\mathbf{Y}_t^{(\ln,2)} = \left(\ln Y_{j,t}^2(\boldsymbol{s}_i)\right)_{i=1,\dots,n, j=1,\dots,p}$ denotes the $n \times p$-dimensional matrix of log-squared observations. In all that follows, the superscript means that the transformation is applied element-wise to each matrix entry. The matrix $\mathbf{H}_t$ is the spatial equivalent of the conditional volatility (see Otto et al. 2019), i.e., the volatility given all past and neighbouring observations (more precisely, log-squared observations as a measure of volatility). Thus, it can serve as a spatiotemporal risk measure on economic markets, e.g., real estate market risks for each location, time point and property type, as we will demonstrate in Section 4. The Hadamard product is denoted by $\circ$. Moreover, the $n \times p$-dimensional matrix of disturbances is denoted by $\boldsymbol{\Xi}_t = (\boldsymbol{\varepsilon}_{1,t}, \dots, \boldsymbol{\varepsilon}_{p,t})$ with independent and identically distributed random vectors $\boldsymbol{\varepsilon}_{j,t} = (\varepsilon_{j,t}(\boldsymbol{s}_1), \dots, \varepsilon_{j,t}(\boldsymbol{s}_n))'$ with $E(\boldsymbol{\varepsilon}_{j,t}) = \mathbf{0}$ and $Cov(\boldsymbol{\varepsilon}_{j,t}) = \mathbf{I}_n$ for all $j = 1, \dots, p$ and $t = 1, \dots, T$, where $\mathbf{I}_n$ is the n-dimensional identity matrix. The $n \times p$-dimensional intercept matrix of the volatility equations is denoted by $\mathbf{A}$. The elements are the constant terms of the conditional volatilities of each response variable for all locations. The weight $n \times n$-dimensional matrix $\mathbf{W}$ defines the spatial dependence structure, i.e., which locations are considered to be adjacent. Moreover, the cross-variable spatial effects are represented by the off-diagonal elements of $\boldsymbol{\Psi}$, and the temporally lagged cross-variable effects are given by the off-diagonal elements of $\boldsymbol{\Pi}$. Both matrices have dimension $p \times p$. In addition, the own-variable spatial and temporal autoregressive ARCH effects are summarised by the diagonal entries of $\boldsymbol{\Psi}$ and $\boldsymbol{\Pi}$, respectively.

The multivariate spatiotemporal ARCH model can be written as a multivariate spatiotemporal autoregressive process by applying a log-squared transformation,

$$\mathbf{Y}_t^{(\ln,2)} = \mathbf{H}_t^{(\ln)} + \boldsymbol{\Xi}_t^{(\ln,2)}.$$

Then, we get that

$$\mathbf{Y}_t^{(\ln,2)} = \mathbf{A} + \mathbf{W}\mathbf{Y}_t^{(\ln,2)}\boldsymbol{\Psi} + \mathbf{Y}_{t-1}^{(\ln,2)}\boldsymbol{\Pi} + \boldsymbol{\Xi}_t^{(\ln,2)}.$$

With $\mathbf{U}_t = \boldsymbol{\Xi}_t^{(\ln,2)} - E\left(\boldsymbol{\Xi}_t^{(\ln,2)}\right)$ and $\tilde{\mathbf{A}} = \mathbf{A} + E\left(\boldsymbol{\Xi}_t^{(\ln,2)}\right)$, the model can be rewritten as

$$\mathbf{Y}_t^{(\ln,2)} = \tilde{\mathbf{A}} + \mathbf{W}\mathbf{Y}_t^{(\ln,2)}\boldsymbol{\Psi} + \mathbf{Y}_{t-1}^{(\ln,2)}\boldsymbol{\Pi} + \mathbf{U}_t. \tag{2}$$

Assuming a standard normal error matrix $\boldsymbol{\Xi}_t$, $E\left(\boldsymbol{\Xi}_t^{(\ln,2)}\right)$ is the expectation of a log-Gamma distribution, i.e., the expectation of each element is equal to $E\left(\ln \varepsilon_{j,t}^2(\boldsymbol{s}_i)\right) = -\gamma - \log(2) \approx -1.27$ for all i, j, t. Then, $\mathbf{A}$ can be determined from $\tilde{\mathbf{A}}$, which facilitates the interpretation. Note that we do not need the normality assumption later to derive the consistency of the QML estimator. According to (2), we see that the vec-spARCH model coincides with a multivariate spatiotemporal autoregressive process of the log-squared transformed process $\mathbf{Y}_t^{(\ln,2)}$. For the multivariate but purely spatial case, Yang and Lee (2017) has derived conditions for identification and the consistency and asymptotic normality of a QML estimator. Furthermore, Yu et al. (2008) derive asymptotic results for a QML estimator of the spatiotemporal but univariate process ($p = 1$) when both n and T are large. We combine these two results to propose a QML estimator for the spatiotemporal, multivariate ARCH model in Section 2.3.

Moreover, analogue to multivariate vec-ARCH time-series model of Engle and Kroner (1995), we can rewrite (1) to get the vectorised form

$$vec(\mathbf{H}_t^{(\ln)}) = vec(\mathbf{A}) + (\boldsymbol{\Psi}' \otimes \mathbf{W})vec(\mathbf{Y}_t^{(\ln,2)}) + (\boldsymbol{\Pi}' \otimes \mathbf{I}_n)vec(\mathbf{Y}_{t-1}^{(\ln,2)}). \tag{3}$$

The Kronecker product is denoted by $\otimes$, and vec is the vectorisation of a matrix. Interestingly, using such vec-representation, one can see that the multivariate ARCH model is a special case of a (univariate) np-dimensional spatiotemporal autoregressive model with a weight matrix $\boldsymbol{\Psi}' \otimes \mathbf{W}$. Thus, spatial GARCH and E-GARCH models can be constructed similarly, and all results of Otto and Schmid (2019) can directly be applied. However, this will not be the focus of this paper.

Below, let $\ddot{\boldsymbol{Y}}_t = vec(\mathbf{Y}_t^{(\ln,2)})$ for an easier notation. With $\mathbf{S}_{np} = \mathbf{I}_{np} - \boldsymbol{\Psi}' \otimes \mathbf{W}$, we can derive the sample log-likelihood for the spatiotemporal model given by (2) with T time points, i.e.,

$$\begin{aligned}\ln \mathcal{L}(\mathbf{A}, \boldsymbol{\Psi}, \boldsymbol{\Pi}|\mathbf{Y}_0) = &-\frac{Tnp}{2}\ln(2\pi) + \frac{n\ln\sigma_u^2}{2p} + \frac{T}{np}\ln|\mathbf{S}_{np}| \\ &- \frac{1}{2np\sigma_u^2}\sum_{t=1}^{T}\left[\mathbf{S}_{np}\ddot{\boldsymbol{Y}}_t - vec(\tilde{\mathbf{A}}) - (\mathbf{I}_p \otimes \mathbf{Y}_{t-1}^{(\ln,2)})vec(\boldsymbol{\Pi})\right]' \\ &\qquad \times \left[\mathbf{S}_{np}\ddot{\boldsymbol{Y}}_t - vec(\tilde{\mathbf{A}}) - (\mathbf{I}_p \otimes \mathbf{Y}_{t-1}^{(\ln,2)})vec(\boldsymbol{\Pi})\right],\end{aligned}$$

where σ_u^2 is the variance of the transformed errors (i.e., all elements of $\mathbf{U}_t$), which is a known quantity in our case. Otherwise, $\tilde{\mathbf{A}}$ would not be identifiable. Furthermore, for standard normal errors $\boldsymbol{\Xi}_t$, we get $\sigma_u^2 = \psi(1/2) \approx 4.93$, where ψ denotes the trigamma function. It is worth mentioning that we derived the Gaussian log-likelihood, but the transformed errors $\mathbf{U}_t$ are, in fact, skewed because of the logarithmic transformation. In the following Section 2.2, however, we suppose much weaker conditions on the moments of $\mathbf{U}_t$, which are fulfilled in the case of standard normal errors $\boldsymbol{\Xi}_t$ but also for other distributional assumptions.

Below, let $\mathbf{A}_0$, $\boldsymbol{\Psi}_0$, and $\boldsymbol{\Pi}_0$ be the true data-generating parameters of $\mathbf{A}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Pi}$, respectively. Furthermore, $\mathbf{S}_{np0} = \mathbf{I}_{np} - \boldsymbol{\Psi}_0' \otimes \mathbf{W}$. With $E(\ddot{\boldsymbol{Y}}_t) = \mathbf{S}_{np0}^{-1}\left(vec(\tilde{\mathbf{A}}_0) + \boldsymbol{\Pi}_0' \otimes \mathbf{Y}_{t-1}^{(\ln,2)}\right)$, we get the expected log-likelihood as

$$\begin{aligned}E(\ln \mathcal{L}(\mathbf{A}, \boldsymbol{\Psi}, \boldsymbol{\Pi}|\mathbf{Y}_0)) = &-\frac{Tnp}{2}\ln(2\pi) + \frac{n\ln\sigma_u^2}{2p} + \frac{T}{np}\ln|\mathbf{S}_{np}| \\ &- \frac{1}{2np\sigma_u^2}\sum_{t=1}^{T}\left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(vec(\tilde{\mathbf{A}}_0 - \tilde{\mathbf{A}}) + ((\boldsymbol{\Pi}_0' - \boldsymbol{\Pi}') \otimes \mathbf{I}_n)\ddot{\boldsymbol{Y}}_{t-1})\right]' \\ &\qquad \times \left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(vec(\tilde{\mathbf{A}}_0 - \tilde{\mathbf{A}}) + ((\boldsymbol{\Pi}_0' - \boldsymbol{\Pi}') \otimes \mathbf{I}_n)\ddot{\boldsymbol{Y}}_{t-1})\right] \\ &- \frac{T}{2np\sigma_u^2}tr\left(\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'\right).\end{aligned}$$

2.2. Assumptions and parameter space

Below, we discuss important model assumptions that are needed to show the consistency of the QML estimators.

Assumption 1. Suppose that each element of $\boldsymbol{\Xi}_t$ is not equal to zero with probability one for all $t = 1, \ldots, T$.

To be able to apply the log-squared transformation of the observed process, we must assume that the response is not equal to zero with probability one. This is true for any continuous error process $\boldsymbol{\Xi}_t$. In practice, there is sometimes an excess of zeros due to missing values. In this case, a small number is often added to the zero values, such that the logarithmic transformation is feasible (see, e.g., Francq et al. 2013, Francq and Zakoian 2011). If there are empirically observed zeros—still occurring with probability zero, though—Sucarrat and Escribano (2018) proposed an expectation-maximisation algorithm for estimation in the time-series case. This would be an interesting extension for future research. Further, we need some basic assumptions on the transformed error process $\mathbf{U}_t$ to apply the results of Yang and Lee (2017) and Yu et al. (2008).

Assumption 2. Assume that each column $j = 1, \dots, p$ of $\mathbf{U}_t$, say $\boldsymbol{U}_{j,t} = (u_{j,t}(\boldsymbol{s}_1), \dots, u_{j,t}(\boldsymbol{s}_n))'$, is a random vector with zero mean and covariance $\sigma_u^2 \mathbf{I}_n$ that is i.i.d. across time. Additionally suppose that $E(|u_{t,ik}u_{t,il}u_{t,ir}u_{t,is}|^{1/\delta}) < \infty$ for all $i = 1, \dots, n$, $t = 1, \dots, T$ and $k, l, r, s = 1, \dots, p$ and some $\delta > 0$.

Moreover, the parameter space needs to be compact, as formulated in the following assumption, to prove the uniform convergence of the log-likelihood function.

Assumption 3. The parameter spaces for $\tilde{\mathbf{A}}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Pi}$ are compact sets, and all parameters in their interior generate a stable process. Moreover, the data-generating parameters $\tilde{\mathbf{A}}_0$, $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Pi}_0$ are in the interior of corresponding parameter space.

The key question of this assumption is the stability of the process. To analyse the stability, we rewrite the model in its reduced form as

$$
\begin{aligned}
\ddot{\boldsymbol{Y}}_t &= \mathbf{S}_{np}^{-1}(vec(\tilde{\mathbf{A}}) + (\boldsymbol{\Pi}' \otimes \mathbf{I}_n)\ddot{\boldsymbol{Y}}_{t-1} + vec(\mathbf{U}_t)) \\
&= \mathbf{S}_{np}^{-1}vec(\tilde{\mathbf{A}}) + \mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n)\left[\mathbf{S}_{np}^{-1}vec(\tilde{\mathbf{A}}) + \mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n)\ddot{\boldsymbol{Y}}_{t-2} + \mathbf{S}_{np}^{-1}vec(\mathbf{U}_{t-1})\right] + \mathbf{S}_{np}^{-1}vec(\mathbf{U}_t) \\
&\;\;\vdots \\
&= (\mathbf{I}_{np} + \mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n) + \cdots + (\mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n))^j)vec(\tilde{\mathbf{A}}) + (\mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n))^j\ddot{\boldsymbol{Y}}_{t-j} \\
&\quad + \sum_{i=0}^{j-1}(\mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n))^i vec(\mathbf{U}_{t-i})
\end{aligned}
$$

Hence, the stability of the process does not only depend on the temporal parameter matrix $\boldsymbol{\Pi}$ but also on the weight matrix $\mathbf{W}$ (via $\mathbf{S}_{np}^{-1}$), which is assumed to be constant over time. If the above series converges, we get a stable and stationary process.

Proposition 1. *Suppose that $\mathbf{S}_{np} = \mathbf{I}_{np} - \boldsymbol{\Psi}' \otimes \mathbf{W}$ is invertible. If all eigenvalues of $\mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n)$ are smaller than one, the multivariate spatiotemporal ARCH process is stable across time.*

Note that each stable spatiotemporal ARCH process is also weakly stationary. Furthermore, the boundary region of $\boldsymbol{\Psi}$ where $|\mathbf{S}_{np}| = 0$ can be problematic in practice. However, as long as the true parameter $\boldsymbol{\Psi}_0$ is bounded away from this region, the maximisation algorithm will not get to these boundaries with a large probability (see also Yang and Lee 2017).

The stability condition mainly depends on the parameter matrices $\boldsymbol{\Pi}$ and $\boldsymbol{\Psi}$ (via $\mathbf{S}_{np}^{-1}$). Let $\rho(\cdot)$ be the spectral radius and $\|\cdot\|$ be a matrix norm. Using the sufficient condition that

$$
\rho(\mathbf{S}_{np}^{-1}(\boldsymbol{\Pi}' \otimes \mathbf{I}_n)) \leq \rho(\mathbf{S}_{np}^{-1})\rho(\boldsymbol{\Pi}' \otimes \mathbf{I}_n) \leq \|\mathbf{S}_{np}^{-1}\| \cdot \|\boldsymbol{\Pi}' \otimes \mathbf{I}_n\| < 1\,, \tag{4}
$$

we can obtain suitable ranges of the parameter matrices (see Horn and Johnson, 2012, spectral radius theorem). Since any eigenvalue of a Kronecker product is some product of the eigenvalues of the two matrices, the eigenvalues of $(\boldsymbol{\Pi}' \otimes \mathbf{I}_n)$ are less than one if the eigenvalues of $\boldsymbol{\Pi}$ are less than one. For example, this is the case if the radius $\rho(\boldsymbol{\Pi})$ is smaller than one or if $\boldsymbol{\Pi}$ is strictly diagonally dominant, i.e., the diagonal entry of each row—the temporal autoregressive effect of the own component—is smaller than the sum of the magnitudes of all the other entries in this row—the cross-component temporal autoregressive effects (see Horn and Johnson, 2012, Gershgorin circle theorem). Moreover,

$$
\begin{aligned}
\|\mathbf{S}_{np}^{-1}\| &= \|\mathbf{I}_{np} + \boldsymbol{\Psi}' \otimes \mathbf{W} + (\boldsymbol{\Psi}' \otimes \mathbf{W})^2 + \cdots\| && (5)\\
&\leq \|\mathbf{I}_{np}\| + \|\boldsymbol{\Psi}' \otimes \mathbf{W}\| + \|(\boldsymbol{\Psi}' \otimes \mathbf{W})\|^2 + \cdots && (6)\\
&= \frac{1}{1 - \|\boldsymbol{\Psi}' \otimes \mathbf{W}\|}\,. && (7)
\end{aligned}
$$

If the spectral radius of $\boldsymbol{\Psi}' \otimes \mathbf{W}$ is less than one, the spatial component is stable. For instance, if $\mathbf{W}$ is a row-standardised spatial weight matrix, implying that the largest eigenvalue of $\mathbf{W}$ is one, similar conditions can be considered for $\boldsymbol{\Psi}$ like for the temporal parameter matrix. If $\boldsymbol{\Psi}$ is strictly diagonally dominant, i.e., the own-component spatial effects are smaller than the sum of all absolute cross-component spatial effects, the spectral radius is smaller than one and $\|\boldsymbol{\Psi}' \otimes \mathbf{W}\| < 1$. Thus, we can the condition that $\|\boldsymbol{\Pi}\| / (1 - \|\boldsymbol{\Psi}\|) < 1$, under which the process is stable. Note that this is a more restrictive condition than in Proposition 1.

Assumption 4. The row and column sums of $\mathbf{W}$ in absolute values are uniformly bounded in n. Moreover, $\mathbf{S}_{np}$ is invertible for all possible matrices $\boldsymbol{\Psi}$ in the parameter space and $\mathbf{S}_{np}^{-1}$ is uniformly bounded in absolute row and column sums.

Assumption 4 is classical in spatial statistics to obtain a stable process across space (cf. Yang and Lee 2017, Kelejian and Prucha 1998, Lee 2004). Here, we could adopt the assumption as formulated in Yang and Lee (2017) for multivariate spatial autoregressive models. Together with Proposition 1, we obtain a stable process across space and time. In practice, the spatial weight matrices are often standardised to meet these regularity conditions, e.g. the most widely adopted row-wise standardisation.

Assumption 5. Let n be a non-decreasing function of T and $T \to \infty$.

Assumption 5 implies that $n, T \to \infty$ simultaneously.

2.3. Consistency of the QML estimator

Due to the presence of endogenous variables, i.e., the instantaneous spatial interactions, the identification of spatial models is generally more difficult than in the strict time-series case, where all spatiotemporal interaction may only occur after one time lag. Thus, we initially focus on the identification of the parameters which is needed for the consistency of the QML estimator in the following Theorem 1. Since the identification is inherent with the spatial dimension of the model, we could follow the same strategy as in Yang and Lee (2017) for multivariate spatial autoregressive models. The identification is based on the information inequality, as proposed by Rothenberg (1971).

Proposition 2. *If the Assumptions 1–5 are fulfilled, then* $\tilde{\mathbf{A}}_0$, $\mathbf{\Psi}_0$ *and* $\mathbf{\Pi}_0$ *are uniquely identifiable.*

For the identification, we make use of the fact that the spatial dependence is constant across time, and the temporal dependence is constant for all spatial locations. If either of them varies in space or time, further identifying information would be needed. Moreover, in contrast to Yang and Lee (2017), the errors are uncorrelated by definition, and the error variance is supposed to be known. The assumption of an uncorrelated error process is essential for GARCH models for the identification of the parameters in the conditional volatility equation, i.e., the so-called GARCH effects. Moreover, the assumption of a known error variance σ_u^2 is, of course, restrictive (see also Francq and Zakoian 2011, Brockwell and Davis 2006), and it is often difficult to choose an appropriate value. Ex-post scale adjustments have been proposed for time series to circumvent this assumption (see Bauwens and Sucarrat 2010, Sucarrat et al. 2016). However, this paper follows the classical approach and points to future research for these ex-post scale adjustments. Moreover, for the purely spatial case with $T = 1$, $\tilde{\mathbf{A}}_0$ must be constant across space to be identifiable. That is, $\tilde{\mathbf{A}}_0 = (\tilde{a}_{10}\mathbf{1}_n', \dots, \tilde{a}_{p0}\mathbf{1}_n')'$ is an $n \times p$-dimensional matrix with constant entries for each row. Similarly, we can see that for the classical time series case with $n = 1$ that $\tilde{\mathbf{A}}_0$ is an $1 \times p$ row vector $(\tilde{a}_{10}, \dots, \tilde{a}_{p0})$.

To estimate the parameters, we propose a QML estimator based on the log-likelihood function given by . That is, the parameters $\boldsymbol{\vartheta} = (vec(\tilde{\mathbf{A}})', vec(\mathbf{\Psi})', vec(\mathbf{\Pi})')$ can be estimated by

$$\hat{\boldsymbol{\vartheta}}_{nT} = \arg\max_{\boldsymbol{\vartheta} \in \Theta} \ln \mathcal{L}(\mathbf{A}, \mathbf{\Psi}, \mathbf{\Pi} | \mathbf{Y}_0),$$

where Θ is the parameter space that fulfils Assumption 3. It is worth noting that we need to condition on the observed vector at $t = 0$, $\mathbf{Y}_0$, because of the temporal autoregressive structure. The consistency of this QML estimator is summarised in the following theorem.

Theorem 1. *Under Assumptions 1–5,* $\boldsymbol{\vartheta}_0 = (vec(\tilde{\mathbf{A}}_0)', vec(\mathbf{\Psi}_0)', vec(\mathbf{\Pi}_0)')$ *can be uniquely identified and* $\hat{\boldsymbol{\vartheta}}_{nT} \xrightarrow{p} \boldsymbol{\vartheta}_0$.

All proofs can be found in the appendix.

3. Monte Carlo simulations

In the following section, we present the results of a series of simulations on the consistency of the parameters for finite samples. To give a first visual impression, we display a bivariate spatial ARCH process ($T = 1$, $n = 900$, Rook's contiguity matrix) with and without spatial cross-correlation in Fig. 1. For both examples, the spatial ARCH effects are equal to 0.5, a moderate level of spatial dependence. Therefore, spatial volatility clusters can be seen in both cases. They are indicated by a higher variance, that is, more intensely coloured pixels, whereas clusters of low variance are close to zero indicated by evenly grey-coloured pixels. Now, for the case with a cross-correlation of 0.35 (top panels), these clusters are aligned across the variables, while they are located at different positions in the lower panels with zero cross-correlation.

Our Monte Carlo simulation study simulated three different bivariate models (A, B, C) with two different error distributions (standard normal and t_3) with 1000 replications. For each combination, we successively increased the size of the spatial field $n \in \{25, 49, 100\}$ and the length of the time series $T \in \{30, 100, 200\}$. We simulated the process on a two-dimensional grid as visualised in Fig. 1, and the spatial weight matrix was chosen as a row-standardised Queen's contiguity matrix. The data-generating parameters of the three considered models are as follows:

(A) Spatiotemporal model with a weak spatial cross-correlation: $\mathbf{A}_0 = \mathbf{1}_n\mathbf{1}_p'$, $\mathbf{\Psi}_0 = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}$, and $\mathbf{\Pi}_0 = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.3 \end{pmatrix}$

(B) Spatiotemporal model without temporal dependence, but the same spatial dependence as for Model A: $\mathbf{A}_0 = \mathbf{1}_n\mathbf{1}_p'$, $\mathbf{\Psi}_0 = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}$, and $\mathbf{\Pi}_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

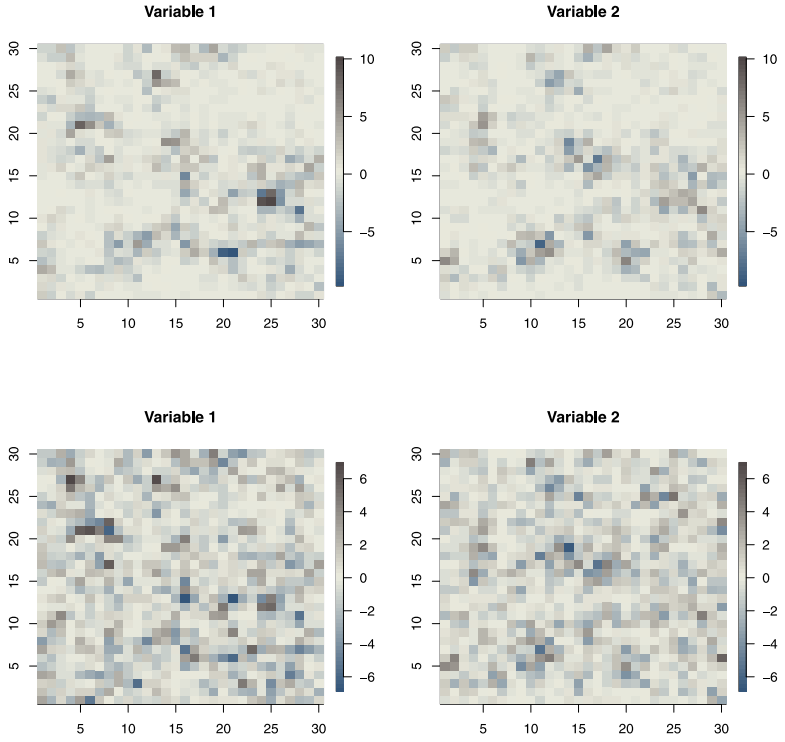


Fig. 1. Simulated random fields (first row: high cross correlation $\psi_{12} = \psi_{21} = 0.35$; second row: no cross correlation $\psi_{12} = \psi_{21} = 0$). The spatial ARCH coefficients are identical for all components and both settings, i.e., $\psi_{11} = \psi_{22} = 0.5$.

(C) Spatiotemporal model with pronounced cross-correlation and weak spatial correlation, same temporal autocorrelation like for Model A: $\mathbf{A}_0 = \mathbf{1}_n\mathbf{1}_p'$, $\mathbf{\Psi}_0 = \begin{pmatrix} 0.2 & 0.4 \\ 0.4 & 0.2 \end{pmatrix}$, and $\mathbf{\Pi}_0 = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.3 \end{pmatrix}$.

For the first simulated model, i.e., Model A with standard normal errors, the parameter estimates are depicted as a series of boxplots for the three increasing sizes $(n, T)' \in \{(25, 30)', (49, 100)', (100, 200)'\}$ in Fig. 2. In all cases, we see that the QML estimator consistently estimates the true values because the boxplots are getting more centred around zero. Moreover, we see the typical bias of the QML estimators for small spatial fields, which rapidly vanishes with increasing sample size. The same behaviour can be observed for all other settings and error distributions. The average bias and the root-mean-square errors (RMSE) of the estimators are reported in Tables 1 and 2, respectively. Both the absolute values of the bias and the RMSE are approaching zero if n and T are increasing.

4. Real-world example: Berlin real-estate prices

In the following section, we will show the application of the process to a real example. For this purpose, we model the changes in the average sales prices of undeveloped land, developed land and condominiums in Berlin. The data are monthly average prices per square metre of land or living space in each post-code region from 2002 to 2014. The average prices across all spatial locations are depicted in Fig. 3 as a time series process.

First of all, it must be noted that there are typically geographical dependencies in the housing market, unlike other financial markets where trading can take place regardless of location. One of the most important factors in a purchase decision is the property's location, whereby the surrounding neighbourhood also influences prices. This dependency is, in turn, influenced by road connections, infrastructure or public transport. Furthermore, the price in the past also plays a role, as is typical for all time series. The temporal proximity creates a causal statistical dependence that decreases the further one looks into the past. These dependencies are observed both in the price process and in the risks in terms of price changes. Moreover, in a regression context (i.e., if the real-estate prices are modelled directly), multivariate spatial ARCH models are appealing error models, especially when spatially correlated variables are omitted from the mean model. These omitted variables lead to statistical dependence in the error variance. This issue is pronounced in spatial statistics where key variables are often omitted due to unavailability or measurement challenges at small spatial scales, e.g., local income levels.

This motivates applying the proposed multivariate spatiotemporal ARCH process to property sales returns. More precisely, we analysed the logarithmic monthly returns of the average sales prices in each category for all $n = 190$ post-code regions in Berlin. All

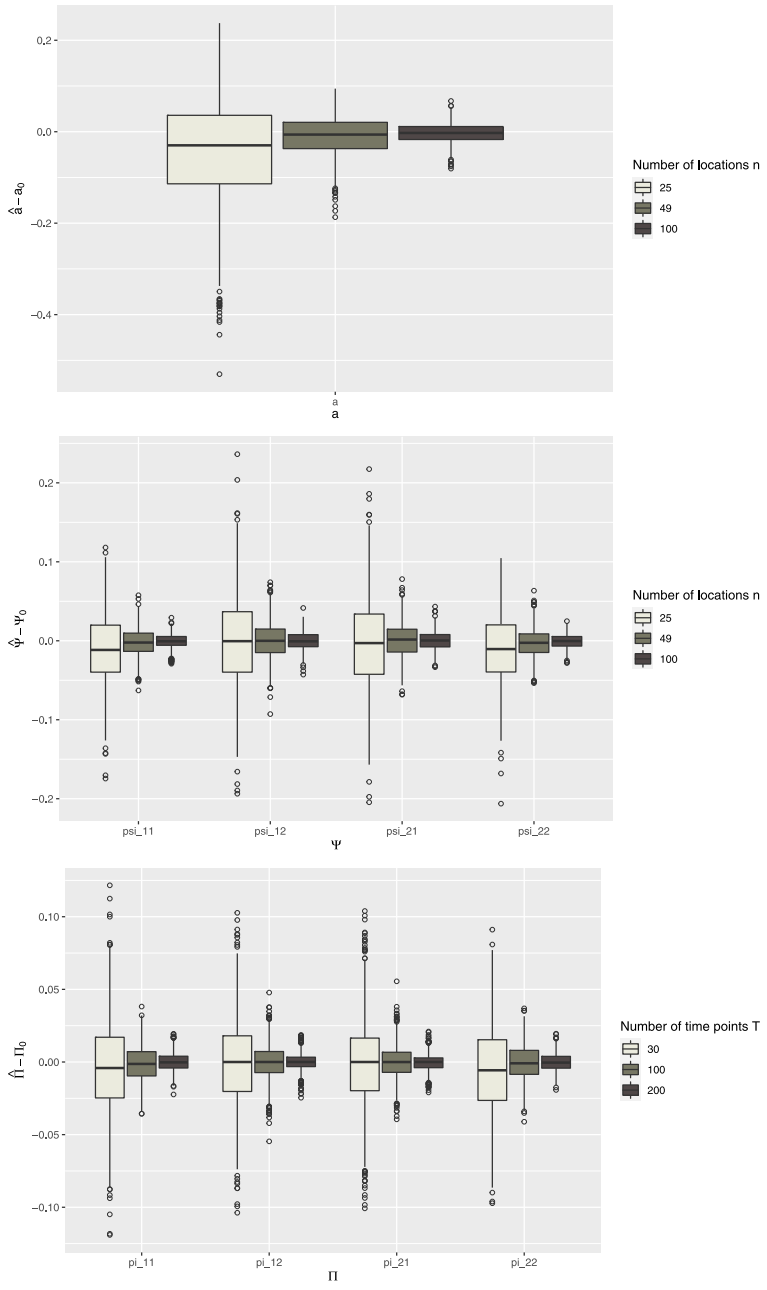


Fig. 2. Estimation performance of the QML estimator for Model A with standard normal errors. Top row: unconditional variance level a, centre: spatial coefficient matrix $\boldsymbol{\Psi}$, bottom: temporal coefficient matrix $\boldsymbol{\Pi}$. For each plot, we show the difference between the parameter estimate and the true data-generating parameter.

regions are displayed in Fig. 4 together with their spatial weights. Specifically, we have chosen $\mathbf{W}$ to be a queen contiguity matrix. Notice that all obtained results need to be interpreted in the light of this assumption, i.e., the estimated spatial parameters represent interactions to the directly adjacent postcode areas. The length of the time series is $T = 156$ and the process is $p = 3$-dimensional. To display the log-return process, Fig. 5 shows the average log returns across all locations in the temporal domain. Especially for the developed and undeveloped land, there were much fewer sales, such that the average returns are more volatile. In the case of no transactions in certain months and areas, we assumed that the average sales price did not change and, thus, the log returns are zero. These missing values (or zero returns), we imputed by randomly simulating from a normal distribution a mean zero and standard deviation of 0.0001. In contrast to sample mean or moving average imputations or replacing zeros by (small) constant values (see also Francq et al., 2013), this preserves the underlying evolution of the real price series. In future, a more detailed analysis, including a zero-transaction model, would be interesting, especially for smaller time granularities and spatial locations. In

Table 1
Average bias of the QML estimates for all considered settings.

Average bias		a	ψ_{11}	ψ_{21}	ψ_{12}	ψ_{22}	π_{11}	π_{21}	π_{12}	π_{22}
Model A, *data-generating parameters*		1	0.5	0.1	0.1	0.5	0.3	0	0	0.3
	$n = 25,\ T = 30$	−0.0464	−0.0115	−0.0021	−0.0003	−0.0112	−0.0036	−0.0010	−0.0008	−0.0055
Normal errors	$n = 49,\ T = 100$	−0.0093	−0.0024	0.0005	0.0002	−0.0028	−0.0013	−0.0002	−0.0002	−0.0004
	$n = 100,\ T = 200$	−0.0027	−0.0005	0.0001	−0.0001	−0.0006	−0.0001	−0.0003	0.0000	−0.0003
	$n = 25,\ T = 30$	0.0269	−0.0106	−0.0026	−0.0040	−0.0103	−0.0042	0.0006	−0.0035	−0.0050
t_3-distr. errors	$n = 49,\ T = 100$	0.0029	−0.0005	−0.0007	−0.0005	−0.0010	−0.0016	−0.0009	−0.0002	−0.0010
	$n = 100,\ T = 200$	0.0009	−0.0005	0.0004	−0.0006	−0.0003	−0.0001	0.0001	0.0001	0.0000
Model B, *data-generating parameters*		1	0.5	0.1	0.1	0.5	0	0	0	0
	$n = 25,\ T = 30$	0.0249	0.0143	0.0194	0.0189	0.0149	−0.0014	−0.0009	−0.0019	−0.0004
Normal errors	$n = 49,\ T = 100$	−0.0021	−0.0008	−0.0005	−0.0003	−0.0004	−0.0001	−0.0003	0.0007	−0.0004
	$n = 100,\ T = 200$	−0.0009	0.0001	−0.0005	0.0004	−0.0005	0.0001	0.0001	0.0000	−0.0002
	$n = 25,\ T = 30$	0.0088	0.0013	0.0006	−0.0019	−0.0011	−0.0026	−0.0017	−0.0003	0.0006
t_3-distr. errors	$n = 49,\ T = 100$	0.0017	0.0010	0.0011	−0.0008	−0.0005	−0.0002	0.0000	−0.0005	0.0001
	$n = 100,\ T = 200$	0.0001	0.0001	−0.0011	0.0017	−0.0003	0.0000	−0.0002	0.0002	0.0002
Model C, *data-generating parameters*		1	0.2	0.4	0.4	0.2	0.3	0	0	0.3
	$n = 25,\ T = 30$	−0.0448	−0.0075	−0.0057	0.0018	−0.0120	−0.0037	−0.0003	−0.0014	−0.0031
Normal errors	$n = 49,\ T = 100$	−0.0095	−0.0025	−0.0003	0.0002	−0.0022	−0.0002	−0.0009	−0.0003	−0.0008
	$n = 100,\ T = 200$	−0.0012	−0.0004	0.0001	0.0003	−0.0001	−0.0002	−0.0001	−0.0002	−0.0001
	$n = 25,\ T = 30$	0.0140	−0.0086	−0.0048	−0.0016	−0.0092	−0.0029	−0.0017	−0.0016	−0.0033
t_3-distr. errors	$n = 49,\ T = 100$	0.0023	−0.0018	−0.0003	−0.0006	−0.0011	0.0001	−0.0002	−0.0003	−0.0009
	$n = 100,\ T = 200$	0.0001	0.0001	−0.0011	0.0005	−0.0009	0.0000	−0.0001	0.0000	−0.0003

Table 2
Root-mean-square errors of the QML estimates for all considered settings.

RMSE		a	ψ_{11}	ψ_{21}	ψ_{12}	ψ_{22}	π_{11}	π_{21}	π_{12}	π_{22}
Model A, *data-generating parameters*		1	0.5	0.1	0.1	0.5	0.3	0	0	0.3
	$n = 25,\ T = 30$	0.1237	0.0456	0.0580	0.0589	0.0444	0.0326	0.0317	0.0313	0.0313
Normal errors	$n = 49,\ T = 100$	0.0442	0.0173	0.0218	0.0228	0.0176	0.0121	0.0119	0.0124	0.0122
	$n = 100,\ T = 200$	0.0221	0.0087	0.0117	0.0110	0.0089	0.0061	0.0060	0.0060	0.0060
	$n = 25,\ T = 30$	0.0948	0.0443	0.0718	0.0712	0.0433	0.0313	0.0292	0.0318	0.0336
t_3-distr. errors	$n = 49,\ T = 100$	0.0288	0.0167	0.0268	0.0256	0.0172	0.0129	0.0115	0.0117	0.0123
	$n = 100,\ T = 200$	0.0148	0.0086	0.0130	0.0130	0.0088	0.0059	0.0052	0.0050	0.0059
Model B, *data-generating parameters*		1	0.5	0.1	0.1	0.5	0	0	0	0
	$n = 25,\ T = 30$	0.1366	0.0912	0.1304	0.1291	0.0941	0.0336	0.0305	0.0295	0.0339
Normal errors	$n = 49,\ T = 100$	0.0252	0.0191	0.0372	0.0368	0.0183	0.0128	0.0124	0.0122	0.0130
	$n = 100,\ T = 200$	0.0120	0.0095	0.0201	0.0197	0.0092	0.0065	0.0058	0.0060	0.0066
	$n = 25,\ T = 30$	0.0717	0.0511	0.1308	0.1278	0.0489	0.0338	0.0302	0.0300	0.0350
t_3-distr. errors	$n = 49,\ T = 100$	0.0266	0.0187	0.0528	0.0540	0.0186	0.0129	0.0112	0.0110	0.0130
	$n = 100,\ T = 200$	0.0130	0.0099	0.0296	0.0296	0.0095	0.0064	0.0053	0.0054	0.0065
Model C, *data-generating parameters*		1	0.2	0.4	0.4	0.2	0.3	0	0	0.3
	$n = 25,\ T = 30$	0.1218	0.0598	0.0813	0.0794	0.0580	0.0350	0.0299	0.0305	0.0341
Normal errors	$n = 49,\ T = 100$	0.0433	0.0236	0.0313	0.0315	0.0235	0.0131	0.0119	0.0120	0.0129
	$n = 100,\ T = 200$	0.0222	0.0122	0.0152	0.0161	0.0117	0.0064	0.0059	0.0059	0.0066
	$n = 25,\ T = 30$	0.0889	0.0596	0.0819	0.0850	0.0597	0.0338	0.0305	0.0303	0.0323
t_3-distr. errors	$n = 49,\ T = 100$	0.0311	0.0233	0.0321	0.0334	0.0245	0.0130	0.0112	0.0118	0.0126
	$n = 100,\ T = 200$	0.0147	0.0118	0.0160	0.0163	0.0120	0.0063	0.0055	0.0055	0.0062

particular, for finer temporal and spatial resolutions, the GARCH effects appear to be more evident, but we would observe more zero transactions, especially for the undeveloped areas. Thus, for this study, we balance the time/spatial granularity with only a small percentage of zero transactions.

The estimated parameters of the multivariate spatiotemporal ARCH process are reported in Table 3 along with their empirical standard errors obtained as Cramer–Rao bounds based on the Hessian matrix of the log-likelihood. The intercept variance levels were assumed to be constant across space but vary with the property types, i.e., $\tilde{\mathbf{A}} = (\tilde{a}_1\mathbf{1}_n', \tilde{a}_2, \tilde{a}_3\mathbf{1}_n')'$. Bearing in mind that we have modelled monthly returns, we observe interesting results. First, spatial dependence is dominated by temporal dependence, which appears to be more important. Second, spatial spill-overs are positive (i.e., we observed clusters of higher variances/risks), but they are only significant for developed land. This highlights the importance of including instantaneous spatial interactions (as mentioned in the introduction). If spatial interactions were only allowed at the first temporal lag, these effects would be erroneously identified as temporally lagged spatial interactions (i.e., spatiotemporal effects). Hence, the resulting parameters for such models should be interpreted with caution. When increasing the temporal intervals from monthly to quarterly data, these spatial interactions will disappear. The same holds when grouping the spatial locations into larger areas. Thus, spatial GARCH models are particularly useful for small spatial units and time granularities (as it is also well-known in finance). Third, cross-variable spill-overs are only significant at the first temporal lag (i.e., after one time period). More precisely, we see significant interactions only between developed and undeveloped land, but not for condominium prices. It is important to remember that the spatial and temporal ARCH effects will

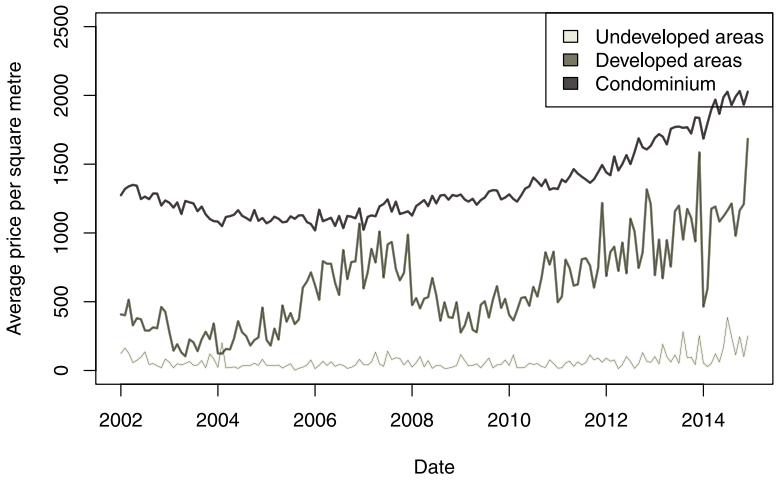


Fig. 3. Monthly average prices (Euro/m^2) of undeveloped, developed land, and condominium across 190 post-code regions in Berlin from January 2002 to December 2014.

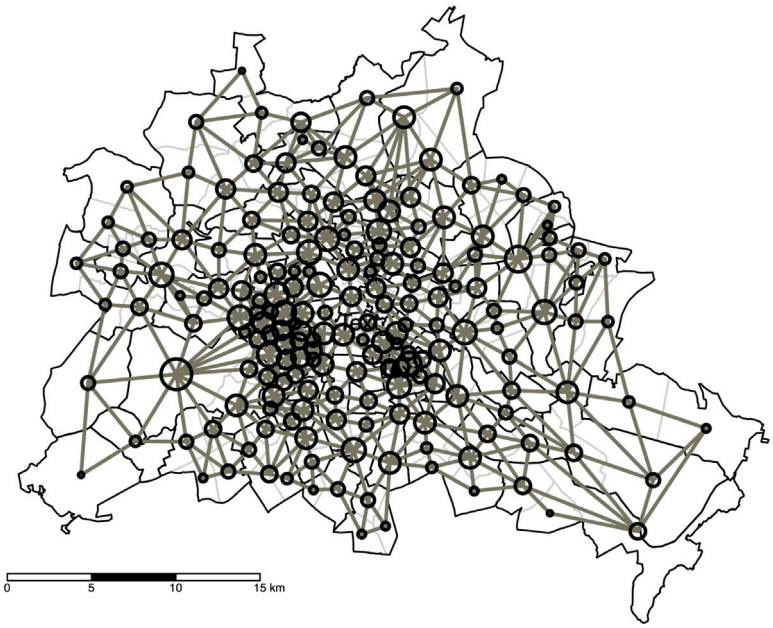


Fig. 4. Spatial locations including their spatial weights matrix **W** (green). The size of the centroids is proportional to the number of neighbouring locations.

also cover changes in the variance due to latent variables. Fourth, the unconditional variance varies across the property types, with developed land experiencing the highest variance, followed by condominium and undeveloped land. Note that undeveloped land usually does not have and will not get building permission.

Moreover, the models allow us to estimate the conditional log-volatility level for each location, time point, and response variable. These volatilities are given in the $n \times p$ matrix $\hat{\mathbf{H}}_t$. Fig. 6 shows the averages across space and time. Several conclusions can be drawn from these figures. Firstly, the conditional volatilities are the lowest for undeveloped land, followed by developed areas, and they are the highest for condominiums. Hence, the highest risks are observed for the condominium market. Secondly, we observe increased volatility at the end of the years for developed land and condominiums, mainly for the first half of the considered time horizon. The highest variation in the market risks is observed for developed areas — the most scarce type of real estate in an urban area. Thirdly, when looking at the maps, we see that risks are different across space. For instance, regarding condominiums, the highest average risks are observed for the former East-Berlin areas. For the other two types, a pronounced risk cluster can be seen in the North-Western city, Berlin-Reinickendorf, an area with a low population density and a large share of water and forest areas.

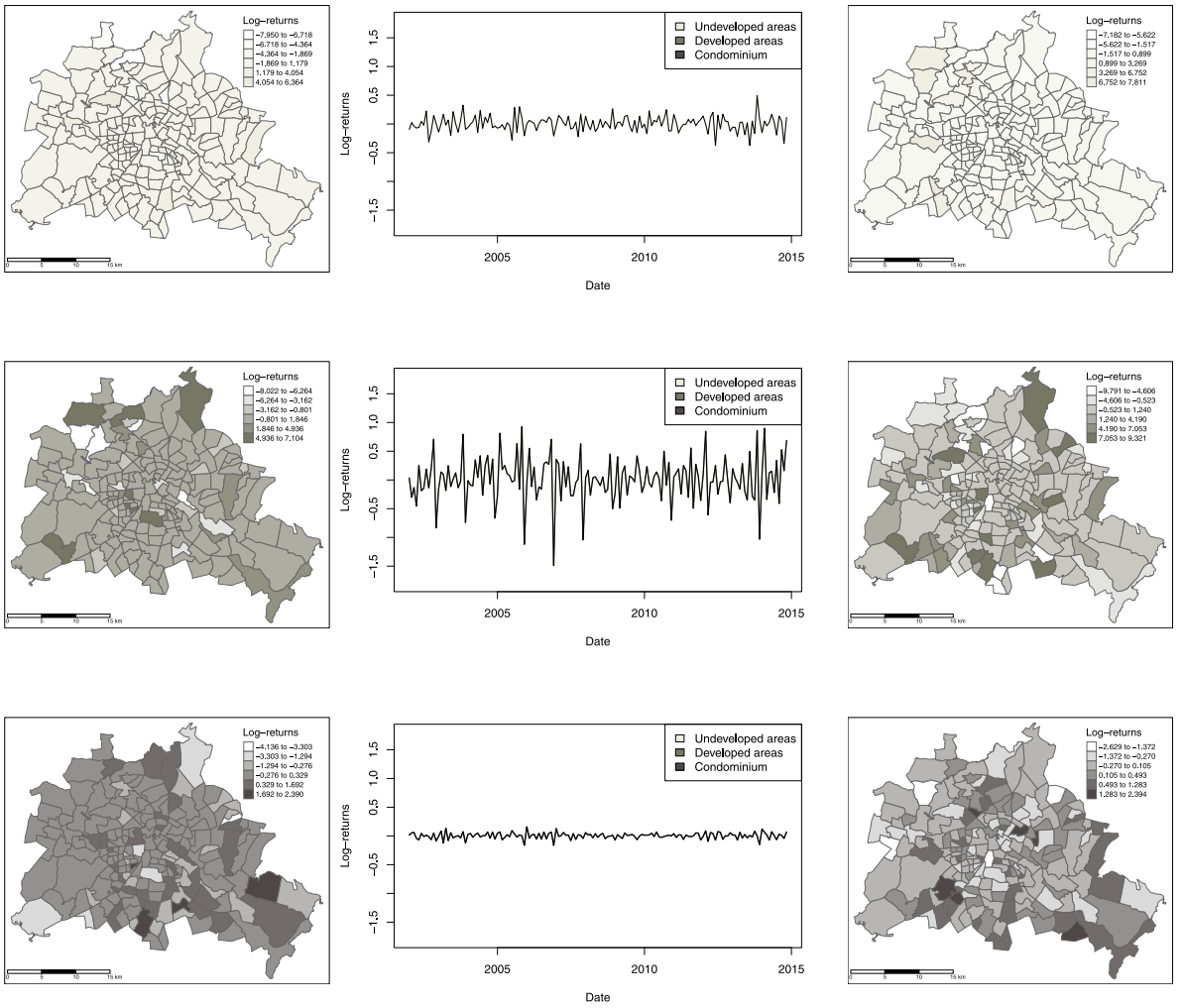


Fig. 5. Monthly log-returns of the price series displayed in Fig. 3 for all property types (from top to bottom: undeveloped land, developed land, condominiums). Left: Log returns across space at the first time point, centre: average log returns as time series (averaged across space), right: log returns across space at the last time point. Note that the log returns were computed from the raw spatiotemporal data and averaged afterwards to be depicted as a time series.

Table 3
QML estimates and standard errors of the empirical example. Spatial ARCH effects are highlighted in light green and temporal ARCH effects in dark green. Significant effects are marked by an asterisk (* t-value > 1.9, ** t-value > 2).

		Undeveloped land		Developed land		Condominium	
		Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
$\tilde{A}$		−4.686**	1.381	0.187	1.372	-2.652*	1.337
Ψ	Undeveloped land	0.111	0.074	0.016	0.075	−0.057	0.074
	Developed land	0.014	0.064	0.144**	0.062	0.000	0.064
	Condominium	−0.085	0.090	0.008	0.090	0.113	0.086
Π	Undeveloped land	0.583**	0.038	0.129**	0.038	−0.014	0.038
	Developed land	0.080**	0.031	0.553**	0.031	0.027	0.031
	Condominium	−0.028	0.044	0.078	0.044	0.606**	0.044

5. Summary and conclusion

In this paper, we have introduced a multivariate spatiotemporal autoregressive model for conditional heteroscedasticity (multivariate vec-spARCH). While ARCH and GARCH models are well-known in time-series econometrics and finance, there are only a few spatiotemporal extensions that typically do not account for spatial simultaneity. For any geographical phenomena, spatial interactions occur instantaneously due to the spatial proximity between the observations. Instead, previous papers typically only

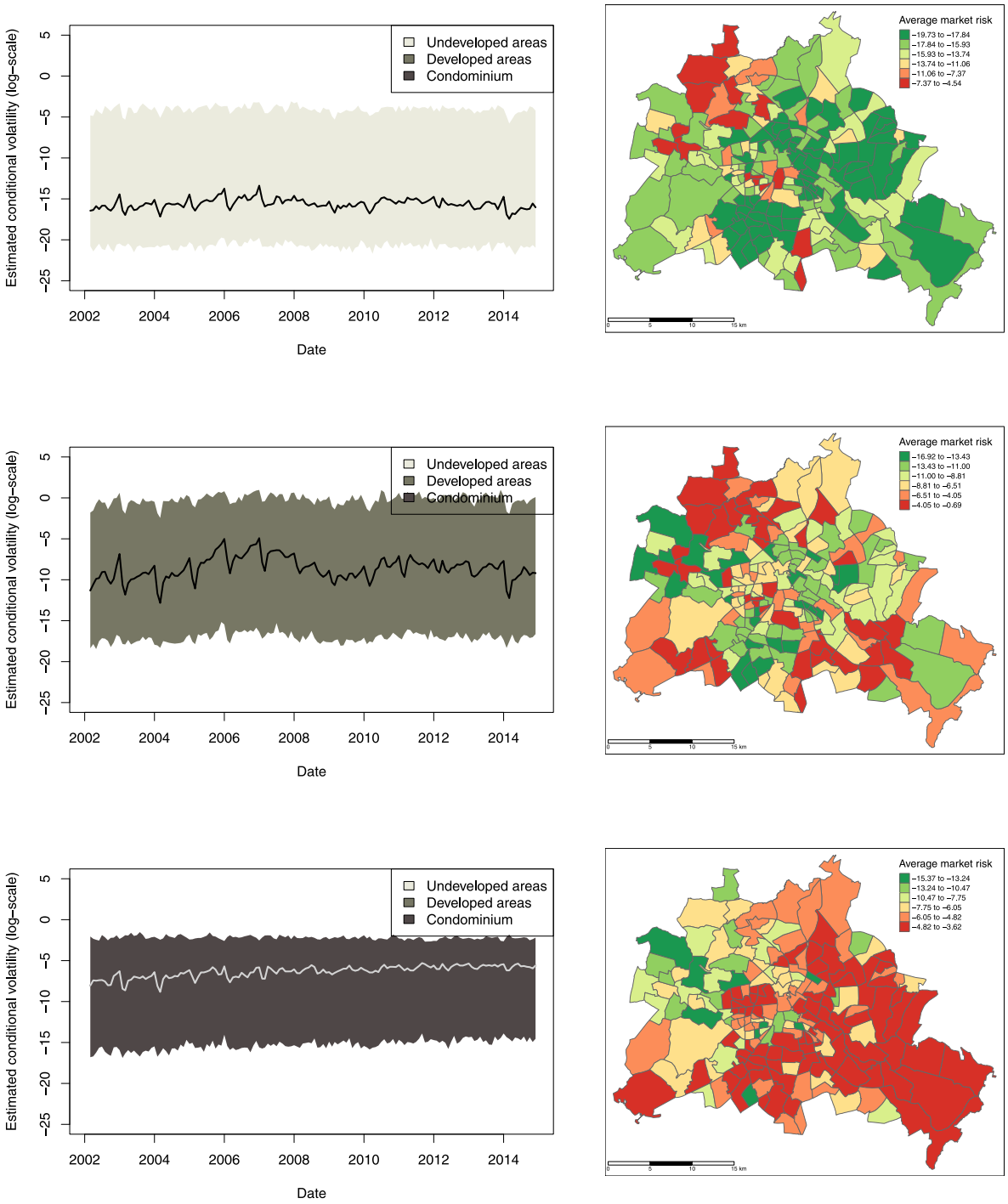


Fig. 6. Estimated conditional log-volatilities $\hat{\mathbf{H}}_t$. Left: Average conditional log-volatilities of all spatial locations plotted across time for all three response variables (from top to bottom: undeveloped land, developed land, condominiums). The shaded area corresponds to the area spanned by the 5% and 95% quantiles. Right: Average conditional log volatility of all time points plotted on a map. The colour scale represents the average market risks $\hat{\mathbf{H}}_t$ from low (green), average (yellow) to high (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

allowed for time-lagged spatial dependence. The model introduced in this paper explicitly accounts for instantaneous spatial and cross-variable interactions and temporal dependence in conditional variance. Thus, the model would also be suitable for purely spatial data without needing observations over time. In the empirical application, it is evident that the log returns of several real estate types are spatially autocorrelated. This indicates local clusters of increased volatilities and market risks – even though

the temporal dependence appears to be more important. Thus, we could show that there are temporally and spatially varying volatilities. Furthermore, we found significant cross-variable dependence in the first temporal lags but no significant instantaneous cross-variable interactions. Again, this motivates applying a multivariate spatiotemporal ARCH model in such studies instead of multivariate time-series models.

For this new model, we discussed parameter estimation using a QML approach. For this reason, the process is reformulated in a vec-representation, and a log-squared transformation is applied to obtain a multivariate spatiotemporal autoregressive process. We showed the consistency of the QML estimator under regular assumptions for the error process when the spatial and temporal dimensions increase. In the finite-sample case, we could see rapidly decreasing RMSEs in a series of simulations with different model specifications and error distributions. All our simulations could be performed in a reasonable amount of time using a standard computer. The required computational resources are usually the bottleneck of the QML approach due to the computation of the log-determinant of the Jacobian matrix.

There are many further directions for future research and potential fields of applications. First, we only considered logarithmic structures in the volatility models, but no classical ARCH structures. However, since the multivariate spatiotemporal model could be transformed into purely spatial models using the vec-representation, previous results of spatial ARCH and GARCH models could be applied. Furthermore, all these spatial econometric models rely on a (correctly) specified spatial weight matrix, which is mostly unknown in practice. Hence, estimation methods for the entire spatial dependence structures (i.e., each spatial weight) are desirable from a practical perspective. Penalised methods are promising in this case because many links can be considered to be zero. Considering long time series, it is also likely that the spatial dependence changes over time (see, e.g., Catania and Billé, 2017, for dynamic spatial panel models) and it would be interesting to consider time-varying spatial parameters or weight matrices. Despite their theoretical benefits of the non-negativity of $\mathbf{H}_t$, logarithmic ARCH approaches require strictly non-zero observations. Similar to the approach in Sucarrat and Escribano (2018), handling zero observations in univariate and multivariate spatiotemporal logarithmic ARCH/GARCH models should be a priority of future research.

Apart from applications in econometrics, environmental and climate processes would be interesting and potential fields for applying the spatiotemporal multivariate ARCH model. The process parameters can be interpreted as local risk measures, which are highly relevant in environmental studies. Furthermore, when considering our model as an error process, the statistical model can reflect spatially and temporally varying measurement or modelling uncertainties. For example, this might be of interest to GNSS positioning in urban environments. Epidemiological and medical studies are other fields where local risks and cross-variable interactions are highly relevant.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — 501539976.

Appendix. Proofs

Proof of Proposition 1. If all eigenvalues of $\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n)$ are smaller than one and $(\mathbf{I}_{np}-\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n))^j \to 0$ for an increasing power j, we get that

$$(\mathbf{I}_{np}+\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n)+\cdots+(\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n))^j)vec(\tilde{\mathbf{A}}) \to (\mathbf{I}_{np}-\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n))^{-1} \tag{8}$$

and

$$\ddot{\boldsymbol{Y}}_t = (\mathbf{I}_{np}-\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n))^{-1} + \sum_{i=0}^{\infty}(\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n))^i\mathbf{U}_{t-i}\,. \tag{9}$$

The stability follows from the convergence of the $(\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n))^i$. If the spectral radius of $\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n)$ is smaller than one, $\mathbf{S}_{np}^{-1}(\mathbf{\Pi}'\otimes\mathbf{I}_n)\to 0$ (e.g., Gentle 2017). □

Proof of Proposition 2. We have to show that

$$\frac{1}{Tnp}E\left(\ln\mathcal{L}(\tilde{\mathbf{A}},\mathbf{\Psi},\mathbf{\Pi}|\mathbf{Y}_0)\right)-\frac{1}{Tnp}E\left(\ln\mathcal{L}(\tilde{\mathbf{A}}_0,\mathbf{\Psi}_0,\mathbf{\Pi}_0|\mathbf{Y}_0)\right)\le 0,$$

where the equality holds if and only if $\tilde{\mathbf{A}}=\tilde{\mathbf{A}}_0$, $\mathbf{\Psi}=\mathbf{\Psi}_0$, and $\mathbf{\Pi}=\mathbf{\Pi}_0$.

$$\begin{aligned}
&\frac{1}{Tnp}E\left(\ln\mathcal{L}(\tilde{\mathbf{A}},\mathbf{\Psi},\mathbf{\Pi}|\mathbf{Y}_0)\right)-\frac{1}{Tnp}E\left(\ln\mathcal{L}(\tilde{\mathbf{A}}_0,\mathbf{\Psi}_0,\mathbf{\Pi}_0|\mathbf{Y}_0)\right)\\
=\;&\frac{1}{np}(\ln|\mathbf{S}_{np}|-\ln|\mathbf{S}_{np0}|)\\
-\;&\frac{1}{2np}\sum_{t=1}^{T}\left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(vec(\tilde{\mathbf{A}}_0-\tilde{\mathbf{A}})+((\mathbf{\Pi}_0'-\mathbf{\Pi}')\otimes\mathbf{I}_n)\ddot{\boldsymbol{Y}}_{t-1})\right]'\\
&\qquad\times\left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(vec(\tilde{\mathbf{A}}_0-\tilde{\mathbf{A}})+((\mathbf{\Pi}_0'-\mathbf{\Pi}')\otimes\mathbf{I}_n)\ddot{\boldsymbol{Y}}_{t-1})\right]\\
-\;&\frac{1}{2np}tr\left(\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{np}\ln|\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'|^{1/np} \\
&- \frac{1}{2np}\sum_{t=1}^{T}\left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(vec(\tilde{\mathbf{A}}_0-\tilde{\mathbf{A}})+((\mathbf{\Pi}_0'-\mathbf{\Pi}')\otimes\mathbf{I}_n)\ddot{\mathbf{Y}}_{t-1})\right]' \\
&\qquad\times\left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(vec(\tilde{\mathbf{A}}_0-\tilde{\mathbf{A}})+((\mathbf{\Pi}_0'-\mathbf{\Pi}')\otimes\mathbf{I}_n)\ddot{\mathbf{Y}}_{t-1})\right] \\
&- \frac{1}{2np}tr\left(\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'\right)
\end{aligned}
$$

First, we focus on the convergence of the quadratic term

$$\frac{1}{2np}\sum_{t=1}^{T}\mathbf{V}_t'\mathbf{V}_t \tag{10}$$

with

$$
\begin{aligned}
\mathbf{V}_t &= \mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(vec(\tilde{\mathbf{A}}_0-\tilde{\mathbf{A}})+((\mathbf{\Pi}_0'-\mathbf{\Pi}')\otimes\mathbf{I}_n)\ddot{\mathbf{Y}}_{t-1}) \\
&= \left(\mathbf{I}_{np}-((\mathbf{\Psi}'-\mathbf{\Psi}_0')\otimes\mathbf{W})\mathbf{S}_{np0}^{-1}\right)(vec(\tilde{\mathbf{A}}_0-\tilde{\mathbf{A}})+((\mathbf{\Pi}_0'-\mathbf{\Pi}')\otimes\mathbf{I}_n)\ddot{\mathbf{Y}}_{t-1}).
\end{aligned}
$$

Thus, under Assumption 5, (10) is equal to zero if and only if $\tilde{\mathbf{A}}=\tilde{\mathbf{A}}_0$ and $\mathbf{\Pi}=\mathbf{\Pi}_0$. Note that $\tilde{\mathbf{A}}_0$ is constant across time, while $\ddot{\mathbf{Y}}_t$ is varying due to the random variation in $\mathbf{\Xi}_t$. Thus, if $T=1$, $\tilde{\mathbf{A}}_0$ must be assumed constant across space to obtain identifiability.

Second,

$$\frac{1}{2np}tr\left(\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'\right)$$

is only a function of $\mathbf{\Psi}$ and

$$\frac{1}{np}tr\left(\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'\right)\geq|\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'|^{1/np}. \tag{11}$$

by the arithmetic and geometric means inequality of eigenvalues of $\mathbf{S}_{np}$. Further,

$$\mathbf{S}_{np}(\mathbf{\Psi})\mathbf{S}_{np0}^{-1}=\mathbf{I}_{np}-((\mathbf{\Psi}'-\mathbf{\Psi}_0')\otimes\mathbf{W})\mathbf{S}_{np0}^{-1}$$

is equal to $\mathbf{I}_{np}$ if and only if $\mathbf{\Psi}=\mathbf{\Psi}_0$ as $n\to\infty$. Then, the equality of (11) holds.

As a consequence, is equal to zero if and only if the parameters coincide with their true values $\mathbf{\Psi}_0$, $\mathbf{\Pi}_0$, and $\tilde{\mathbf{A}}_0$. Hence, the parameters are uniquely identifiable. □

Lemma 1 (*Yang and Lee (2017), Lemma 1*)**.** *The sequences $\mathbf{S}_{np}$ and $\mathbf{S}_{np}^{-1}$ are uniformly bounded in column sum norm, uniformly in $\mathbf{\Psi}$, if $\sup_{\mathbf{\Psi},n}\|\mathbf{\Psi}'\otimes\mathbf{W}_n\|_1<1$. They are uniformly bounded in row sum norm, uniformly in $\mathbf{\Psi}$, if $\sup_{\mathbf{\Psi},n}\|\mathbf{\Psi}'\otimes\mathbf{W}_n\|_\infty<1$.*

Lemma 2 (*Yu et al. (2008), Lemma 9*)**.** *Under Assumptions 1, 4 and 5, it holds for an np-dimensional non-stochastic, uniformly bounded matrix $\mathbf{B}_{np}$ that*

$$\frac{1}{npT}\sum_{t=1}^{T}\ddot{\mathbf{Y}}_t'\mathbf{B}_{np}\ddot{\mathbf{Y}}_t-\frac{1}{npT}E\left[\sum_{t=1}^{T}\ddot{\mathbf{Y}}_t'\mathbf{B}_{np}\ddot{\mathbf{Y}}_t\right]=O_p\left(\frac{1}{\sqrt{npT}}\right), \tag{12}$$

$$\frac{1}{npT}\sum_{t=1}^{T}\ddot{\mathbf{Y}}_t'\mathbf{B}_{np}vec(\mathbf{U}_t)-\frac{1}{npT}E\left[\sum_{t=1}^{T}\ddot{\mathbf{Y}}_t'\mathbf{B}_{np}vec(\mathbf{U}_t)\right]=O_p\left(\frac{1}{\sqrt{npT}}\right), \tag{13}$$

$$\frac{1}{npT}\sum_{t=1}^{T}vec(\mathbf{U}_t)'\mathbf{B}_{np}vec(\mathbf{U}_t)-\frac{1}{npT}E\left[\sum_{t=1}^{T}vec(\mathbf{U}_t)'\mathbf{B}_{np}vec(\mathbf{U}_t)\right]=O_p\left(\frac{1}{\sqrt{npT}}\right), \tag{14}$$

where $E\left[\sum_{t=1}^{T}\ddot{\mathbf{Y}}_t'\mathbf{B}_{np}\ddot{\mathbf{Y}}_t\right]$ is $O(1)$, $E\left[\sum_{t=1}^{T}\ddot{\mathbf{Y}}_t'\mathbf{B}_{np}vec(\mathbf{U}_t)\right]$ is $O(1/T)$ and $E\left[\sum_{t=1}^{T}vec(\mathbf{U}_t)'\mathbf{B}_{np}vec(\mathbf{U}_t)\right]$ is $O(1)$.

Proof of Theorem 1. The proof of the theorem consists of two parts; first, the identification of the parameters, and, second, the uniform and equicontinuous convergence of $\frac{1}{npT}\ln\mathcal{L}_{nT}(\boldsymbol{\vartheta}|\boldsymbol{Y}_0)$ to $\frac{1}{nT}Q(\boldsymbol{\vartheta}|\boldsymbol{Y}_0)$ in probability with $\boldsymbol{\vartheta}_0$ being a unique maximiser of $Q(\boldsymbol{\vartheta})$. Then, the consistency of the QML estimator follows.

1. The unique identification of the parameters is shown in the proof of Proposition 2.
2. Let

$$\tilde{\boldsymbol{U}}_t(\boldsymbol{\vartheta})=\mathbf{S}_{np}\ln vec(\mathbf{Y}_t^{(2)})-vec(\tilde{\mathbf{A}})-(\mathbf{\Pi}'\otimes\mathbf{I}_n)\ln vec(\mathbf{Y}_{t-1}^{(2)})$$

and

$$\boldsymbol{U}_t=\mathbf{S}_{np0}\ln vec(\mathbf{Y}_t^{(2)})-vec(\tilde{\mathbf{A}}_0)-(\mathbf{\Pi}_0'\otimes\mathbf{I}_n)\ln vec(\mathbf{Y}_{t-1}^{(2)}),$$

the true error vector of the transformed process, i.e., $vec(\Xi_t^{(\ln,2)})$. Furthermore, let

$$\tilde{\boldsymbol{U}}_t(\xi) = \boldsymbol{U}_t - (\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')\otimes \mathbf{W}\ddot{\boldsymbol{Y}}_t - vec(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_0) - ((\boldsymbol{\Pi}' - \boldsymbol{\Pi}_0')\otimes \mathbf{I}_n)\ddot{\boldsymbol{Y}}_{t-1}$$

with ξ being the differences in the parameters. Then,

$$\begin{aligned}
\tilde{\boldsymbol{U}}_t(\xi)'\tilde{\boldsymbol{U}}_t(\xi) \;=\; & \boldsymbol{U}_t'\boldsymbol{U}_t - vec(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_0)'vec(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_0) \\
& + \quad (\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')'\otimes \mathbf{W}'\ddot{\boldsymbol{Y}}_t'\ddot{\boldsymbol{Y}}_t\mathbf{W}\otimes(\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0') \\
& - \quad ((\boldsymbol{\Pi}' - \boldsymbol{\Pi}_0')\otimes \mathbf{I}_n)'\ddot{\boldsymbol{Y}}_{t-1}'\ddot{\boldsymbol{Y}}_{t-1}((\boldsymbol{\Pi}' - \boldsymbol{\Pi}_0')\otimes \mathbf{I}_n) \\
& + \quad 2(\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')'\otimes \mathbf{W}'\ddot{\boldsymbol{Y}}_t'\ddot{\boldsymbol{Y}}_{t-1}((\boldsymbol{\Pi}' - \boldsymbol{\Pi}_0')\otimes \mathbf{I}_n) \\
& - \quad 2(\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')'\otimes \mathbf{W}'\ddot{\boldsymbol{Y}}_t'\boldsymbol{U}_t \\
& - \quad 2((\boldsymbol{\Pi}' - \boldsymbol{\Pi}_0')\otimes \mathbf{I}_n)'\ddot{\boldsymbol{Y}}_{t-1}'\boldsymbol{U}_t \\
& + \quad 2(\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')'\otimes \mathbf{W}'\ddot{\boldsymbol{Y}}_t'vec(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_0) \\
& + \quad 2((\boldsymbol{\Pi}' - \boldsymbol{\Pi}_0')\otimes \mathbf{I}_n)'\ddot{\boldsymbol{Y}}_{t-1}'vec(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_0) \\
& - \quad 2vec(\tilde{\mathbf{A}} - \tilde{\mathbf{A}}_0)'\boldsymbol{U}_t\,.
\end{aligned}$$

Using Lemmas 1 and 2, it follows that

* $\frac{1}{npT}\sum_{t=1}^{T}\boldsymbol{U}_t'\boldsymbol{U}_t - \frac{1}{npT}E\left[\sum_{t=1}^{T}\boldsymbol{U}_t'\boldsymbol{U}_t\right] \xrightarrow{p} 0$,
* $\frac{1}{npT}\sum_{t=1}^{T}(\mathbf{W}\otimes\mathbf{I}_p)'\ddot{\boldsymbol{Y}}_t'\ddot{\boldsymbol{Y}}_t(\mathbf{W}\otimes\mathbf{I}_p) - \frac{1}{npT}E\left[\sum_{t=1}^{T}(\mathbf{W}\otimes\mathbf{I}_p)'\ddot{\boldsymbol{Y}}_t'\ddot{\boldsymbol{Y}}_t(\mathbf{W}\otimes\mathbf{I}_p)\right] \xrightarrow{p} 0$,
* $\frac{1}{npT}\sum_{t=1}^{T}\ddot{\boldsymbol{Y}}_{t-1}'\ddot{\boldsymbol{Y}}_{t-1} - \frac{1}{npT}E\left[\sum_{t=1}^{T}\ddot{\boldsymbol{Y}}_{t-1}'\ddot{\boldsymbol{Y}}_{t-1}\right] \xrightarrow{p} 0$,
* $\frac{1}{npT}\sum_{t=1}^{T}(\mathbf{W}\otimes\mathbf{I}_p)'\ddot{\boldsymbol{Y}}_t'\ddot{\boldsymbol{Y}}_{t-1} - \frac{1}{npT}E\left[\sum_{t=1}^{T}(\mathbf{W}\otimes\mathbf{I}_p)'\ddot{\boldsymbol{Y}}_t'\ddot{\boldsymbol{Y}}_{t-1}\right] \xrightarrow{p} 0$,
* $\frac{1}{npT}\sum_{t=1}^{T}(\mathbf{W}\otimes\mathbf{I}_p)'\ddot{\boldsymbol{Y}}_t'\boldsymbol{U}_t - \frac{1}{npT}E\left[\sum_{t=1}^{T}(\mathbf{W}\otimes\mathbf{I}_p)'\ddot{\boldsymbol{Y}}_t'\boldsymbol{U}_t\right] \xrightarrow{p} 0$, and
* $\frac{1}{npT}\sum_{t=1}^{T}\ddot{\boldsymbol{Y}}_{t-1}'\boldsymbol{U}_t - \frac{1}{npT}E\left[\sum_{t=1}^{T}\ddot{\boldsymbol{Y}}_{t-1}'\boldsymbol{U}_t\right] \xrightarrow{p} 0$.

Because $\tilde{\mathbf{A}}-\tilde{\mathbf{A}}_0$ is uniformly bounded, the remaining terms converge to zero in probability by Chebycheff's inequality. Moreover, as $\boldsymbol{\vartheta} = (vec(\tilde{\mathbf{A}})', vec(\boldsymbol{\Psi})', vec(\boldsymbol{\Pi})')'$ is bounded in $\boldsymbol{\Theta}$, we get that

$$\frac{1}{npT}\sum_{t=1}^{T}\tilde{\boldsymbol{U}}_t(\xi)'\tilde{\boldsymbol{U}}_t(\xi) - \frac{1}{npT}E\left[\sum_{t=1}^{T}\tilde{\boldsymbol{U}}_t(\xi)'\tilde{\boldsymbol{U}}_t(\xi)\right] \xrightarrow{p} 0$$

uniformly in $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}$, and, thus,

$$\frac{1}{npT}\ln\mathcal{L}_{nT}(\tilde{\mathbf{A}}, \boldsymbol{\Psi}, \boldsymbol{\Pi}|\mathbf{Y}_0) - \frac{1}{npT}Q(\tilde{\mathbf{A}}, \boldsymbol{\Psi}, \boldsymbol{\Pi}|\mathbf{Y}_0) \xrightarrow{p} 0$$

uniformly in $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}$.

Further, the equicontinuity of the expected likelihood must be shown. Let ι_{ij} be a matrix with the (i,j)-th entry equal to one, and all other entries are zero. First, $\frac{1}{np}\frac{\partial\ln|\mathbf{S}_{np}|}{\partial\psi_{ij}} = \frac{1}{np}tr(\mathbf{S}_{np}'^{-1}(\iota_{ij}\otimes\mathbf{W}))$ is uniformly bounded by a constant, uniformly in $\boldsymbol{\Psi}$, because $\mathbf{S}_{np}^{-1}$ is uniformly bounded according to Lemma 1. Secondly, $\frac{1}{np}\ln|\mathbf{S}_{np}|$ is a Lipschitz function in $\boldsymbol{\Psi}$ and, thus, uniformly equicontinuous. Thirdly,

$$\begin{aligned}
\sum_{t=1}^{T}\tilde{\boldsymbol{U}}_t(\xi)'\tilde{\boldsymbol{U}}_t(\xi) \;=\; & \left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}vec(\tilde{\mathbf{A}}_0 - \tilde{\mathbf{A}}) + \mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(\boldsymbol{\Pi}_0' - \boldsymbol{\Pi}')\ddot{\boldsymbol{Y}}_{t-1}\right]' \\
& \times\left[\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}vec(\tilde{\mathbf{A}}_0 - \tilde{\mathbf{A}}) + \mathbf{S}_{np}\mathbf{S}_{np0}^{-1}(\boldsymbol{\Pi}_0' - \boldsymbol{\Pi}')\ddot{\boldsymbol{Y}}_{t-1}\right]
\end{aligned}$$

is uniformly equicontinuous, because $\tilde{\mathbf{A}}$ and $\boldsymbol{\Pi}$ are bounded, $\mathbf{S}_{np}(\boldsymbol{\Psi})$ is uniformly bounded in $\boldsymbol{\Psi}$ and $\ddot{\boldsymbol{Y}}_t'\ddot{\boldsymbol{Y}}_t$ is $O(1)$ in $\boldsymbol{\vartheta}$ according to Lemma 2. Then, since $\mathbf{S}_{np}^{-1}$ is $O(1)$ in $\boldsymbol{\Psi}$ and $\mathbf{S}_{np}(\boldsymbol{\Psi})\mathbf{S}_{np0}^{-1} = \mathbf{I}_{np} - ((\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')\otimes\mathbf{W})\mathbf{S}_{np0}^{-1}$, also

$$\frac{1}{2np}tr\left(\mathbf{S}_{np}\mathbf{S}_{np0}^{-1}\mathbf{S}_{np0}'^{-1}\mathbf{S}_{np}'\right) = \frac{1}{2np}tr\left((\mathbf{I}_{np} - ((\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')\otimes\mathbf{W})\mathbf{S}_{np0}^{-1})(\mathbf{I}_{np} - ((\boldsymbol{\Psi}' - \boldsymbol{\Psi}_0')\otimes\mathbf{W})\mathbf{S}_{np0}^{-1})'\right)$$

is a Lipschitz function in $\boldsymbol{\Psi}$. Thus, this term is uniformly equicontinuous.

Because all terms are uniformly equicontinuous, also $\frac{1}{npT}Q(\tilde{\mathbf{A}}, \boldsymbol{\Psi}, \boldsymbol{\Pi}|\boldsymbol{Y}_0)$ is uniformly equicontinuous.

Because $\boldsymbol{\vartheta}_0$ is uniquely identified and the log-likelihood uniformly converges to the uniformly equicontinuous $\frac{1}{npT}Q(\tilde{\mathbf{A}}, \boldsymbol{\Psi}, \boldsymbol{\Pi}|\mathbf{Y}_0)$ in $\boldsymbol{\vartheta} = (vec(\tilde{\mathbf{A}})', vec(\boldsymbol{\Psi})', vec(\boldsymbol{\Pi})')'$, the consistency follows. This completes the proof. □

References

Bauwens, L., Sucarrat, G., 2010. General-to-specific modelling of exchange rate volatility: A forecast evaluation. Int. J. Forecast. 26 (4), 885–907.
Berrisch, J., Ziel, F., 2024. Multivariate probabilistic CRPS learning with an application to day-ahead electricity prices. Int. J. Forecast..

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31 (3), 307–327.
Borovkova, S., Lopuhaa, R., 2012. Spatial GARCH: A spatial approach to multivariate volatility modeling. Available at SSRN 2176781.
Brockwell, P.J., Davis, R.A., 2006. Introduction to Time Series and Forecasting. Springer Science & Business Media.
Caporin, M., Paruolo, P., 2015. Proximity-structured multivariate volatility models. Econometric Rev. 34 (5), 559–593.
Catania, L., Billé, A.G., 2017. Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. J. Appl. Econometrics 32 (6), 1178–1196.
Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. Wiley.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.
Engle, R., Kelly, B., 2012. Dynamic equicorrelation. J. Bus. Econom. Statist. 30 (2), 212–228.
Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11 (1), 122–150.
Espa, G., Arbia, G., Benedetti, R., et al., 1996. Effects of the MAUP on image classification. Geograp. Syst. 3 (2–3), 123–141.
Francq, C., Wintenberger, O., Zakoian, J.-M., 2013. Garch models without positivity constraints: Exponential or log GARCH? J. Econometrics 177 (1), 34–46.
Francq, C., Zakoian, J.-M., 2011. GARCH Models: Structure, Statistical Inference and Financial Applications. John Wiley & Sons.
Gentle, J.E., 2017. Matrix transformations and factorizations. In: Matrix Algebra. Springer, pp. 227–263.
Hølleland, S., Karlsen, H.A., 2020. A stationary spatio-temporal GARCH model. J. Time Series Anal. 41 (2), 177–209.
Horn, R.A., Johnson, C.R., 2012. Matrix Analysis. Cambridge University Press.
Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autorgegressive disturbance. J. Real Estate Finance Econ. 17 (1), 99–121.
Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72 (6), 1899–1925.
Otto, P., 2019. spGARCH: An R-package for spatial and spatiotemporal ARCH models. R J. 11 (2), 401–420.
Otto, P., Schmid, W., 2019. Spatial and spatiotemporal GARCH models – a unified approach. arXiv:1908.08320.
Otto, P., Schmid, W., Garthoff, R., 2018. Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity. Spatial Stat. 26, 125–145.
Otto, P., Schmid, W., Garthoff, R., 2019. Stochastic properties of spatial and spatiotemporal ARCH models. Statistical Papers.
Robinson, P.M., 2009. Large-sample inference on spatial dependence. Econom. J. 12.
Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 577–591.
Sato, T., Matsuda, Y., 2017. Spatial autoregressive conditional heteroskedasticity models. J. Japan Statist. Soc. 47 (2), 221–236.
Sato, T., Matsuda, Y., 2021. Spatial extension of generalized autoregressive conditional heteroskedasticity models. Spatial Econ. Anal. 16 (2), 148–160.
Silvennoinen, A., Teräsvirta, T., 2009. Multivariate garch models. In: Handbook of financial time series, Springer, pp. 201–229.
Sucarrat, G., Escribano, A., 2018. Estimation of log-GARCH models in the presence of zero returns. Eur. J. Finance 24 (10), 809–827.
Sucarrat, G., Grønneberg, S., Escribano, A., 2016. Estimation and inference in univariate and multivariate log-GARCH-X models when the conditional density is unknown. Comput. Stat. Data Anal. 100, 582–594.
Taspinar, S., Doğan, O., Chae, J., Bera, A.K., 2021. Bayesian Inference in Spatial Stochastic Volatility Models: An Application to House Price Returns in Chicago. Oxford Bulletin of Economics and Statistics.
Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46 (sup1), 234–240.
Yang, K., Lee, L.-F., 2017. Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. J. Econometrics 196 (1), 196–214.
Yu, J., de Jong, R., Lee, L.-F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. J. Econometrics 146 (1), 118–134.
Ziel, F., Weron, R., 2018. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. Energy Econ. 70, 396–420.