A general framework for spatial GARCH models

Philipp Otto[1] · **Wolfgang Schmid[2]**

Abstract

In time-series analysis, particularly in finance, generalized autoregressive conditional heteroscedasticity (GARCH) models are widely applied statistical tools for modelling volatility clusters (i.e., periods of increased or decreased risk). In contrast, it has not been considered to be of critical importance until now to model spatial dependence in the conditional second moments. Only a few models have been proposed for modelling local clusters of increased risks. In this paper, we introduce a novel spatial GARCH process in a unified spatial and spatiotemporal GARCH framework, which also covers all previously proposed spatial ARCH models, exponential spatial GARCH, and time-series GARCH models. In contrast to previous spatiotemporal and time series models, this spatial GARCH allows for instantaneous spill-overs across all spatial units. For this common modelling framework, estimators are derived based on a non-linear least-squares approach. Eventually, the use of the model is demonstrated by a Monte Carlo simulation study and by an empirical example that focuses on real estate prices from 1995 to 2014 across the postal code areas of Berlin. A spatial autoregressive model is applied to the data to illustrate how locally varying model uncertainties (e.g., due to latent regressors) can be captured by the spatial GARCH-type models.

✉ Philipp Otto
  otto@ikg.uni-hannover.de

✉ Wolfgang Schmid
  schmid@europa-uni.de

1   Institute of Cartography and Geoinformatics, Leibniz University Hannover, Appelstraße 9a, 30167 Hannover, Lower Saxony, Germany

2   Department of Statistics, European University Viadrina, Große Scharrnstraße 59, 15230 Frankfurt, Brandenburg, Germany

⌖ Springer

1 Introduction

Recent literature have dealt with the extension of generalized autoregressive conditional heteroscedasticity (GARCH) models to spatial and spatiotemporal processes (e.g., Otto et al. 2018, 2021; Sato and Matsuda 2020). Whereas the classical ARCH model is defined as a process over time, these random processes have a multidimensional support. Thus, they allow for spatially dependent second-order moments, while the observations are uncorrelated and the mean is constant in space (see Otto et al. 2021). For all geo-referenced processes, it is important to allow for instantaneous spatial interactions, as "near things are more related than distant things" (Tobler's first law of geography, Tobler 1970). That is, observations nearby are more similar than observations with larger distances. With regard to autoregressive dependencies, the focus was mostly on the mean process, but not on the spatial conditional heteroscedasticity. Thus, Otto et al. (2018) suggested first a purely spatial ARCH model. Furthermore, Sato and Matsuda (2017, 2020) introduced a random process incorporating elements of GARCH and exponential GARCH (E-GARCH) processes, which is, however, neither a GARCH nor an E-GARCH process—also in a one-dimensional space, where the model should collapse to a classical time-series GARCH model. Their model can be rather considered as symmetric spatial log-GARCH process. Moreover, Otto et al. (2018) only focussed on spatial ARCH processes without considering the influences from the realized, conditional variance at neighbouring locations. Direct extensions of GARCH and E-GARCH processes to spatial settings do not exist among current research.

Thus, we introduce a completely novel generalized spatial ARCH model (spatial GARCH or spGARCH) in this paper. Because a general definition of this model is used, time-series GARCH models (Bollerslev 1986), the previously introduced spatial ARCH (Otto et al. 2018) and the symmetric log-GARCH (Sato and Matsuda 2020) are nested. This definition also allows us to define a spatial exponential spatial GARCH model, which will be the subject of a future paper. Moreover, other GARCH-type models, like threshold or multivariate GARCH models, can easily be constructed. This unified spatial GARCH process is a completely new class of models in spatial statistics/econometrics, for which we derive consistent estimators based on a nonlinear least-squares approach. In addition, all models are computationally implemented in one library, the R-package spGARCH (version > 2.0).

From a practical perspective, this unified spatial GARCH model can be used to model spill-over effects in the conditional variances across the spatial units. That means that an increasing variance in a certain region of the considered space would lead to an increase or decrease in the adjacent regions, depending on the direction (sign) of the spatial dependence. Compared to previous spatiotemporal GARCH models, these spill-overs are instantaneous. That is, there is no time lag needed. All previously proposed spatiotemporal GARCH models include the spatial autoregressive dependence in the first temporal lag (i.e., spatial spillover are temporally delayed), so that these models can be seen as special cases of multivariate time-series GARCH models (e.g., Hølleland and Karlsen 2020; Borovkova and Lopuhaa 2012; Caporin and Paruolo 2015, or Billio et al. 2021 for networks). Contrary to these spatiotemporal models, the proposed spatial GARCH model could also be applied to purely spatial data, like for

modelling local climate risks, such as fluctuations in the temperature and precipitation, or financial risks in spatially constrained markets, such as real estate or labour. Furthermore, spatial GARCH-type models can be used as error models for any linear or non-linear spatial regression model to account for local model uncertainties (i.e., areas in which the considered models perform worse than in others). These model uncertainties can be considered to be a kind of local risk. Instead of modelling such autoregressive dependence as a GARCH process, stochastic volatility models could also be used. Taşpınar et al. (2021) considered a stochastic volatility approach for spatial settings and showed its applicability for U.S. house prices.

The remainder of this paper is structured as follows. In the next section, we introduce the generalized framework of spatial and spatiotemporal autoregressive conditional heteroscedasticity models and discuss two examples nested within this approach, more precisely, the novel spatial GARCH (as an equivalent to the time-series GARCH models) and the spatial log-GARCH processes by Sato and Matsuda (2017, 2020). Following from there, a non-linear least-squares procedure is introduced for this model class. These theoretical sections are followed up with a discussion of the insights gained from simulation studies. The paper then supplies a real-world example, namely the real estate prices in the German capital city of Berlin. In Sect. 6, we stress some important extensions for future research before concluding the paper.

2 Spatial and spatiotemporal GARCH-type models

Let $\{Y(s) \in \mathbb{R} : s \in D_s\}$ be a univariate stochastic process, where D_s represents a set of possible locations in a q-dimensional space. Thus, spatial and spatiotemporal models are both covered by this approach. With regards to spatiotemporal processes, the temporal dimension can be easily considered as one of the q dimensions. In addition, time-series GARCH models are included for $q = 1$.

Let $s_1, \ldots, s_n$ denote all locations, and let Y stand for the vector of observations $(Y(s_i))_{i=1,\ldots,n}$. The commonly applied spatial autoregressive (SAR) model implies that the conditional variance $\text{Var}(Y(s_i)|Y(s_j), j \neq i)$ is constant (cf. Cressie 1993; Cressie and Wikle 2011) and does not depend on the observations of neighbouring locations. This approach is extended by assuming the changes in the volatility can spill over to neighbouring regions and that conditional variances can vary over space, resulting in clusters of high and low variance. As in time-series ARCH models developed by Engle (1982), the vector of observations is given by the non-linear relationship

$$Y = \text{diag}(h)^{1/2} \varepsilon \tag{1}$$

where $h = (h(s_1), \ldots, h(s_n))'$ and $\varepsilon = (\varepsilon(s_1), \ldots, \varepsilon(s_n))'$ is a noise component, which is later specified in more detail.

Moreover, we assume that a known function f exists, which relates h to a vector $F = (f(h(s_1)), \ldots, f(h(s_n)))'$. This general approach is beneficial because different spatial GARCH-type processes can be defined by choosing f and a suitable model of F. For instance, they could have additive or multiplicative dynamics, or the spill-over effects in the conditional variances could be global or locally constrained to direct

neighbouring observations. In this paper, we extend the class of spatial ARCH models to generalized spatial GARCH models (spGARCH), which are analogously defined to the time-series GARCH models and have additive dynamics (cf. Bollerslev 1986). Besides, previously introduced spatial ARCH models are nested within this general approach (e.g., Otto et al. 2018; Sato and Matsuda 2020; see also Examples 1 and 3 in Sect. 2.3).

2.1 Generalized approach

Below, we introduce a general approach covering some important spatial and spatiotemporal GARCH-type models, namely the spatial ARCH model of Otto et al. (2018) and the logarithmic model of Sato and Matsuda (2020). For these models, vector $\boldsymbol{F}$ is chosen as

$$\boldsymbol{F} = \boldsymbol{\alpha} + \mathbf{W}_1 \boldsymbol{\gamma}(\boldsymbol{Y}^{(2)}) + \mathbf{W}_2 \boldsymbol{F} \tag{2}$$

with a measurable function $\boldsymbol{\gamma}(\boldsymbol{x}) = (\gamma_1(\boldsymbol{x}), \ldots, \gamma_n(\boldsymbol{x}))'$ and $\boldsymbol{Y}^{(2)} = (Y(\boldsymbol{s}_1)^2, \ldots, Y(\boldsymbol{s}_n)^2)'$. The weighting matrices $\mathbf{W}_1 = (w_{1,ij})_{i,j=1,\ldots,n}$ and $\mathbf{W}_2 = (w_{2,ij})_{i,j=1,\ldots,n}$ are assumed to be non-negative with zeros on the diagonal (i.e., $w_{v,ij} \geq 0$ and $w_{v,ii} = 0$ for all $i, j = 1, \ldots, n$ and $v = 1, 2$). Moreover, let $\boldsymbol{\alpha} = (\alpha_i)_{i=1,\ldots,n}$ be a positive vector.

First, we discuss under what conditions the process is well-defined. To do this, we make use of the Banach fixed point theorem for random processes. The field of random fixed point theorems has been studied by several authors (e.g., Hanš 1957; Bharucha-Reid et al. 1976; Tan and Yuan 1997).

In the following, we make use of the notation $\mathbf{E} = \operatorname{diag}(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)$. Considering the operator

$$\boldsymbol{T}(\omega) \circ z = \boldsymbol{\alpha} + \mathbf{W}_1 \boldsymbol{\gamma}(\mathbf{E}(\omega)z) - (\mathbf{I} - \mathbf{W}_2)(f(z_i))_{i=1,\ldots,n} + z \tag{3}$$

defined on $I\!R^n$ with a norm $||.||$, the following conditions can be derived such that the process is well-defined. Furthermore, ω is an element in the probability space of the error process and $\circ$ denotes the Hadamard product.

Theorem 1 *Suppose that the operator $\boldsymbol{T}(\omega)$ defined in (3) is a continuous random operator on $(I\!R^n, ||.||)$ to itself and that there is a a non-negative real-valued random variable $L_n(\omega) < 1$ a.s. such that $||\boldsymbol{T}(\omega) \circ z_1 - \boldsymbol{T}(\omega) \circ z_2|| \leq L_n(\omega)||z_1 - z_2||$ for all $z_1, z_2 \in I\!R^n$. Then the equations (1) and (2) have exactly one real-valued measurable solution z.*

The proof of this theorem is given in the Appendix. Note that the condition of Theorem 1 is fulfilled if, for example, $\boldsymbol{\gamma}$ satisfies a Lipschitz condition with constant L_1, $(f(z_i))_{i=1,\ldots,n}$ satisfies a Lipschitz condition with a constant L_2 and $L_n := L_1||\mathbf{W}_1\mathbf{E}|| + L_2||\mathbf{I} - \mathbf{W}_2|| < 1$ where we make use of the matrix norm induced by the vector norm. However, in order to guarantee that L_n does not depend on n, we need stronger conditions. If we take the 1-norm and if the matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are row-standardized then $||\mathbf{I} - \mathbf{W}_2|| < 2$. To ensure that $||\mathbf{W}_1\mathbf{E}||$ is bounded we have to assume that the $\varepsilon(\boldsymbol{s}_i)$ are uniformly bounded. We refer to Otto et al. (2018)

where this problem is discussed for a spatial ARCH process in more detail. Moreover, it is important to note that the operator $T(\omega)$ is continuous if f and γ are continuous.

Further, the solution of (3) which reflects h should be non-negative such that the process Y is a well-defined real-valued process. In many applications, the functions f and γ are defined to be zero for negative values, such that h is always positive if $\alpha > 0$. We will come back to this point later.

In addition, the fixed-point theorem of Banach implies that the sequence $z_m = T(z_{m-1})$, $m \geq 1$ converges to h for given α, E, W_1, and W_2. Consequently, this result represents one way to simulate such a process.

2.2 Properties of spatial GARCH models

Below, we discuss some important properties of this process including the following condition for stationarity.

Corollary 2 *Suppose that the assumptions of Theorem 1 are fulfilled and that the solution of (3) is non-negative. If $(\varepsilon(s_1), \ldots, \varepsilon(s_n))'$ is strictly stationary, then $(Y(s_1), \ldots, Y(s_n))'$ is strictly stationary as well.*

Moreover, the observations $Y(s)$ are uncorrelated with a mean of zero, as we will show in the following theorem. Thus, spatial GARCH models are suitable error models for use with other linear or non-linear spatial regression models, such as spatial autoregressive or spatial error models (see also Elhorst 2010), without affecting the mean equation. In this way, locally varying model uncertainties can be captured.

Theorem 3 *Let $i \in \{1, \ldots, n\}$. Suppose that the assumptions of Theorem 1 are satisfied and that the solution of (3) is non-negative. Further let ε be sign-symmetric, i.e.,*

$$\varepsilon \overset{d}{=} ((-1)^{v_1} \varepsilon(s_1), \ldots, (-1)^{v_n} \varepsilon(s_n)) \quad \text{for all} \quad v_1, \ldots, v_n \in \{0, 1\}.$$

(a) Then $Y(s_i)$ is a symmetric random variable. All odd moments and all conditional odd moments of $Y(s_i)$ are zero, provided that they exist.
(b) It holds that $Cov(Y(s_i), Y(s_j)) = 0$ for $i \neq j$ if the second moment exists.

In the spatial setting, however, the conditional variance $\text{Var}(Y(s_i)|Y(s_j), j \neq i)$ is not exactly equal to $h(s_i)$ (see Otto et al. 2021). In principle, this is due to the fact that there is no clear (causal) ordering of observations as in time series, where only past observations can influence future observations, but not vice versa. In case of a directional spatial dependence, however, $h(s_i)$ is equal to the conditional variance at location s_i. A detailed analysis of this point can be found in Otto et al. (2021). Nevertheless, the interpretation of h is similar to the conditional variance. In locations s, where $h(s)$ is large, the conditional variance is also large and vice versa (see Otto et al. 2021, Fig. 1). That means that the local risk or level of uncertainty of this particular region is high compared to its neighbours. Such regions could be identified via h; this could be of interest in terms of the valuation of real estate or other immovable assets since it provides insights into an individual location's risk.

In addition to this, the spatial GARCH coefficients measure potential risk spillovers from neighbouring locations. It is worth noting that in the case of directional spatial processes, h is equal to the conditional variances. Thus, it can be interpreted in the same way as with time-series GARCH models.

2.3 Examples of spatial GARCH models

This general framework allows for a large range of GARCH-type models. Depending on the definition of f and $\boldsymbol{\gamma}$, the resulting spatial GARCH-type models have different stochastic properties. We discuss some important special cases below, starting with the spatial ARCH model, which is a direct extension of the ARCH process of Engle (1982) to spatial and spatiotemporal processes. It was originally introduced by Otto et al. (2018). For more details on its stochastic properties, we refer to (Otto et al. 2021).

Example 1 (Spatial ARCH process of Otto et al. 2018) Choosing $f(x) = x I_{[0,\infty)}(x)$, $\gamma_i(\boldsymbol{x}) = x_i I_{[0,\infty)}(x_i)$ for $i = 1, \ldots, n$, and $\mathbf{W}_2 = \mathbf{0}$ the spatial ARCH (spARCH) process is obtained. It is given by

$$Y(\boldsymbol{s}_i) = \sqrt{h(\boldsymbol{s}_i)}\varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n$$

with

$$\boldsymbol{h} = \boldsymbol{\alpha} + \mathbf{W}_1 \boldsymbol{Y}^{(2)} \,.$$

Here, $I_A(x)$ denotes the indicator function on a set A. The process is well-defined if $||\mathbf{W}_1 \mathbf{E}|| < 1$. This is an immediate consequence of Theorem 1. Indeed, the spatial ARCH process can be easily extended to a spatial GARCH process by considering the realized values of $h(\cdot)$ in adjacent locations. This novel spatial GARCH process is defined in the following example.

Example 2 (Spatial GARCH process) Taking $f(x) = x I_{[0,\infty)}(x)$ and $\gamma_i(\boldsymbol{x}) = x_i I_{[0,\infty)}(x_i)$ for $i = 1, \ldots, n$ a spatial GARCH (spGARCH) process is obtained. That is,

$$Y(\boldsymbol{s}_i) = \sqrt{h(\boldsymbol{s}_i)}\varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n$$

with

$$\boldsymbol{h} = \boldsymbol{\alpha} + \mathbf{W}_1 \boldsymbol{Y}^{(2)} + \mathbf{W}_2 \boldsymbol{h} \,.$$

Since $\boldsymbol{Y}^{(2)} = \mathbf{E}\boldsymbol{h}$, the quantity $\boldsymbol{h}$ can be specified as

$$\boldsymbol{h} = (\mathbf{I} - \mathbf{W}_1 \mathbf{E} - \mathbf{W}_2)^{-1}\boldsymbol{\alpha} \,, \tag{4}$$

if the inverse exists. For this simple example, there is a unique solution if $||\mathbf{W}_1 \mathbf{E} + \mathbf{W}_2|| < 1$, as it is already expressed in Theorem 1. Alternatively, the condition is fulfilled if the process is directional. In this case, $\mathbf{W}_1$ and $\mathbf{W}_2$ are lower or upper triangular matrices (cf. Basak et al. 2018; Merk and Otto 2021).

The case of triangular matrices also includes causal temporal and spatiotemporal GARCH processes. Let us first consider time-series GARCH models. In this case, the

dimension of the underlying domain D_s is equal to $q = 1$ and there is strict causal relation between the observations. That is, current observations are only influenced by past observations. This implies a triangular structure of the weight matrices. For instance, for a GARCH(1, 1) process, the first subdiagonal elements of $\mathbf{W}_1$ and $\mathbf{W}_2$ are equal to one if $s_1 = 1, \ldots, s_n = T$ representing the consecutive time points from one to T. Similarly, for spatiotemporal processes, the weight matrices are triangular if instantaneous spatial interactions are excluded, such as for the models of Hølleland and Karlsen (2020), Borovkova and Lopuhaa (2012), and Caporin and Paruolo (2006). In contrast to these causal processes, the current observations of this new spatiotemporal GARCH model are not only influenced by past realisations of the process, but also by their neighbouring observations. In this way, the model can account for instantaneous spatial interactions as implied by the first law of geography.

Contrary to the previous examples, Sato and Matsuda (2017, 2020) have considered a slightly different choice of f, and have used the log-transformation to avoid any non-negativity problems of $\boldsymbol{h}$. Thus, their model combines the GARCH and the E-GARCH attempts and can be regarded as the spatial extension of a (symmetric) log-GARCH process (see also Francq et al. 2013). Let $\boldsymbol{h}_L = (\log(h(\boldsymbol{s}_i)))_{i=1,\ldots,n}$ and $\boldsymbol{Y}_L^{(2)} = (\log(Y(\boldsymbol{s}_j)^2))_{i=1,\ldots,n}$.

Example 3 (Spatial log-GARCH process of Sato and Matsuda 2017) Choosing $f(x) = \log(x)$ and $\gamma_i(\boldsymbol{x}) = \log(x_i)$ the symmetric spatial log-GARCH (log-spGARCH) process is obtained, i.e.,

$$Y(\boldsymbol{s}_i) = \sqrt{h(\boldsymbol{s}_i)}\varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n$$

with

$$\boldsymbol{h}_L = \boldsymbol{\alpha} + \mathbf{W}_1 \boldsymbol{Y}_L^{(2)} + \mathbf{W}_2 \boldsymbol{h}_L.$$

The process has a unique solution if $||\mathbf{W}_1 + \mathbf{W}_2|| < 1$. This is also an immediate consequence of Theorem 1. It is obtained by setting $\gamma(\boldsymbol{x}) = (\log(x_i))_{i=1,\ldots,n}$, $f(x) = \log(x)$, and considering the right side of (3) to be a function of $\log(z)$. We see that the condition on the existence of a solution is much simpler than for the spGARCH process, since it only depends on the weight matrices and not on the random matrix $\mathbf{E}$. This simplification is due to the fact that we have an additive decomposition of the function γ, i.e., $\gamma(\mathbf{E}z) = \boldsymbol{\phi}_1(\mathbf{E}) + \boldsymbol{\phi}_2(z)$ with certain functions $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. This functional equation is solved by the logarithm function. However, the behaviour of the log-spGARCH is different to that of the spGARCH. Thus, the one or the other could be preferable for empirical applications. To summarize this section, we provide an overview on all nested spatial GARCH-type models and important other GARCH model, e.g., the classical time-series GARCH models, in Table 1.

3 Statistical inference

In the following section, we firstly discuss the choice of the weight matrices in more detail. In a general setting, $\mathbf{W}_1$ and $\mathbf{W}_2$ have $n(n - 1)$ free parameters, while only

Table 1 Overview of nested spatial GARCH-type models for $\gamma_i(\boldsymbol{Y}) = Y(s_i)^2$

Model	f	$\gamma_i(\boldsymbol{Y})$	$\mathbf{W}_1$	$\mathbf{W}_2$
Dimension $q > 1$ (non-triangular weight matrices)				
Spatial ARCH model, Ex. 1, Otto et al. (2018)	$f(x) = x$	$Y(s_i)^2$	Standard spatial weight matrix	Zero matrix
Spatial GARCH model, Ex. 2	$f(x) = x$	$Y(s_i)^2$	Standard spatial weight matrices	
Symmetric spatial log-GARCH model, Ex. 3, Sato and Matsuda (2020)	$f(x) = \log(x)$	$Y(s_i)^2$	Standard spatial weight matrices	
Dimension $q > 1$ (triangular weight matrices)				
Directional/causal spatial GARCH models, Merk and Otto (2021)	$f(x) = x$	$Y(s_i)^2$	Strictly triangular matrices	
Spatiotemporal GARCH models with lagged spatial effects (see also Hølleland and Karlsen 2020)	$f(x) = x$	$Y(s_i)^2$	Strictly triangular matrices (vec-representation nT-dimensional)	
Dimension $q = 1$ (triangular weight matrices)				
Time-series GARCH, Bollerslev (1986)	$f(x) = x$	$Y(s_i)^2$	Strictly triangular matrices	
Time-series ARCH, Engle (1982)	$f(x) = x$	$Y(s_i)^2$	Strictly triangular matrices	Zero matrix

n values are observed. In spatial econometrics, these matrices are therefore usually replaced with a parametric model to control the influence of adjacent regions. Note that the parameter estimates are typically biased if the weights are misspecified. Alternatively, they might instead be estimated using statistical learning approaches, e.g., lasso-type estimators under the assumption of a certain degree of sparsity (e.g., Bhattacharjee and Jensen-Butler 2013; Otto and Steinert 2021). In this section of the paper, however, we will focus on a classical parametric model. For this, we develop an estimation method based on non-linear least squares estimators and show the consistency of these estimators. In practice, the weight matrices must carefully be selected.

3.1 Choice of weight matrices

There is great flexibility in the choice of the weight matrices (see Getis 2009 for an overview). In practice, these are usually dependent upon additional parameters and spatial locations. Frequently, it is assumed that $\mathbf{W}_1 = \rho\mathbf{W}_1^*$ and $\mathbf{W}_2 = \lambda\mathbf{W}_2^*$ with the predefined, known matrices $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$. That is, $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ describe the structure of the spatial dependence, with the weights as a multiple of these specific matrices. In settings such as these, it is easy to test whether a random process exhibits such a spatial dependence, by testing the parameters ρ and λ. As with time-series GARCH models, ρ measures the extent to which a volatility shock in one region spills over to neighbouring regions, while $\rho + \lambda$ gives an impression how fast this effect will fade out

in space (see, e.g., Campbell et al. 1997). It is important to note that spill-over effects of shocks always happen simultaneously in purely spatial setting, i.e., without any temporal delay. For spatial autoregressive models, a similar distinction between local and global spill-over effects can be made (see Fingleton 2009, 2008). A more general approach can be obtained by choosing $\mathbf{W}_k = \text{diag}(\rho_1, \ldots, \rho_1, \ldots, \rho_k, \ldots, \rho_k)\mathbf{W}^*$ as the weights for $k \in \{1, 2\}$. Here, different areas are weighted in different ways. For instance, all counties of state i are weighted by ρ_i, while counties of another state, j, get a different weighting factor, ρ_j. Alternatively, $\mathbf{W}_k^*$ could be chosen as $(K_\theta(\mathbf{s}_i - \mathbf{s}_j))_{i,j=1,\ldots,n}$ for $k \in \{1, 2\}$ with a known function K. In this case, the spatial correlation depends on the distance between two locations. For instance, inverse distance weighting schemes $K(\mathbf{x}) = ||\mathbf{x}||^{-k}$ with k being estimated, or anisotropic weighting schemes dependent upon the bearing between two locations.

3.2 Parameter estimation

Below, we assume that the weight matrices have the structure

$$\mathbf{W}_1 = \rho\mathbf{W}_1^*, \mathbf{W}_2 = \lambda\mathbf{W}_2^*, \boldsymbol{\alpha} = \alpha\mathbf{1}. \tag{5}$$

Thus, the model has three parameters to be estimated, $\boldsymbol{\vartheta} = (\rho, \lambda, \alpha)'$. Let $\boldsymbol{\vartheta}_0$ denote the true parameters. In the following we use the symbol $||\mathbf{x}||_2$ for the Euclidean norm of a vector $\mathbf{x}$ and $||\mathbf{A}||_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$ for the matrix norm of an $n \times n$ matrix $\mathbf{A}$, which is induced by a maximum norm.

Even though this seems to be a strong restriction compared to the general definition of the model in (1) and (2), it is probably the most widely applied specification in practice. Generally, the estimation method could also be applied for more complex specifications of the spatial interactions, such as the above-mentioned choices or higher-order spatial lags. In this case, particular attention should be paid to the identifiability of the process parameters (cf. Manski 1993) and to the consistence of the parameter estimators.

One possible method for estimating the parameters is the non-linear least-squares approach (NLSE). Squaring the components of (1) and taking the logarithms, we get that for $i = 1, \ldots, n$

$$\log(Y(\mathbf{s}_i)^2) = \log(h(\mathbf{s}_i)) + \log(\varepsilon(\mathbf{s}_i)^2)$$
$$= \mathbb{E}(\log(\varepsilon(\mathbf{s}_i)^2)) + \log(h(\mathbf{s}_i)) + \eta(\mathbf{s}_i)$$

with $\eta(\mathbf{s}_i) = \log(\varepsilon(\mathbf{s}_i)^2) - \mathbb{E}(\log(\varepsilon(\mathbf{s}_i)^2))$. Now $\eta(\mathbf{s}_i), i = 1, \ldots, n$ is a white noise process. Moreover, it follows with $\tau(x) = f(\exp(x))$ that

$$\mathbf{F} = (\tau(\log(h(\mathbf{s}_i)))_{i=1,\ldots,n} = (\mathbf{I} - \lambda\mathbf{W}_2^*)^{-1}(\alpha\mathbf{1} + \rho\mathbf{W}_1^*\boldsymbol{\gamma}(Y^{(2)}))$$
$$= (\mathbf{I} - \lambda\mathbf{W}_2^*)^{-1}(\alpha\mathbf{1} + \rho\mathbf{W}_1^*\tilde{\boldsymbol{\gamma}}(\log(Y^{(2)})))$$

where $\tilde{\boldsymbol{\gamma}}(\mathbf{x}) = (\gamma_i(\exp(x_1), \ldots, \exp(x_n)))_{i=1,\ldots,n}$. Note that $(\mathbf{I} - \lambda\mathbf{W}_2^*)^{-1}$ exists if $||\lambda\mathbf{W}_2^*|| < 1$.

Now, let $(c_i(\lambda))_{i=1,\dots,n} = (\mathbf{I} - \lambda\mathbf{W}_2^*)^{-1}\mathbf{1}$ and $(d_i(\lambda)')_{i=1,\dots,n} = (\mathbf{I} - \lambda\mathbf{W}_2^*)^{-1}\mathbf{W}_1^*$. In order to denote the dependence on $\boldsymbol{\vartheta}$ we write $h_{\boldsymbol{\vartheta}}(s_i)$, $i = 1, \dots, n$. Then,

$$\log(h_{\boldsymbol{\vartheta}}(s_i)) = \tau^{-1}(\alpha\, c_i(\lambda) + \rho\, d_i(\lambda)'\, \tilde{\boldsymbol{\gamma}}(\log(\mathbf{Y}^{(2)}))).$$

Here, we assume that $c = \mathbb{E}(\log(\varepsilon(s_i)^2))$ is a known quantity. Using $H_i = \log(Y(s_i)^2) - \mathbb{E}(\log(\varepsilon(s_i)^2))$ and $\mathbf{H} = (H_i)_{i=1,\dots,n}$ the estimators of the parameters α, λ, and ρ are obtained by minimizing the non-linear sum of squares

$$\sum_{i=1}^{n} (H_i - \log(h_{\boldsymbol{\vartheta}}(s_i)))^2 = \sum_{i=1}^{n} \left(H_i - \tau^{-1}(\alpha\, c_i(\lambda) + \rho\, d_i(\lambda)'\, \tilde{\boldsymbol{\gamma}}(\mathbf{H} + c\mathbf{1})) \right)^2$$

with respect to $\boldsymbol{\vartheta}$.

Although τ^{-1} is a known function, this minimization problem is complex. Thus, we will impose further assumptions which are fulfilled for all relevant special cases. We will suppose that $\boldsymbol{\gamma}(x) = (\gamma(x_i))$ with a known function γ. Consequently, $\tilde{\boldsymbol{\gamma}}(x) = (\tilde{\gamma}(x_i))_{i=1,\dots,n}$ with $\tilde{\gamma}(x) = \gamma(\exp(x))$, which leads to the easier minimization of

$$Q_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{i=1}^{n}\left(H_i - \tau^{-1}\left(\alpha\, c_i(\lambda) + \rho\, d_i(\lambda)'\,(\tilde{\gamma}(H_v + c))\right)_{v=1,\dots,n}\right)^2. \quad (6)$$

Note that since the (i,i)-th element of $\mathbf{W}_1^*$ is zero it follows that $d_i(\lambda)'\,(\tilde{\gamma}(H_v + c))_{v=1,\dots,n}$ is no function of H_i. Minimization problems of that type have been studied in detail in, e.g., Amemiya (1985), Pötscher and Prucha (1997), and Newey and McFadden (1994). There, sufficient conditions are given for the consistency and asymptotic normality of the resulting estimators under various conditions. Note that, in the present case, $\{H_i\}$ is a strictly stationary process. Moreover, in most papers on this topic, the regression function is assumed to be a deterministic function depending on certain parameters. In the present case, however, it is a function depending on the observations $\{H_i\}$ which makes the analysis of the asymptotic behaviour of the estimators much harder. Further, it must be noted that a spatial problem is present. The positions s_i are points in a space and we need a certain distance measure between these points to assess the dependence of the observations.

Theorem 4 *Suppose that $\boldsymbol{\vartheta}_0 \in \Theta = [\rho_l, \rho_u] \times [\lambda_l, \lambda_u] \times [\alpha_l, \alpha_u] \subseteq [0, 1) \times [0, 1) \times [0, \infty)$. Let $\{\varepsilon(s_i) : i \in \mathbb{N}\}$ be independent and identically distributed random variables with existing moment $\mathbb{E}((\log(|\varepsilon(s_1)|))^2)$. Let f be a differentiable and invertible function with $f' > 0$ on $(0, \infty)$ and let γ be a measurable function on $[0, \infty)$ with $\mathrm{Var}(\gamma(Y(s_1)^2)) < \infty$. Suppose that $f'(h_{\boldsymbol{\vartheta}}(s_i))\, h_{\boldsymbol{\vartheta}}(s_i) \geq L > 0$ for all i, $\boldsymbol{\vartheta} \in \Theta, \omega$. Further assume that for $\boldsymbol{\vartheta} = (\rho, \lambda, \alpha) \in \Theta$*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}(\log(h_{\boldsymbol{\vartheta}}(s_i))) - \mathbb{E}(\log(h_{\boldsymbol{\vartheta}_0}(s_i)))\right)^2 \quad exists, \quad (7)$$

$$\frac{1}{n} \sum_{i=1}^{n} (\log(h_{\vartheta}(s_i)) - \mathbb{E}(\log(h_{\vartheta}(s_i))))^2 \xrightarrow{p} 0 \tag{8}$$

as n tends to infinity and that the limit function in (7) has a unique minimum at ϑ_0. Moreover, let $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ be row-standardized, i.e., $\mathbf{W}_1^* \mathbf{1} = \mathbf{W}_2^* \mathbf{1} = \mathbf{1}$. Then the minimization problem (6) has a solution $\hat{\vartheta}_n$ and it holds that $\hat{\vartheta}_n \xrightarrow{p} \vartheta_0$ as $n \to \infty$.

Note that the solution of (6) does not have to be unique. For more details, we refer to Sect. 4 of Amemiya (1985).

Moreover, for a spGARCH process it holds that $f'(x)x = x$ and $h_{\vartheta}(s_i) \geq \alpha_l > 0$ and thus the above condition is fulfilled. For a log-spGARCH process we have that $f'(x)x = 1$ and thus it is fulfilled as well. Further, for a log-spGARCH process the condition (7) can be easily seen to be fulfilled since for a strictly stationary process $\{Y(s_i)\}$ the quantity $\mathbb{E}(\log(h_{\vartheta}(s_i)))$ does not depend on i at all.

To prove the consistency of the local minimum, the roots of the first derivative of the sum of non-linear squares with respect to the parameters must be zero, i.e.,

$$\frac{\partial Q_n(\vartheta)}{\partial \vartheta} = 0.$$

Theorem 5 *Suppose that the conditions of Theorem 4 are fulfilled and (7) and (8) hold in an open neighbourhood N of ϑ_0.*

Let Θ_T denote the set of roots of the equation

$$\frac{\partial Q_n(\vartheta)}{\partial \vartheta} = 0.$$

Then it holds for all $\varepsilon > 0$ that

$$\lim_{n \to \infty} P \left(\inf_{\vartheta \in \Theta_T} (\vartheta - \vartheta_0)'(\vartheta - \vartheta_0) > \varepsilon \right) = 0.$$

As in the previous section, we will now consider a special case of the general framework, namely an spGARCH model. That is, we choose $f(x) = x I_{[0,\infty]}(x)$ while γ is an arbitrary function satisfying certain conditions.

Lemma 6 *Let $\{Y(s_i)\}$ be an spGARCH process. Suppose that $\Theta = (0, 1) \times (0, 1) \times (0, \infty)$. Let $\{\varepsilon(s_i) : i \in \mathbb{N}\}$ be independent and identically distributed random variables with existing moment $\mathbb{E}(\log(|\varepsilon(s_1)|)^2)$. Let γ be a non-negative measurable function on $[0, \infty)$ with $\mathrm{Var}(\gamma(Y(s_1)^2)) < \infty$. Suppose that $\{Y(s_i)\}$ is strictly stationary. Moreover, let $\mathbf{W}_1$ and $\mathbf{W}_2$ be row-standardized.*

(a) If there is an open neighbourhood N of ϑ_0 such that for all $\vartheta = (\rho, \lambda, \alpha) \in N$ it holds with $\Delta = \left(\gamma(Y(s_j)^2) - \mathbb{E}(\gamma(Y(s_j)^2)) \right)_{j=1,\ldots,n}$ that

$$\frac{1}{n} \Delta' \mathbf{W}_1^{*'} (\mathbf{I} - \lambda \mathbf{W}_2^{*'})^{-1} (\mathbf{I} - \lambda \mathbf{W}_2^*)^{-1} \mathbf{W}_1^* \Delta \xrightarrow{p} 0 \tag{9}$$

as n tends to infinity then the assumption (8) is fulfilled.

(b) If

$$\frac{1}{n}\mathbf{1}'(\mathbf{I} - \lambda\mathbf{W}_2^*)^{-1}\mathbf{W}_1^* \text{Cov}(\mathbf{\Delta})\mathbf{W}_1^{*'}(\mathbf{I} - \lambda\mathbf{W}_2^{*'})^{-1}\mathbf{1} \to 0 \tag{10}$$

as n tends to infinity then the condition (9) is fulfilled.

Note that the assumption (9) is a statement about the topological structure of the underlying space. Moreover, (10) shows that it can be interpreted as an assumption on the underlying autocorrelation structure of the process. It is fulfilled if $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ are sparse to limit the spatial dependence to a manageable degree. Here the choice of the weight matrices is restricted. It is also satisfied if the autocorrelation is weak.

4 Computational implementation and simulation studies

In the following section, we assume the simple parametric setting given by (5). We simulated a spatial GARCH process as specified in Example 2 and the weighting matrices $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ were set as row-standardized Rook contiguity matrices, for which the upper triangular elements were set to zero to avoid negative values of $h(\mathbf{s}_i)$. Thus, the conditions of Theorem 1 are fulfilled. In practice, such processes are relevant to model directional processes, for instance. It is worth mentioning at this point that the condition for invertibility of the spatial GARCH process is restrictive, i.e., to represent $Y(\mathbf{s})$ as a function of all $\varepsilon(\mathbf{s})$ in a closed-form, such that the solutions $Y(\mathbf{s})$ are real-valued. This condition is, however, only sufficient. For estimation, a function for the residual process $\varepsilon(\mathbf{s})$ given all (real-valued) observations $y(\mathbf{s})$ is needed, which is easier for spatial GARCH models; and therefore, we can relax on the assumption of a directional process.

The simulation study is performed on a $d \times d$ spatial unit grid (i.e., $D_{\mathbf{s}} = \{\mathbf{s} = (s_1, s_2)' \in \mathbb{Z}^2 : 1 \leq s_1, s_2 \leq d\}$), resulting in $n = d^2$ observations, with $m = 10000$ replications. The size of this spatial field has been successively increased with $d \in \{5, 10, 15\}$. Moreover, we have considered different settings depending on the data-generating parameters $\boldsymbol{\theta}_0 = (\rho_0, \lambda_0, \alpha_0)'$. For all settings, the errors were independently drawn from a standard normal distribution and α_0 equals 1 (i.e., the spatially constant term), whereas ρ_0 and λ_0 varied across the settings. To be precise, $\rho_0 \in \{0.2, 0.4, 0.7\}$ and $\lambda_0 \in \{0.2, 0.4, 0.7\}$ to have settings with a weak, moderate and large dependence in the conditional spatial heteroscedasticity.

For all three parameters ρ, λ, and α, where $\rho_0 + \lambda_0 < 1$, the average bias and RMSE are shown Tables 2 and 3. We see that the RMSE is decreasing with an increasing number of observations in all cases, while the absolute bias is decreasing in almost all cases. The exceptions are observed for cases where the bias is small. Moreover, the non-linear least squares approach can efficiently be implemented and runs very fast on standard computers, even if the number of observations is large. On average, the computing time for estimating the parameters ranges from 3.4 to 3.8 s for 25 observations, 3.2–4.8 s for 100 observations, and 5.5–8.5 s for 225 observations on a standard notebook. This method is implemented in the R package `spGARCH` (version > 2.0, see also Otto 2019).

Table 2 Bias of the estimates $\hat{\rho}$, $\hat{\lambda}$, $\hat{\alpha}$ for different settings with $\alpha_0 = 1$

		$\rho_0 = 0.2$			$\rho_0 = 0.4$			$\rho_0 = 0.7$		
		$\hat{\rho}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\alpha}$
$\lambda_0 = 0.2$	$d = 5$	0.009	0.072	−0.054	−0.039	0.069	0.034	−0.086	0.070	0.207
	$d = 10$	0.004	0.053	−0.056	−0.017	0.033	0.012	−0.031	0.016	0.141
	$d = 15$	−0.001	0.053	−0.064	−0.011	0.020	0.006	−0.017	0.002	0.103
$\lambda_0 = 0.4$	$d = 5$	0.010	−0.044	0.188	−0.037	−0.004	0.323			
	$d = 10$	0.007	−0.056	0.171	−0.011	−0.029	0.248			
	$d = 15$	0.003	−0.049	0.140	−0.005	−0.028	0.186			
$\lambda_0 = 0.7$	$d = 5$	0.001	−0.067	0.565						
	$d = 10$	−0.002	−0.054	0.486						
	$d = 15$	−0.004	−0.039	0.384						

Table 3 RMSE of the estimates $\hat{\rho}$, $\hat{\lambda}$, $\hat{\alpha}$ for different settings with $\alpha_0 = 1$

		$\rho_0 = 0.2$			$\rho_0 = 0.4$			$\rho_0 = 0.7$		
		$\hat{\rho}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\alpha}$
$\lambda_0 = 0.2$	$d = 5$	0.229	0.332	0.536	0.293	0.318	0.639	0.379	0.300	0.913
	$d = 10$	0.167	0.294	0.468	0.209	0.250	0.505	0.250	0.190	0.654
	$d = 15$	0.131	0.276	0.431	0.158	0.211	0.431	0.181	0.139	0.510
$\lambda_0 = 0.4$	$d = 5$	0.233	0.345	0.784	0.304	0.333	1.055			
	$d = 10$	0.169	0.306	0.697	0.213	0.257	0.835			
	$d = 15$	0.131	0.281	0.633	0.158	0.210	0.692			
$\lambda_0 = 0.7$	$d = 5$	0.244	0.344	1.500						
	$d = 10$	0.170	0.263	1.343						
	$d = 15$	0.126	0.210	1.173						

5 Real-world application: condominium prices in Berlin

In markets that are constrained in space, one can typically expect to find locally varying risks. Typical examples of such markets are real estate and labour. For the former, the property prices are highly dependent upon the location of the real estate and prices in the surrounding areas. Similarly, for the latter, this market is also often constrained in space due to the limited mobility of labourers.

On the one hand, we observe conditional mean levels that vary in space, so-called spatial clusters. That is, both clustered areas of higher prices and lower prices can be observed. On the other hand, we may also expect to find locally varying price risks, which can be considered as local volatility clusters. The proposed spatial GARCH-type models are capable of capturing such spatial dependencies in the conditional variance. This motivates why we consider condominium prices (average prices per square metre) at a fine spatial scale. In particular, we will analyze the relative changes from 1995 to 2014 across all Berlin postcode regions (i.e., $n = 190$), more precisely, the logarithmic returns over entire period to obtain a purely spatial data set of the long-term changes. The data are depicted in Fig. 1. The sample mean for these price changes is 0.8103 with a median of 0.6965. In total, the price changes range from -2.5650 to 7.3131. We can observe a spatial cluster of positive values in the north-western postcode regions. To model this spatial dependence, we consider first-order contiguity matrices $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ giving equal weights to each directly neighbouring region. This choice appears to be the best with respect to the Akaike and Bayesian information criterion. For estimation, we do not need further restrictions on the error process or the weight matrix, such as the triangular weighting scheme in the simulation study. There is a well-defined mapping from the observed process $\{Y(s)\}$ to the errors $\{\varepsilon(s)\}$, and the restrictions were only needed for the inverse relation (i.e., the mapping from $\{\varepsilon(s)\}$ to $\{Y(s)\}$).

However, this fine spatial scale of all postcode regions causes another problem, namely exogenous regressors are often not available or cannot be assigned in the given small-scale resolution. For instance, the average household income could play
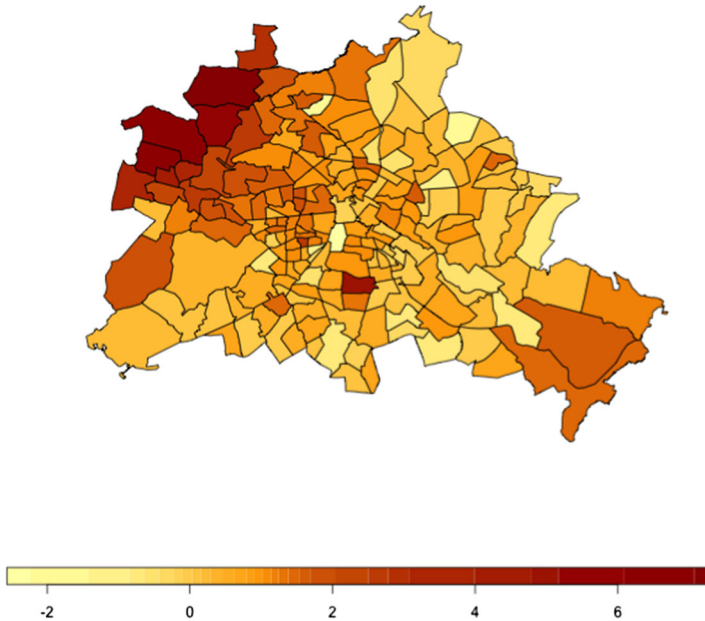
Fig. 1 Long-term logarithmic returns of the condominium prices for all Berlin postcode regions

an important role in the increase of the condominium prices, but the place of living (in terms of postal code areas) does not usually coincide with the place of work. There is no reliable way to associate quantities like personal or household income with postcode areas. Moreover, local infrastructure like schools, leisure facilities, parks is not limited to the residents of the respective areas with the same postal code. Thus, modelling an empirical process on such a small spatial scale is typically prone to heteroscedasticity induced by latent variables.

To illustrate these effects, we applied the developed spatial GARCH model to the residuals of a spatial autoregressive model, briefly SAR (see, e.g. Halleck Vega and Elhorst 2015; Lee 2004), with and without exogenous regressors. More precisely, we select the regressors from a set of potential covariates available for different spatial scales, including the number of crimes (in so-called life-world oriented spaces, LOR, 161 units), number of schools, kinder gardens (postcode-area level), percentage of migrants (LOR level), number of inhabitants (LOR level), size of areas used for infrastructure, living, water, vegetation (district level, 12 units), and average net income per household (district level). The included regressors were chosen such that the Akaike information criterion is minimal. The results of these two models are shown in Table 4 along with the spGARCH coefficients of the error process. All covariates were standardized. Regarding the residuals of these two mean models with and without regressors (before fitting a spGARCH model to the residuals), we observe that both of them are not autocorrelated in space (Moran's $I = -0.0375$ with a p-value of 0.7811 for the intercept-only model, and $I = -0.0124$ with $p = 0.5676$ for the model including covariate effects). That is, the spatial correlation of original data could fully

be modelled ($I = 0.5104$). However, looking at the absolute values of the residuals, we observe that there is significant autocorrelation for both models ($I = 0.1027$ ($p = 0.0042$) and $I = 0.0808$ ($p = 0.0181$) for the intercept-only model and regressive model, respectively).

In the final regressive model, only four regressors were selected, namely a proxy for the available free space (i.e., the proportion of settlement area to total area), the net household income in each district, and linear trends in the east-west direction, and north-south direction have been included (i.e., the coordinates of the centroids of each postal code unit). While a significant negative trend can be observed from west to east, the increase in the north-south direction seems to be of minor importance. Furthermore, the average household income clearly influences the price development of condominiums in Berlin. Condominiums in high-class districts in terms of average income have increased relatively less than condominiums in lower-income areas.

It is worth noting that the dependence in the conditional heteroscedasticity could partly be covered by these covariates and the spatial autocorrelation in the absolute residuals was reduced. Nevertheless, latent effects are present which were not modelled by including these regressors. This can also be seen in Fig. 2, where the residuals of the SAR model are displayed in a diverging colour scheme (blue areas indicate negative residuals, red areas are positive residuals). Even though there is no spatial dependence in the residuals (red and blue areas are irregularly located across space without clustering), we observe a certain dependence in the absolute residuals, because darker colours are clustered together while in other areas the residuals are close to zero indicated by yellow-coloured regions. Thus, an spGARCH model has been fitted in a second step to the residuals of both models. The obtained spGARCH parameters can be interpreted as local model uncertainties and simultaneously cover latent variables, which could not be included due to the fine spatial scale of postcode-area levels.

In both cases, we observe significant positive dependence in the conditional second moments. More precisely, the GARCH effects amount to $\hat{\rho} = 0.2136$ and $\hat{\rho} = 0.2022$ with $\hat{\lambda}$ roughly equal to 0.70 for the intercept-only and regression model, respectively. These parameters can similarly be interpreted as in the time-series case, although $\boldsymbol{h}$ does not necessarily coincide with the conditional second moments (see Otto et al. 2021). Furthermore, $\hat{\alpha}$ does not significantly differ from zero. It is worth noting that, according to (4), $\boldsymbol{h}$ is obtained by multiplying $(\mathbf{I} - \mathbf{W}_1 \mathbf{E} - \mathbf{W}_2)^{-1}$ with $\boldsymbol{\alpha}$, such that a small value of $\boldsymbol{\alpha}$ does not necessarily implies small values of $\boldsymbol{h}$. Analyzing the residuals of this combined model shows that the remaining dependence in the heteroscedasticity could be explained. The squared residuals are no longer significantly autocorrelated in space.

The resulting conditional variance is visualized for the model with regressors in Fig. 3. The highest values of $h(\boldsymbol{s}_i)$ can be observed for the outer regions in north-west and another cluster is located in the southern city centre. This indicates that the highest uncertainty in the price changes is observed for these regions, while there is a band around the city centre where the price changes could more accurately be predicted by the regression model (i.e., the estimated values of $h(\boldsymbol{s}_i)$ are lower). These results seem to be very reasonable because the real-estate market was changing the most in these areas. First, due to increasing prices in the centre, new land for building has been created outside the city—mostly along the regional transport tracks which are mainly

Table 4 Estimated parameters of the spGARCH model for the residuals of a spatial autoregressive model with and without regressors, where the dependent variables are the changes in the condominium prices in Berlin

Parameter	Without regressors		With regressors	
	Estimate	Standard error	Estimate	Standard error
Regression coefficients				
Intercept	0.2002	(0.0814)	0.3392	(0.0917)
Available area	–		− 0.1831	(0.0979)
Net income	–		− 0.2131	(0.0995)
West-east gradient	–		− 0.2605	(0.0809)
North-south gradient	–		0.0917	(0.0716)
SAR process				
γ	0.7456	(0.0588)	0.5758	(0.0789)
spGARCH residuals				
α	1.959×10^{-3}	(0.0122)	1.0014×10^{-5}	(0.0123)
ρ	0.2136	(0.0425)	0.2022	(0.0386)
λ	0.7091	(0.0570)	0.7016	(0.0576)
Summary statistics				
AIC of the mean model	539.9281		532.543	
Moran's I of the residuals (p value)	0.0806 (0.9623)		− 0.0376 (0.7774)	
Moran's I of the squared residuals (p value)	− 0.0820 (0.9721)		− 0.0900 (0.9821)	
Average h	0.6039		0.5404	
Maximum	4.8183		4.0335	

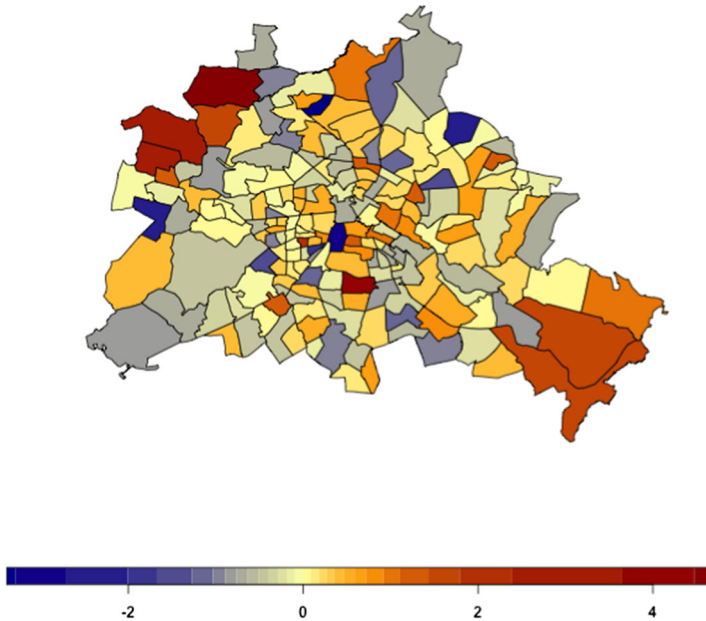All standard errors are given in parentheses

Fig. 2 Residuals of the SAR model

going in an east-west direction. Second, the major airport of West Berlin, namely the airport Berlin-Tempelhof, was closed in 2008 changing the atmosphere in this region, some areas changed from regions in the flight paths to calm regions very close to the city centre, while others were not affected. This explains the second cluster in the centre.

6 Discussion and conclusions

Recently, a few papers have introduced spatial ARCH and GARCH-type models that allow the modelling of an instantaneous spatial autoregressive dependence of heteroscedasticity. In this paper, we propose a generalized spatial ARCH model that additionally covers all previous approaches. Due to the flexible definition of the model as a set of functions, we can derive a common estimation strategy for all these spatial GARCH-type models. It is based on non-linear least squares.

In the second part of the paper, we confirmed our theoretical findings on the consistency of the estimators by means of Monte Carlo simulation studies. The estimation method is computationally implemented in the R package spGARCH. Eventually, the use of the model was demonstrated through an empirical example. More precisely, this paper has shown how the model uncertainties of local price changes in the real estate market in Berlin can be described using an spGARCH model as residuals' process. Though all proposed models are uncorrelated and have a zero mean, potential
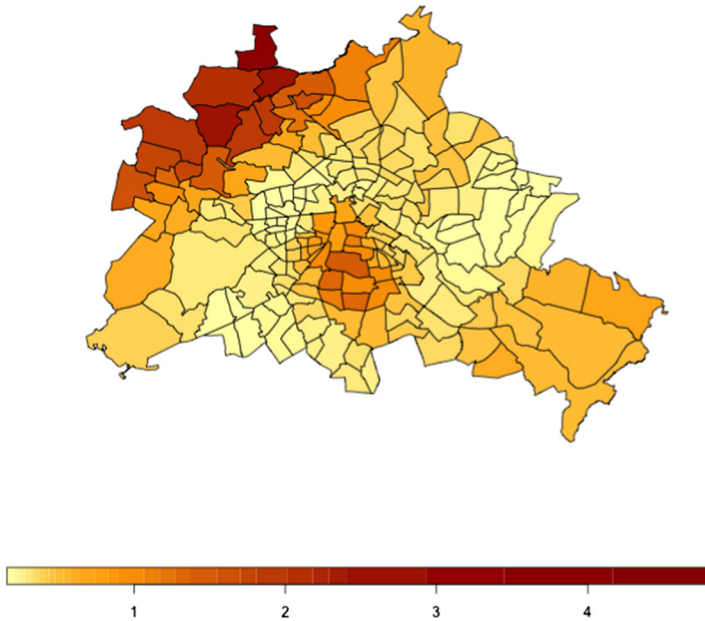
Fig. 3 Spatial equivalent of the conditional variance, $\hat{\boldsymbol{h}}(s)$

interactions between the error process and the mean equation should be analyzed in greater detail in future research.

In addition, we want to stress that the dependence structure does not necessarily have to be interpreted in a spatial sense. Thus, we briefly discuss a further example below, on which the "spatial" proximity could also be defined as the edges of networks. In such cases, $\mathbf{W}_1$ and $\mathbf{W}_2$ would be interpreted as adjacency matrices. For instance, one might consider the financial returns of several stocks as a network, where the only assets that are connected are those that are correlated above a certain threshold. Choosing the threshold equal to 0.5, a financial network as displayed in Fig. 4 can be created. Thus, spGARCH models can be used to analyze various forms of information, whether that might be volatility, risk, or spill-overs from one stock to another, if these assets are close to one another within a certain network. In future research, attempts for modelling volatility clusters within networks, using spatial GARCH models, should be analyzed in greater detail. Moreover, a temporal dimension can be added. The main difference between the spatiotemporal GARCH in our framework and a multivariate vec-GARCH model is that we allow for instantaneous spatial/network interactions, while multivariate GARCH include only temporally lagged cross-variable (i.e., spatial) interactions. Alternatively, in financial applications, the spatial locations could be considered to be unknown. Santi et al. (2021a, b) considered the case of unknown or incompletely known spatial locations for autoregressive models. They propose to approximate the geographical space by another space spanned by certain covariates, which seems to be a promising approach also for spatial GARCH models and financial applications.
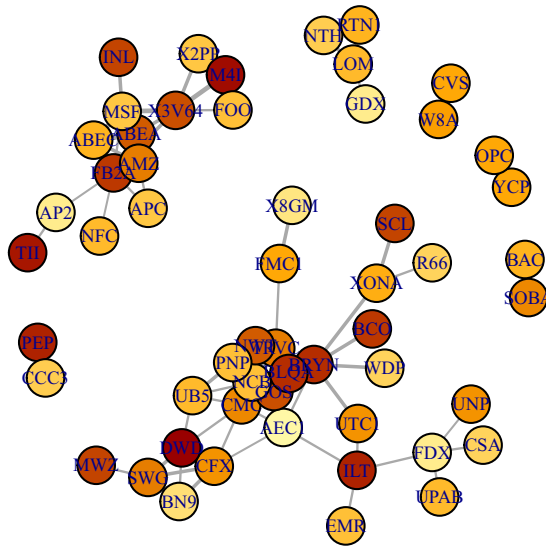
Fig. 4 Financial network of selected stocks of the S&P 500, where the colour of the nodes denotes the annual returns in 2017 with darker colours indicating higher returns

Up to now, we have assumed that suitable functions of the spGARCH model framework are known. Hence, it is possible to maximize certain goodness-of-fit criteria in order to obtain the best-fitting model. However, these functions can also be estimated using a non-parametric approach; for instance by penalized or classical B-splines. Besides, further choices of f have not been discussed in this paper yet, including choices of f to obtain E-spGARCH or logarithmic spGARCH models. Also, multivariate models remain open for future research. This will be the subject of some forthcoming papers.

Appendix A Proofs

Theorem 1 Inserting (1) into (2), we get that

$$F(h) = \alpha + W_1 \gamma(Eh) - (I - W_2) F(h) = 0.$$

Now, we want to know whether there is a solution $\boldsymbol{h}$ of this equation. This is equivalent to the problem of whether or not $\boldsymbol{T}(\omega) \circ \boldsymbol{h}$ has a fixed-point. The existence and the uniqueness of a solution $\boldsymbol{h}$ is an immediate consequence of the Banach fixed-point theorem. Here we make use of the randomized version given as Theorem 7 of Bharucha-Reid et al. (1976). Inserting this solution in (1) provides a measurable solution $\boldsymbol{Y}$. □

Corollary 2 Because $(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))'$ is a measurable function of $(\varepsilon(\boldsymbol{s}_1), \ldots, \varepsilon(\boldsymbol{s}_n))'$, it is strictly stationary as well. □

Theorem 3 (a) Since

$$\boldsymbol{F} = (f(\boldsymbol{h}(\boldsymbol{s}_i)))_{i=1,\ldots,n} = \boldsymbol{\alpha} + \mathbf{W}_1 \boldsymbol{\gamma}(\mathbf{E}\boldsymbol{h}) + \mathbf{W}_2 \boldsymbol{F}$$

its solution $\boldsymbol{h}$ is a function of $\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2$, say $h(\boldsymbol{s}_i) = \xi_i(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)$. Since $Y(\boldsymbol{s}_i) = \varepsilon(\boldsymbol{s}_i)\sqrt{h(\boldsymbol{s}_i)}$ it follows that

$$\begin{aligned}
-Y(\boldsymbol{s}_i) &= -\varepsilon(\boldsymbol{s}_i)\sqrt{\xi_i(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)} \\
&= -\varepsilon(\boldsymbol{s}_i)\sqrt{\xi_i(\varepsilon(\boldsymbol{s}_1)^2, \ldots, (-\varepsilon(\boldsymbol{s}_i))^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)} \\
&\overset{d}{=} Y(\boldsymbol{s}_i)
\end{aligned}$$

since $\boldsymbol{\varepsilon}$ is sign-symmetric. Thus, $Y(\boldsymbol{s}_i)$ is a symmetric random variable. Moreover,

$$(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))' \overset{d}{=} (-Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))'.$$

Thus, $\mathbb{E}(Y(\boldsymbol{s}_1)^{2k-1}|Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)) = \mathbb{E}(-Y(\boldsymbol{s}_1)^{2k-1}|Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n))$. Consequently, this quantity is zero.

(b) Now,

$$\begin{aligned}
Y(\boldsymbol{s}_i)Y(\boldsymbol{s}_j) &= \varepsilon(\boldsymbol{s}_i)\varepsilon(\boldsymbol{s}_j)\sqrt{\xi_i(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)}\sqrt{\xi_j(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)} \\
&\overset{d}{=} -\varepsilon(\boldsymbol{s}_i)\varepsilon(\boldsymbol{s}_j)\sqrt{\xi_i(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)}\sqrt{\xi_j(\varepsilon(\boldsymbol{s}_1)^2, \ldots, \varepsilon(\boldsymbol{s}_n)^2)} \\
&= -Y(\boldsymbol{s}_i)Y(\boldsymbol{s}_j)
\end{aligned}$$

and thus $\text{Cov}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)) = 0$ for $i \neq j$.

□

Proof of Theorem 4 To prove the above theorem we make use of Newey and McFadden (1994). First we observe that Θ is a compact set. Moreover, $Q_n(\boldsymbol{\vartheta})$ (see (6)) is a continuous function for $\boldsymbol{\vartheta} \in \Theta$. Now we have to show that

(i) $Q_n(\boldsymbol{\vartheta}) \overset{p}{\to} Q_0(\boldsymbol{\vartheta})$ for all $\boldsymbol{\vartheta} \in \Theta$

(ii) there is $\tau > 0$ and $\hat{B}_n = O_p(1)$ such that

$$|Q_n(\tilde{\boldsymbol{\vartheta}}) - Q_n(\boldsymbol{\vartheta})| \leq \hat{B}_n ||\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}||^\tau \quad \forall \tilde{\boldsymbol{\vartheta}}, \boldsymbol{\vartheta} \in \Theta.$$

If (i) and (ii) hold then it follows with Lemma 2.9 of Newey and McFadden (1994) that $sup_{\vartheta \in \Theta} |Q_n(\vartheta) - Q_0(\vartheta)| \xrightarrow{P} 0$ as $n \to \infty$.

If additionally

(iii) $Q_0(\vartheta)$ is continuous for $\vartheta \in \Theta$

holds, then we get with Lemma 3 of Amemiya (1985) that $\hat{\vartheta}_n \xrightarrow{P} \vartheta_0$ as $n \to \infty$ and the result is proved.

We start with proving (i). Recall that $H_i = \log(h_{\vartheta_0}(s_i)) + \log(\varepsilon(s_i)^2) - \mathbb{E}(\log(\varepsilon(s_i)^2)) = \eta(s_i) + \log(h_{\vartheta_0}(s_i))$. Thus, in the present case, the function $Q_0(\vartheta)$ can be found by splitting $Q_n(\vartheta)$ into eight parts, as follows

$$Q_n(\vartheta) = I_n + II_n + III_n + IV_n + V_n + VI_n + VII_n + VIII_n$$

with

$$I_n = \frac{1}{n}\sum_{i=1}^{n}\eta(s_i)^2, \quad II_n = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}(\log(h_\vartheta(s_i))) - \mathbb{E}(\log(h_{\vartheta_0}(s_i))))\right)^2,$$

$$III_n = \frac{1}{n}\sum_{i=1}^{n}\left(\log(h_{\vartheta_0}(s_i)) - \mathbb{E}(\log(h_{\vartheta_0}(s_i))))\right)^2$$

$$IV_n = \frac{1}{n}\sum_{i=1}^{n}\left(\log(h_\vartheta(s_i)) - \mathbb{E}(\log(h_\vartheta(s_i))))\right)^2$$

$$V_n = \frac{2}{n}\sum_{i=1}^{n}\eta(s_i)\left(\log(h_{\vartheta_0}(s_i)) - \log(h_\vartheta(s_i))\right)$$

$$VI_n = \frac{2}{n}\sum_{i=1}^{n}(\mathbb{E}(\log(h_{\vartheta_0}(s_i))) - \mathbb{E}(\log(h_\vartheta(s_i))))(\log(h_{\vartheta_0}(s_i)) - \mathbb{E}(\log h_{\vartheta_0}(s_i))),$$

$$VII_n = -\frac{2}{n}\sum_{i=1}^{n}(\mathbb{E}(\log(h_{\vartheta_0}(s_i))) - \mathbb{E}(\log(h_\vartheta(s_i))))(\log(h_\vartheta(s_i)) - \mathbb{E}(\log h_\vartheta(s_i))),$$

$$VIII_n = -\frac{2}{n}\sum_{i=1}^{n}(\log(h_{\vartheta_0}(s_i)) - \mathbb{E}(\log h_{\vartheta_0}(s_i)))(\log(h_\vartheta(s_i)) - \mathbb{E}(\log h_\vartheta(s_i))).$$

The weak law of large numbers (WLLN) implies that I_n converges in probability to $\text{Var}(\log(\varepsilon(s_1)^2))$. II_n is a special case of IV_n. Further

$$V_n/2 = \frac{1}{n}\sum_{i=1}^{n}\eta(s_i)\left(\log(h_{\vartheta_0}(s_i)) - \log(h_\vartheta(s_i))\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\eta(s_i)\left(E\left(\log(h_{\vartheta_0}(s_i))\right) - \mathbb{E}((\log(h_\vartheta(s_i))))\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\eta(s_i)\left(E\left(\log(h_\vartheta(s_i))\right)-\log(h_\vartheta(s_i))\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\eta(s_i)\left(\log(h_{\vartheta_0}(s_i))-E\left(\log(h_{\vartheta_0}(s_i))\right)\right)$$

$$=V_{n,1}+V_{n,2}+V_{n,3}.$$

Since $\{\eta(s_i)\}$ are independent and identically distributed, it follows that

$$P(|V_{n,1}|>\varepsilon)\leq\frac{\mathbb{E}(\eta(s_1)^2)}{n^2\varepsilon^2}\sum_{i=1}^{n}\left(\mathbb{E}(\log(h_\vartheta(s_i)))-\mathbb{E}(\log(h_{\vartheta_0}(s_i)))\right)^2\to0$$

as n tends to ∞ and thus $V_{n,1}$ converges in probability to zero. Using the Cauchy–Schwarz inequality and (8), we get that $V_{n,2}$ and $V_{n,3}$ converge in probability to zero. Thus, $V_n\xrightarrow{p}0$ as $n\to\infty$.

Moreover, the mixed quantities VI_n, VII_n, and $VIII_n$ convergence to zero in probability. This is obtained by applying the Cauchy-Schwarz inequality and by making use of the previous results and assumptions.

Consequently, if $\tilde{Q}_0(\vartheta)$ denotes the limit in (7), then $Q_0(\vartheta)=\mathrm{Var}(\log(\varepsilon(s_1)^2))+\tilde{Q}_0(\vartheta)$ and part i) is proved.

Next, we prove part (ii). We use that $\tau'(\tau^{-1}(x))=f'(f^{-1}(x))f^{-1}(x)$ and, thus, $\tau'(\tau^{-1}(\alpha c_i(\lambda)+\rho d_i(\lambda)(\tilde{\gamma}(H_v+c)))_{v=1,\dots,n})=f'(h_\vartheta(s_i))h_\vartheta(s_i)$. Consequently,

$$Q_n(\tilde{\vartheta})-Q_n(\vartheta)=(\tilde{\vartheta}-\vartheta)'Q_n'(\vartheta^*)$$

with

$$Q_n'(\vartheta^*)=\begin{pmatrix}-\frac{2}{n}\sum_{i=1}^{n}\frac{H_i-\log(h_{\vartheta^*}(s_i))}{f'(h_{\vartheta^*}(s_i))\,h_{\vartheta^*}(s_i)}\,d_i(\lambda^*)'\,(\tilde{\gamma}(H_v+c))\\-\frac{2}{n}\sum_{i=1}^{n}\frac{H_i-\log(h_{\vartheta^*}(s_i))}{f'(h_{\vartheta^*}(s_i))\,h_{\vartheta^*}(s_i)}\,(\alpha^*c_i'(\lambda^*)+\rho^*d_i'(\lambda^*)'\,(\tilde{\gamma}(H_v+c)))\\-\frac{2}{n}\sum_{i=1}^{n}\frac{H_i-\log(h_{\vartheta^*}(s_i))}{f'(h_{\vartheta^*}(s_i))\,h_{\vartheta^*}(s_i)}\,c_i(\lambda^*)\end{pmatrix}.$$

Since $H_i=\eta(s_i)+\log(h_{\vartheta_0}(s_i))$, we get that

$$Q_n'(\vartheta^*)=\begin{pmatrix}-\frac{2}{n}\sum_{i=1}^{n}\frac{\eta(s_i)+\log(h_{\vartheta_0}(s_i))-\log(h_{\vartheta^*}(s_i))}{f'(h_{\vartheta^*}(s_i))\,h_{\vartheta^*}(s_i)}\,d_i(\lambda^*)'\,(\gamma(Y(s_v)^2))\\-\frac{2}{n}\sum_{i=1}^{n}\frac{\eta(s_i)+\log(h_{\vartheta_0}(s_i))-\log(h_{\vartheta^*}(s_i))}{f'(h_{\vartheta^*}(s_i))\,h_{\vartheta^*}(s_i)}\,(\alpha^*c_i'(\lambda^*)+\rho^*d_i'(\lambda^*)'\,(\gamma(Y(s_v)^2)))\\-\frac{2}{n}\sum_{i=1}^{n}\frac{\eta(s_i)+\log(h_{\vartheta_0}(s_i))-\log(h_{\vartheta^*}(s_i))}{f'(h_{\vartheta^*}(s_i))\,h_{\vartheta^*}(s_i)}\,c_i(\lambda^*)\end{pmatrix}.$$

Note that

$$(c_i(\lambda))=\left(\mathbf{I}+\lambda\mathbf{W}_2^*+\lambda^2\mathbf{W}_2^{*2}+\cdots\right)\mathbf{1}=\frac{1}{1-\lambda}\mathbf{1}.$$

This quantity is well defined since $||\lambda\mathbf{W}_2^*||_\infty = \lambda < 1$. Thus, $c_i(\lambda_l) \le c(\lambda) \le c(\lambda_u)$. Further,

$$\boldsymbol{d}_i(\lambda)' = \left(\mathbf{I} + \lambda\mathbf{W}_2^* + \lambda^2\mathbf{W}_2^{*2} + \cdots\right)\mathbf{W}_1^*$$

is a non-decreasing function in λ for each component of the matrix and $\boldsymbol{d}_i(\lambda)'\mathbf{1} = 1/(1-\lambda)$.

We obtain with the inequality of Cauchy–Schwarz that

$$|Q_n(\tilde{\boldsymbol{\vartheta}}) - Q_n(\boldsymbol{\vartheta})| \le ||\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})||_2 ||Q_n'(\boldsymbol{\vartheta}^*)||_2$$

where $||\cdot||_2$ stands for the Euclidean distance. Each component of $Q_n'(\boldsymbol{\vartheta}^*)$ is stochastically bounded. For the first component, this can be seen by applying Cauchy-Schwarz and making use of the fact that

$$\max_{\lambda\in[\lambda_l,\lambda_u]} \frac{1}{n}\sum_{i=1}^n \left(\boldsymbol{d}_i(\lambda)'\left(\gamma(Y(\boldsymbol{s}_v)^2)\right)\right)^2$$
$$\le \frac{1}{n}\sum_{i=1}^n\sum_{j,v=1}^n d_{ij}(\lambda_u)d_{iv}(\lambda_u)|\gamma(Y(\boldsymbol{s}_j)^2)||\gamma(Y(\boldsymbol{s}_v)^2)|$$

and, thus,

$$P\left(|\max_{\lambda\in[\lambda_l,\lambda_u]}\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{d}_i(\lambda)'\left(\gamma(Y(\boldsymbol{s}_v)^2)\right)\right)^2 > K\right) \le \frac{\mathbb{E}((\gamma(Y(\boldsymbol{s}_1)^2))^2)}{(1-\lambda_u)^2K}.$$

For the other components, the argumentation is similar. Thus, we have proved (ii).

Finally, we come to part (iii). Let $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \in \Theta$. Since by (ii)

$$|Q_n(\boldsymbol{\vartheta}_2) - Q_n(\boldsymbol{\vartheta}_1)| \le \hat{B}_n||\boldsymbol{\vartheta}_2 - \boldsymbol{\vartheta}_1||_2$$

with $\hat{B}_n = O_p(1)$ and since because of the uniform convergence there is a subsequence such that $sup_{\boldsymbol{\vartheta}\in\Theta}|Q_{r_n}(\boldsymbol{\vartheta}) - Q_0(\boldsymbol{\vartheta})| \overset{a.s.}{\to} 0$ as $n \to \infty$ it follows with a constant K that

$$|Q_0(\boldsymbol{\vartheta}_2) - Q_0(\boldsymbol{\vartheta}_1)| \le \lim_{n\to\infty}|Q_{r_n}(\boldsymbol{\vartheta}_2) - Q_{r_n}(\boldsymbol{\vartheta}_1)| \le K||\boldsymbol{\vartheta}_2 - \boldsymbol{\vartheta}_1||_2$$

and thus Q_0 is continuous. Part (iii) is proved and the proof of the Theorem is finished. □

Proof of Theorem 5 Following Theorem 4.1.2 in Amemiya (1985), we have to prove that there is an open neighbourhood $N \subseteq \Theta$ of $\boldsymbol{\vartheta}$ with

$$Q_n(\boldsymbol{\vartheta}) \overset{p}{\to} Q_0(\boldsymbol{\vartheta}) \text{ as } n \to \infty \text{ uniformly in } \boldsymbol{\vartheta} \in N.$$

In order to prove this we follow the proof of Theorem 4. It follows under the modified assumptions as above that $Q_n(\vartheta) \xrightarrow{p} Q_0(\vartheta)$ for $\vartheta \in N$ as $n \to \infty$. Using (ii) of the proof of Theorem 4, we get that the convergence is uniform on a closed set $N_0 \subset N$. Thus it is also uniform on an open subset of N_0 since Θ is a subspace of the Euclidean space.

The rest follows as in the proof of Theorem 4. $\qquad\qquad\qquad\qquad\qquad\square$

Proof of Lemma 6 First, we prove part a). Since $f(x) = x$ we get that $\tau^{-1} = \log(x)$ and

$$\log(h_\vartheta(s_i) = \log(\alpha c_i(\lambda) + \rho d_i(\lambda)' \left(\gamma(Y(s_i)^2)\right).$$

It holds for $z > -1$ that

$$\log(1+z) = z + \frac{z^2}{2}\left(-\frac{1}{1+\varsigma^2}\right), \quad |\varsigma| \le z \qquad (A1)$$

and therefore $|\log(1+z) - z| \le z^2$.

Let $Z_{in} = \frac{d_i(\lambda)'}{c_i(\lambda)}\left(\gamma(Y(s_i)^2)\right)$, $\kappa = \rho/\alpha$, and $\kappa_i = \kappa/(1 + \kappa \mathbb{E}(Z_{in}))$. Since $\{\gamma(Y(s_i)^2)\}$ is strictly stationary as well it follows that $\mathbb{E}(Z_{in}) = E(\gamma(Y(s_1)^2))\, d_i(\lambda)'\, \mathbf{1}/c_i(\lambda) = \mathbb{E}(\gamma(Y(s_1)^2))/c_i(\lambda)$ and thus

$$\kappa_i = \frac{\kappa c_i(\lambda)}{c_i(\lambda) + \kappa\, \mathbb{E}(\gamma(Y(s_1)^2))} = \frac{\kappa}{1 + \kappa(1-\lambda)\, \mathbb{E}(\gamma(Y(s_1)^2))}$$

does not depend on i at all. We briefly write κ_1.

Then the left side of (8) is equal to

$$\frac{1}{n}\sum_{i=1}^{n}(\log(1 + \kappa Z_{in}) - \mathbb{E}(\log(1 + \kappa Z_{in}))$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\log(1 + \kappa_1(Z_{in} - \mathbb{E}(Z_{in}))) - E(\log(1 + \kappa_1(Z_{in} - \mathbb{E}(Z_{in})))))$$

$$= I_n - \mathbb{E}(I_n).$$

Using (A1) we get that $I_n = II_n + III_n$ with $II_n = \frac{\kappa_1}{n}\sum_{i=1}^{n}(Z_{in} - \mathbb{E}(Z_{in}))$ and

$$|III_n| \le IV_n = \frac{\kappa_1^2}{n}\sum_{i=1}^{n}(Z_{in} - \mathbb{E}(Z_{in}))^2).$$

Since

$$\sum_{i=1}^{n}(Z_{in} - \mathbb{E}(Z_{in}))^2) = (1-\lambda)^2\sum_{i=1}^{n}(d_i(\lambda)'\boldsymbol{\Delta}_i)^2$$

it follows with (9) that $IV_n \overset{p}{\to} 0$. Because

$$\left| \frac{1}{n} \sum_{i=1}^{n} (Z_{in} - \mathbb{E}(Z_{in})) \right| \leq \left(\frac{1}{n} \sum_{i=1}^{n} (Z_{in} - \mathbb{E}(Z_{in}))^2 \right)^{1/2}$$

it holds that $II_n \overset{p}{\to} 0$ as well and part a) is proved.

To prove part b) the Markov inequality is applied. It holds that

$$P \left(\frac{1}{n} \sum_{i=1}^{n} (Z_{in} - \mathbb{E}(Z_{in})) \right)^2 > \varepsilon \right) \leq \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(Z_{in})$$

and

$$\sum_{i=1}^{n} \mathrm{Var}(Z_{in}) = (1 - \lambda)^2 \mathbf{1}'(\mathbf{I} - \lambda \mathbf{W}_2^*)^{-1} \mathbf{W}_1^* \mathrm{Cov}(\mathbf{\Delta}) \mathbf{W}_1^{*\prime} (\mathbf{I} - \lambda \mathbf{W}_2^{*\prime})^{-1} \mathbf{1}$$

and thus the result follows. $\qquad\square$

References

Amemiya T (1985) Advanced econometrics. Harvard University Press, Cambridge

Basak GK, Bhattacharjee A, Das S (2018) Causal ordering and inference on acyclic networks. Empir Econ 55(1):213–232

Bharucha-Reid AT et al (1976) Fixed point theorems in probabilistic analysis. Bull Am Math Soc 82(5):641–657

Bhattacharjee A, Jensen-Butler C (2013) Estimation of the spatial weights matrix under structural constraints. Reg Sci Urban Econ 43(4):617–634

Billio M, Caporin M, Frattarolo L et al (2021) Networks in risk spillovers: a multivariate GARCH perspective. Econ Stat

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

Borovkova S, Lopuhaa R (2012) Spatial GARCH: a spatial approach to multivariate volatility modeling. Available at SSRN 2176781

Campbell JY, Lo AW, MacKinlay AC (1997) The econometrics of financial markets. Princeton University Press, Princeton

Caporin M, Paruolo P (2006) GARCH models with spatial structure. SIS Stat 447–450

Caporin M, Paruolo P (2015) Proximity-structured multivariate volatility models. Economet Rev 34(5):559–593

Cressie N (1993) Statistics for spatial data. Wiley

Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, Hoboken

Elhorst JP (2010) Applied spatial econometrics: raising the bar. Spat Econ Anal 5(1):9–28. https://doi.org/10.1080/17421770903541772

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

Fingleton B (2008) A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. Spat Econ Anal 3(1):27–44. https://doi.org/10.1080/17421770701774922

Fingleton B (2009) Spatial autoregression. Geogr Anal 41(4):385–391

Francq C, Wintenberger O, Zakoian JM (2013) GARCH models without positivity constraints: exponential or log GARCH? J Econ 177(1):34–46

Getis A (2009) Spatial weights matrices. Geogr Anal 41(4):404–410

Halleck Vega S, Elhorst JP (2015) The SLX model. J Reg Sci 55(3):339–363

Hanš O (1957) Reduzierende zufällige Transformationen. Czechoslov Math J 7(1):154–158

Hølleland S, Karlsen HA (2020) A stationary spatio-temporal GARCH model. J Time Ser Anal 41(2):177–209

Lee LF (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6):1899–1925. https://doi.org/10.1111/j.1468-0262.2004.00558.x

Manski CF (1993) Identification of endogenous social effects: the reflection problem. Rev Econ Stud 60(3):531–542. https://doi.org/10.2307/2298123

Merk MS, Otto P (2021) Directional spatial autoregressive dependence in the conditional first-and second-order moments. Spatial Stat 41

Newey K, McFadden D (1994) Large sample estimation and hypothesis. In: McFadden DL, Engle RF (ed) Handbook of econometrics, IV, pp 2112–2245

Otto P (2019) spGARCH: An R-package for spatial and spatiotemporal ARCH models. R J 11(2):401–420

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 (2021) Stochastic properties of spatial and spatiotemporal ARCH models. Stat Pap 62(2):623–638

Otto P, Steinert R (2022) Estimation of the spatial weighting matrix for spatiotemporal data under the presence of structural breaks. J Comput Graph Stat. https://doi.org/10.1080/10618600.2022.2107530

Pötscher BM, Prucha IR (1997) Dynamic nonlinear econometric models: asymptotic theory. Springer, New York

Santi F, Dickson MM, Espa G Taufer E, Mazzitelli A (2021a) Handling spatial dependence under unknown unit locations. Spat Econ Anal 16(2):194–216

Santi F, Dickson MM, Giuliani D, Arbia G, Espa G (2021b) Reduced-bias estimation of spatial autoregressive models with incompletely geocoded data. Comput Stat 36(4):2563–2590

Sato T, Matsuda Y (2017) Spatial autoregressive conditional heteroskedasticity models. J Jpn Stat Soc 47(2):221–236

Sato T, Matsuda Y (2020) Spatial extension of generalized autoregressive conditional heteroskedasticity models. Spatial Econ Anal 1–13

Taşpınar S, Doğan O, Chae J et al (2021) Bayesian inference in spatial stochastic volatility models: an application to house price returns in Chicago. Oxford Bull Econ Stat

Tan KK, Yuan XZ (1997) Random fixed point theorems and approximation. Stoch Anal Appl 15(1):103–123

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(sup1):234–240