Dynamic Spatial Autoregressive Models with Time–varying Spatial Weighting Matrices

Anna Gloria Billé[a], Francisco Blasques[b], Leopoldo Catania*[c]

[a] *Department of Statistical Sciences, University of Padova, Padova, Italy*
[b] *Vrije Universiteit Amsterdam, ITSEF and Tinbergen Institute, Amsterdam, The Netherlands*
[c] *Department of Economics and Business Economics and CREATES, Aarhus University, Aarhus, Denmark*

Abstract

We propose a new spatio–temporal model with time–varying spatial weighting matrices. The filtering procedure of the time–varying unknown parameters is performed using the information contained in the score of the conditional distribution of the observables. We provide conditions for the stationarity and ergodicity of the filtered sequence of the spatial matrices as well as for the consistency and asymptotic normality of the maximum likelihood estimator (MLE). An extensive Monte Carlo simulation study to investigate the finite sample properties of the maximum likelihood estimator is also reported. We finally analyze the association between eight European countries' perceived risk, suggesting that the economically strong countries have their perceived risk increased due to their spatial connection with the economically weaker countries. We also investigate the evolution of the spatial connection between the house prices in different areas of the UK, identifying periods when the usually adopted sparse weighting matrix is not sufficient to describe the underlying spatial process.

Keywords: Dynamic spatial autoregressive models, Time–varying weighting matrices, Distance decay functions

* Department of Economics and Business Economics and CREATES, Aarhus University, Aarhus, Denmark. leopoldo.catania@econ.au.dk, Phone: +45 8716 5536, web page: http://pure.au.dk/portal/en/leopoldo.catania@econ.au.dk.

Francisco Blasques is thankful to the Dutch Science Foundation NWO for financial support (Vidi.195.099).

Preprint submitted to *October 28, 2020*

1. Introduction

Estimation of high–dimensional covariance matrices has received a greater attention, especially in the era of Big Data. Within the spatial statistics literature, the spatio–temporal covariance and cross–covariance functions among spatial units are estimated by typically assuming an isotropic spatial process (Porcu et al., 2016). Most of these model specifications are useful in geostatistical sciences where the surface of the random field is continuous. With discrete random fields, i.e. where the spatial statistical units are referred to regular/irregular areal data, the weighting matrix is a way to connect these units by using different well–known criteria (i.e. contiguity criteria, k–nearest neighbors, distances, etc.). However, in most of the cases the weighting matrix is assumed to be exogenous, typically based on distances between geographical locations.

In the spatial econometrics literature, estimating the W matrices poses at least two problems. The first is that, by treating W as a parameter, the spatial autoregressive coefficient, ρ, is generally not identified. The second is that, in the case of cross–sectional or panel data with $T << N$, the estimation of $N \times (N-1)/2$ parameters is intractable. A possible solution has been proposed by LeSage and Pace (2007), who rely on a different spatial econometric model called *matrix exponential spatial specification (MESS)* which allows us to estimate the $N \times (N-1)/2$ parameters of W. However, the estimation is feasible only as long as the spatial dimension N is relatively small. In this regard, a two–step residual regression estimator based on spectral decomposition of the error covariance matrix in the first stage has recently been proposed by Bhattacharjee and Jensen-Butler (2013). The reader is also referred to the LASSO procedure proposed by Ahrens and Bhattacharjee (2015) and Otto and Steinert (2018).

Alternatively, spatial effects can be based on the specification of distance decay functions rather than by considering simple first–order sparse matrices. Indeed, with negative exponential or inverse distance functions, all the spatial units are connected each other but spatial units located at higher distances are less related. However, even if there are theoretical reasons indicating that spatial interaction effects can be described as a smooth variation, it is often not clear the degree at which the spatial dependence between units diminishes as distance increases. With weighting matrices defined as inverse distance functions to the power of an unknown parameter, say γ, we can interpret γ as a distance decay parameter and determine the radius–effect in space through which the interaction effects tend to rapidly diminish. Alternatively, one can use negative exponential functions or any

continuous monotonically decreasing function that ensures decreasing weights as distances increase. Then, we generally consider the possibility that interactions may continue even over a first or second–order neighborhood, but we do not know the radius at which we cut off the spatial series because of the unknown power that scales the relative importance of the geographic distances.

In this paper, we propose to estimate, simultaneously with the other parameters, the evolution of the distance decay parameter γ_t of a dynamic symmetric $\mathsf{W}(\gamma_t, \mathbf{d})$ matrix parameterized in terms of a generic distance function, by exploiting the recent advancements in Score Driven (SD) models. The SD framework of Creal et al. (2013) and Harvey (2013) allows us to update a set of time–varying unobserved parameters using the information contained in the scaled score of the conditional distribution of the observables. The use of the score to track the conditional distribution of a random variable over time has been proved to be optimal in a realized Kullback–Leibler sense, see e.g. Blasques et al. (2015). In this way, our paper provides a new promising method for analyzing and forecasting spatial correlation structures and could be used in a wide range of empirical spatio–temporal applications. Our model specification is also computationally feasible even with large N and T, and it can be viewed as a generalization of the ones proposed by Blasques et al. (2016b) and Catania and Billé (2017), in which the W is simply prespecified. Convergence of the initialized spatial time varying matrix recursion to a unique stationary and ergodic solution is proven along with consistency and asymptotic normality of the Maximum Likelihood Estimator (MLE). Theoretical results are presented for N fixed and T large ($T \to \infty$), which match the configuration of our empirical analysis. A Monte Carlo study investigates the finite sample properties of the MLE as well as the ability of the model to recover several artificial dynamics of γ_t.

We finally provide two empirical examples showing how the proposed model can be applied. In the first application we study the spatial association between the European perceived risk proxied by the credit default swap spreads of eight European countries. We show that the examined European countries are heterogeneous with respect to their optimal level of connection with other countries, and that economically strong countries like Germany, the Netherlands, and France have their perceived risk increased due to their spatial connection with economically weaker countries like Ireland, Portugal, and Spain. Our second empirical application aims at providing general advice for common spatial econometrics analysis based on a pre–specified distance matrix by using UK House Price datasets. Through the evolution of γ_t, we highlight periods of time in which a sparse W matrix

is not an appropriate choice to correctly model the spillover effects.

The rest of the paper is structured in the following way. Section 2 explains the relation between the present paper and its closest recent literature. Section 3 detailes the Dynamic Spatial Weighting Matrix (DSWM) model and describes the filtering procedure for the time–varying distance decay parameter. Section 4 shows different possible parametrization of the time–varying weighting matrix and explains the role of the time–varying distance decay parameter. Section 5 reports the statistical properties of the proposed model. Section 6 investigates the finite sample properties of the ML estimator and reports a simulation experiment illustrating the filtering ability of the proposed model. Section 7 analyzes two different empirical applications. Finally, Section 8 concludes.

2. Literature Review

In this section, we briefly review the closest contributions in the literature, distinguishing them according to the type of weighting matrix, estimation method and model specification. In all the following cases the elements of the weighting matrix are estimated in some way.

For the endogenous treatment of W in static cross–sectional spatial models the reader is referred to Qu and Lee (2015) and Benjanuvatra et al. (2015). Qu and Lee (2015) proposed three estimation procedures – QMLE, 2SIV, GMM – for a SAR(1) specification where the weighting matrix depends on some observables. However, with the exception of the QMLE case, additional orthogonality conditions are needed when considering additional exogenous regressors to determine W. Indeed, those estimators might not achieve the desidered statistical properties unless the additional assumptions are validated. Moreover, finding proper exogenous regressors is empirically most of the time not feasible. An alternative is the parametrisation proposed by Benjanuvatra et al. (2015). In this case, the weighting matrix is defined as a function of a static unknown parameter $\mathsf{W}(\gamma)$ and the estimation procedure lies in the QMLE framework. In terms of parametrization, our paper is the natural spatio–temporal extension of the paper by Benjanuvatra et al. (2015).

Model estimation methods with time–varying W_t matrices within spatial dynamic panel data models have been recently introduced by Lee and Yu (2012), Kelejian and Piras (2014), Wang and Yu (2015), Han and Lee (2016), Qu et al. (2016), and Qu et al. (2017)[1]. Kelejian and Piras (2014)

[1]For contributions that also consider common factors to estimate the weighting matrix, the reader is referred to Pesaran and Tosetti (2011) and Shi and Lee (2017a).

and Qu et al. (2016) proposed IV estimators, whereas the works by Han and Lee (2016) and Qu et al. (2017) relied on Bayesian estimation and QMLE, respectively. In particular, similarly to Qu and Lee (2015), Qu et al. (2017) suggested to define the weighting matrix as a bounded function of some time–varying observables which follow a VAR process. In this case, a $p \times p$ matrix of coefficients should be estimated togheter with the other parameters, where p is the number of additional observables, making this procedure not computationally efficient unless p is small.

Finally, Lasso estimation is another way to estimate the elements of the weighting scheme, see Ahrens and Bhattacharjee (2015) and Otto and Steinert (2018). In the former case the authors developed a two–step procedure for a spatial panel data model with $N >> T$, whereas in the latter case a two–step approach for a spatio–temporal model with $N << T$ and potential structural breaks in W has been implemented.

In this paper, we propose to parametrise the weighting matrix such that it is defined as a monotonically decreasing function of a time–varying parameter, i.e. γ_t. Showing the evolution of γ_t is basically an alternative way to estimate the elements of the weighting matrix over time by exploiting the time series information rather than that coming from additional exogenous regressors. Moreover, our score driven model specification allows us to simultaneously estimating all the parametrs at each point in time, and also updating the estimates through the filtering process over time, rather than using two–step procedures.

3. The Dynamic Spatial Weighting Matrix Model

In this section we extend a (first–order) spatial autoregressive model with heteroskedastic innovations, i.e. SAR(1), by allowing for a time–varying spatial weighting matrix of unknown order of proximity. Let $\mathbf{y}_t$ be an N–dimensional stochastic vector of spatial variables located on a possibly unevenly spaced lattice $Z \subseteq \Re^N$ at time t. We assume that $\mathbf{y}_t$ is generated according to a SAR(1) model, see e.g. Bao and Ullah (2007). The Dynamic Spatial Weighting Matrix (DSWM) model is then defined as

$$\mathbf{y}_t = \rho \mathsf{W}\left(\gamma_t, \mathbf{d}\right) \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{D}\left(\mathbf{0}, \boldsymbol{\Sigma}, \psi\right), \tag{1}$$

where $\mathbf{X}_t = (\mathbf{x}_{j,t}; j = 1, \ldots, K)$ is an $N \times K$ matrix of exogenous covariates with the associated vector of coefficients $\boldsymbol{\beta} = (\beta_j; j = 1, \ldots, K)'$, $\boldsymbol{\varepsilon}_t$ is an N–valued stochastic vector of spatially located innovations at time t with continuous distribution $\mathcal{D}(\mathbf{0}, \boldsymbol{\Sigma}, \psi)$ with shape parameter ψ, such that

$\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma} = \text{diag}\left(\sigma_i^2; i = 1, \ldots, N\right)'$.[2] In Subsection 3.1 we extend model (1) to the case with (time) heteroscedastic innovations, i.e. $\boldsymbol{\Sigma}_t$.

In this paper, we consider two different distributions $\mathcal{D}(\cdot)$: (i) the multivariate Normal distribution $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and, (ii) the multivariate Student's t distribution with $\psi = \nu > 2$ degrees of freedom $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{T}(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$.[3] The Normal distribution is the usual choice in the spatial econometrics literature. However, when the data exhibits features that are against the Normal assumption, such as outliers or tail dependence among the spatial units, the assumption of conditional Normality becomes too restrictive. The multivariate Student's t distribution can be used in the above cases. Furthermore, the choice of $\mathcal{D}(\cdot)$ has important implications for our model as we will discuss later.

In Equation (1), $\mathsf{W}(\gamma_t, \mathbf{d})$ is a spatial weighting matrix at time t with the associated static spatial autoregressive coefficient ρ. The dynamic spatial weighting matrix $\mathsf{W}(\gamma_t, \mathbf{d})$ is an N–dimensional symmetric square matrix whose elements are defined as follows

Definition 3.1. $\mathsf{W}(\gamma_t, \mathbf{d}) = \mathsf{W}_t = \{\omega_{ij,t}\}_{i,j=1}^N$*:*
(a) $\omega_{ij,t} = \omega_{ji,t}, \quad \omega_{ij,t} > 0$, $\omega_{ii,t} = 0 \quad \forall i, j = 1, \ldots, N$*; (b)* $\mathbf{d} = (d_{ij}; i, j = 1, \ldots, N)$, $d_{ii} = 0, \quad \forall i = 1, \ldots, N$ *and* $d_{ij} = d_{ji}$, $d_{ij} > 0$ *for all* $i, j = 1, \ldots, N, \quad i \neq j$*; (c)* $\gamma_t > 0$.

Definition 3.1(a) ensures that all the elements of W_t are real positive entries with zeros on the main diagonal, whereas Definition 3.1(b) states that the element $\mathbf{d}$ is an N^2 vector of strictly exogenous non–stochastic variables representing a metric among the spatial units. In this paper, we do not allow for non–symmetric weighting matrices before normalization, and we concentrate our analysis on distance decay functions to define the weights. The element γ_t is the time–varying decay parameter measuring the relative importance of the higher–order neighborhoods in the spatial process at each point in time t. In Subsection 4.3 we will discuss the role of γ_t and its parameter space in detail.

Now let $\mathcal{F}_t = \sigma(\mathbf{y}_{t-s}, \mathbf{X}_{t-s+1}, s \geq 0)$ be the information set up to time t. We consider several parameterizations of W_t as a function of distances. For this purpose, we define the elements of W_t in the following way:

[2]Note that we allow for cross–sectional heteroscedasticity if $\sigma_i^2 \neq \sigma_j^2$ for at least one $i \neq j$, $i, j = 1, \ldots, N$ and cross–sectional homoscedasticity if $\sigma_i^2 = \sigma_j^2, \forall i, j$. Spatial units are often heterogeneous in important characteristics, e.g. the size, and for that reason it is important to consider a model that allows for the innovations to be heteroscedastic.

[3]The condition $\nu > 2$ follows from the parametrization we use for the multivariate Student's t distribution which is in terms of the covariance matrix $\boldsymbol{\Sigma}$.

Definition 3.2. *$\omega_{ij,t} = f(\gamma_t, d_{ij})$, where $f : \Re_+ \to \Re_+$ is an $\mathcal{F}_t$–measurable differentiable distance decay function among the spatial units i and j:*
(a) $f(\gamma_t, d_{ij}) > 0$ for $i \neq j$; (b) $f(\gamma_t, d_{ij}) = 0$ for $i = j$; (c)$f(\gamma_t, d_{ij}) < f(\gamma_t, d_{hk}) \iff d_{ij} > d_{hk}$; (d) $\lim_{d_{ij}\to\infty} f(\gamma_t, d_{ij}) = 0$, $\lim_{d_{ij}\to 0} f(\gamma_t, d_{ij}) = 1$, $\lim_{\gamma_t\to\infty} f(\gamma_t, d_{ij}) = 0$, $\lim_{\gamma_t\to 0} f(\gamma_t, d_{ij}) = 1$.

The choice of $f(\cdot,\cdot)$ implies a particular parametrization of W_t, and usually depends on the particular problem that applied econometricians face. The function $f(\cdot,\cdot)$ can be defined as any monotonically decreasing function that ensures lower weights as distances among spatial units increase. However, it is worth noting that, misspecification of $f(\cdot,\cdot)$ might result in biased estimates of the spatial autocorrelation coefficient, ρ, see for example Billé and Leorato (2020). Details on this parametrization are referred to Subsection 4.1.

Due to the simultaneous nature of spatial autoregressive processes, spatial models are typically specified in reduced forms. We require that:

Lemma 3.1. *Let $\lambda(\mathsf{W}_t)$ denote the spectral radius of the square N–dimensional W_t matrix at time t: $\lambda(\mathsf{W}_t) = max\{|e_{1t}|, ..., |e_{Nt}|\}$, where $e_{1t}, ..., e_{Nt}$ are the eigenvalues of W_t at time t. Then, $(\mathbb{I}_N - \rho\mathsf{W}_t)$ is non–singular for all the values of ρ in the interval $(-1/\lambda(\mathsf{W}_t), 1/\lambda(\mathsf{W}_t))$ at each time t.*

Assumption 1. *$\rho \in \left(-\frac{1}{\lambda(\mathsf{W}_t)}, \frac{1}{\lambda(\mathsf{W}_t)}\right)_{\setminus\{0\}}, \forall t$, where $\lambda(\mathsf{W}_t)$ is defined by Lemma 3.1.*

Assumption 1 ensures that the model in (1) can be uniquely defined by Lemma 3.1. Note that we explicitly exclude the case of $\rho = 0$ to model identifiability. Unfortunately, as will be discussed later, the ρ coefficient affects the evolution of γ_t, which prevents us from defining a priori the parameter space $\left(-\frac{1}{\lambda(\mathsf{W}_t)}, \frac{1}{\lambda(\mathsf{W}_t)}\right)_{\setminus\{0\}}, \forall t$. As is usually done in the spatial econometric literature to circumvent this problem, we consequently establish proper normalization rules of the symmetric N–dimensional weighting matrix W_t exploiting the following definition:

Definition 3.3. *$g(\mathsf{W}_t) = \mathsf{W}_t^*$, where $g(\cdot)$ is a differentiable, Lipschitz continuous, $\mathcal{F}_t$–measurable normalizing function such that $\lambda(\mathsf{W}_t^*) = 1, \ \forall t$.*

The $g(\cdot)$ function normalizes W_t through transformations of its elements $\omega_{ij,t}, i, j = 1, \ldots, N$.[4] Model (1) is then written as:

$$\mathbf{y}_t = \rho^* \mathsf{W}_t^* \mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{D}(\mathbf{0}, \boldsymbol{\Sigma}, \psi), \tag{2}$$

[4]We require that the same conditions for $\omega_{ij,t}$, such as their ordering with respect to d_{ij} and their limit behavior with respect to γ_t and d_{ij}, also hold for $\omega^*_{ij,t}$ under every choice of $g(\cdot)$.

where now $\rho^* \in (-1,1)_{\setminus\{0\}}$, since $\lambda(\mathsf{W}_t^*) = 1$ holds by construction at each t. We should note that ρ^* is always a function of the original ρ in model (1) apart from the normalizing factor/factors used to define the model in Equation 2. Under specific choices of $g(\cdot)$, model (1) can always be represented by considering a proper equivalent model with normalized weights. In a different context, the functional form of $g(\cdot)$ does not imply the equivalence between models (2) and (1). We detail these two scenarios and explain their consequences in Subsection 4.2.

Proper normalization rules, such as the ones mentioned above, ensure that the inverse matrix $(\mathbb{I}_N - \rho^* \mathsf{W}_t^*)^{-1}$ exists for all values of ρ^* in the interval $(-1,1)$. Therefore, model (2) can be expressed through its reduced form

$$\mathbf{y}_t = \mathbf{A}_t^{-1}\mathbf{X}_t\boldsymbol{\beta} + \mathbf{A}_t^{-1}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{D}(\mathbf{0}, \boldsymbol{\Sigma}, \psi) \tag{3}$$

with $\mathbf{A}_t = \mathbb{I}_N - \rho^* \mathsf{W}_t^*$, where $\mathbb{I}_N$ is the N–dimensional identity matrix. According to (3), the conditional distribution of $\mathbf{y}_t$ is

$$\mathbf{y}_t | \mathcal{F}_{t-1} \sim \mathcal{D}\left(\mathbf{y}_t; \widetilde{\boldsymbol{\mu}}_t, \widetilde{\boldsymbol{\Sigma}}_t, \psi\right), \tag{4}$$

where $\widetilde{\boldsymbol{\mu}}_t = \mathbf{A}_t^{-1}\mathbf{X}_t\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{\Sigma}}_t = \mathbf{A}_t^{-1}\boldsymbol{\Sigma}\mathbf{A}_t^{-1\prime}$.

It follows that the log likelihood contribution at time t of the observation $\mathbf{y}_t$ is proportional to

$$\log p(\mathbf{y}_t | \gamma_t, \boldsymbol{\eta}, \mathbf{X}_t) \propto -\frac{1}{2}\log|\boldsymbol{\Sigma}| + \log|\mathbf{A}_t| - (1/2)\,\mathbf{z}_t'\mathbf{z}_t, \tag{5}$$

with $\boldsymbol{\eta} = \left(\rho^*, \operatorname{diag}(\boldsymbol{\Sigma})', \boldsymbol{\beta}'\right)'$ for the Normal case, and

$$\begin{aligned} \log p(\mathbf{y}_t | \gamma_t, \boldsymbol{\eta}, \mathbf{X}_t) \propto & \log\Gamma\left(\frac{\nu+n}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2}\log(\nu-2) \\ & + \log|\mathbf{A}_t| - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \frac{\nu+n}{2}\log\left(1 + \frac{\mathbf{z}_t'\mathbf{z}_t}{\nu-2}\right), \end{aligned} \tag{6}$$

with $\boldsymbol{\eta} = \left(\rho^*, \operatorname{diag}(\boldsymbol{\Sigma})', \boldsymbol{\beta}', \nu\right)'$ for the Student's t case. Finally, the quantity $\mathbf{z}_t'\mathbf{z}_t$ is defined as

$$\mathbf{z}_t'\mathbf{z}_t = (\mathbf{A}_t\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta})'\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{A}_t\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}). \tag{7}$$

This modeling framework allows us to capture nonlinear dynamics of the spatial interactions of the variables of interest over time.

3.1. Filtering procedure for time–varying parameters

We let the time–varying distance decay parameter γ_t be updated through a filter based on the scaled score of the conditional density (4), exploiting the recent advantages of the fast growing

literature on Score Driven models, see Creal et al. (2013) and Harvey (2013). Specifically, let $\widetilde{\gamma}_t = \ln(\gamma_t)$, which excludes what Halleck Vega and Elhorst (2015) called the "perfect solution problem" (i.e. $\gamma_t = 0$), and let

$$\widetilde{\gamma}_{t+1} = (1-\xi)\,\kappa + \alpha \widetilde{s}_t + \xi \widetilde{\gamma}_t, \tag{8}$$

where κ, α and ξ are constants coefficients to be estimated, and

$$\widetilde{s}_t = \widetilde{\mathcal{I}}_t\left(\widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right)^{-1/2} \widetilde{\nabla}_t\left(\mathbf{y}_t, \widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right), \tag{9}$$

with

$$\widetilde{\nabla}_t\left(\mathbf{y}_t, \widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right) = \frac{\partial \log p\left(\mathbf{y}_t; \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_t\right)}{\partial \widetilde{\gamma}}\Big|_{\widetilde{\gamma}=\widetilde{\gamma}_t} \tag{10}$$

$$\widetilde{\mathcal{I}}_t\left(\widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right) = \mathbb{E}_{t-1}\left[\widetilde{\nabla}_t\left(\mathbf{y}_t, \widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right)^2\right], \tag{11}$$

where $\mathbb{E}_{t-1}$ represents the expectation with respect to the conditional distribution of the spatial units at time $t-1$. The quantities $\widetilde{\nabla}_t\left(\mathbf{y}_t, \widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right)$ and $\widetilde{\mathcal{I}}_t\left(\widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right)$ represent the score and the information quantity of the reparametrized spatial decay parameter $\widetilde{\gamma}_t$ with respect to the conditional distribution of $\mathbf{y}_t$ defined in Equation (4), respectively. These two quantities can be easily recovered starting from their analogous version evaluated with respect to the original decay parameter γ_t indicated as $\nabla_t\left(\mathbf{y}_t, \gamma_t, \mathbf{X}_t\right)$ and $\mathcal{I}_t\left(\gamma_t, \mathbf{X}_t\right)$. Indeed, the following relations hold:

$$\widetilde{\nabla}_t\left(\mathbf{y}_t, \gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right) = \frac{\partial \widetilde{\gamma}_t}{\partial \gamma_t} \nabla_t\left(\mathbf{y}_t, \gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right) \tag{12}$$

$$\widetilde{\mathcal{I}}_t\left(\widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right) = \left(\frac{\partial \widetilde{\gamma}_t}{\partial \gamma_t}\right)^2 \mathcal{I}_t\left(\gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right). \tag{13}$$

The quantities $\nabla_t\left(\mathbf{y}_t, \gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right)$ and $\mathcal{I}_t\left(\gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right)$ are defined in Appendix B for both the Normal and the Student's t cases.

Given a spatial sample of observations of length T, the model parameters can be easily estimated through Maximum Likelihood (ML), i.e.

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\max} \sum_{t=1}^{T} \log p\left(\mathbf{y}_t | \boldsymbol{\theta}, \mathbf{X}_t\right), \tag{14}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\eta}', \kappa, \alpha, \xi\right)'$, and $\log p\left(\mathbf{y}_t | \boldsymbol{\theta}, \mathbf{X}_t\right)$ is the log likelihood contribution of observation $\mathbf{y}_t$ defined in Equations (5) and (6) for the Normal and Student's t cases, respectively. In (14), $\boldsymbol{\Theta}$ represents a compact set where the model parameters take values, see Harvey (2013) and Blasques et al. (2014) for ML estimation of Score Driven models.

4. Different parameterizations of W_t

Given that choosing the true spatial weight matrix is surely an unrealistic goal, part of the relevant literature is now looking on quantifying the estimation bias to a particular weighting matrix selection, and checking for robustness of the parameters of interest. Estimating W_t throughout a proper parametrization is the solution adopted in this paper. In the following two subsections we specifically describe the way we define monotonically decreasing functions of distances and normalizing functions for the equivalence of the models.

4.1. Distance decay functions

In line with Halleck Vega and Elhorst (2015), we parametrize the weighting matrix with a generic distance decay function, i.e. $f(\gamma_t, d_{ij})$, in order to allow for its time–varying version with a limited number of parameters to be filtered, i.e. γ_t. We consider two alternatives, namely: (i) inverse distance, (ii) negative exponential. In the first case we have

$$f(\gamma_t, d_{ij}) = \frac{1}{d_{ij}^{\gamma_t}}, \tag{15}$$

which implies that the intensity of the relationships among pairs of units in space is inversely proportional to the unknown time–varying γ_t–power of their distances. Similarly, the negative exponential function is given by

$$f(\gamma_t, d_{ij}) = \exp(-\gamma_t d_{ij}), \tag{16}$$

where, again, the higher the estimate of the time–varying parameter γ_t, the lower the role played by the spatial units at greater distances (i.e. higher–order neighborhoods). It follows that, in both cases, γ_t can be interpreted as a time–varying distance decay parameter, and, for small estimated values, this is an indication that higher–order neighbors have to be recognized for better describing spatial dependencies.

4.2. Normalizing functions

In this section we explain two different choices of the function $g(\cdot)$ starting from definition 3.3, namely: (i) spectral normalization, (ii) row–normalization. The spectral–normalisation rule proposed by Kelejian and Prucha (2010) for a static spatial autoregressive model with autoregressive and

heteroscedastic disturbances, i.e. SARAR(1,1), is such that $\mathsf{W}^* = \mathsf{W}/\lambda(\mathsf{W})$ and $\rho^* = \rho \times \lambda(\mathsf{W})$, with $\lambda(\cdot)$ defined by Lemma 3.1. Under this parameterization, the function $g(\mathsf{W}_t)$ takes the form

$$g(\mathsf{W}_t) = \frac{1}{\lambda(\mathsf{W}_t)}\mathsf{W}_t, \tag{17}$$

which ensures that $\lambda(\mathsf{W}_t^*) = 1, \ \forall t$. It follows that model (1) has the representation given in (2).[5] However, the spatial interaction coefficient ρ^*, which corresponds to the spectral–normalized weights matrix, will in general depend on the spatial dimension due to the fact that the normalizing factor depends on the spatial dimension as well.

Alternatively, but with a different interpretation of the spatial weighting matrix, one can consider the well-known row–normalization of W_t. The row–normalization rule is such that $\rho^* \in (-1, 1)$, but the resulting model is no more equivalent to the original one. Consequently, unless theoretical issues suggest a row–normalized weight matrix, this approach will in general lead to a misspecified model[6]. Under this parameterization, the function $g(\cdot)$ takes the form $g(\mathsf{W}_t) = \mathsf{W}_t^* = \{\omega_{ij,t}^*\}_{i,j=1}^N$, where

$$\omega_{ij,t}^* = \frac{\omega_{ij,t}}{\sum_j \omega_{ij,t}}, \quad \forall t, \tag{18}$$

which ensures an upper bound of the parameter space for ρ^* equal to 1 and suggests a useful interpretation of the interactions between spatial units as a weighted average of their neighbors.

In this paper we consider the spectral–normalization rule for W_t as in Equation (17), preserving also the interpretation of the distance decay functions in absolute rather than in relative terms.

4.3. The role of γ_t

In a purely spatial parametric framework, i.e. with $\gamma_t = \gamma$ and no time information, a nonlinear procedure to estimate this parameter has been proposed by Fischer et al. (2009). Rather than estimate the parameter γ, Kostov (2010) specified a "plausible" range of values of $\gamma \in [0.4, 4]$ with increments of 0.1, leading to a large discrete number of possible candidate weighting matrices to be evaluated through a grid search procedure, for which the Bayesian model choice method of Hepple

[5]Note, however, that the parameter γ, and its dynamic version γ_t, depends on the measurement unit used for geographical distances when row-normalization of the weighting matrix is not allowed for.

[6]The row–normalization rule can only ensure the equivalence between the two model specifications in (1) and (2) with pre–specified spatial weighting matrices if a k–nearest neighbors (i.e. k–nn) approach is used to define the order of proximity of each spatial unit.

(2004) can be used. In the so-called spatially lagged–X (SLX) model, if the estimate $\hat{\gamma}$ is reasonably small, then this can be interpreted as an indication that interactions may continue even over a first or second–order neighborhood, suggesting that a first–order contiguity matrix or a k–nn approach[7] may be not an appropriate way to represent the true spatial dependence in that case (Halleck Vega and Elhorst, 2015). Estimating a static γ is then useful to select an appropriate weighting scheme, i.e. between *sparse* and *dense* W matrices, to avoid as much as possible model misspecifications implied by wrongly assumed spatial processes to the extent possible.

The reason why estimation of a distance decay parameter γ is useful is that we are generally interested in knowing the degree at which the spatial dependence between units diminishes as distance increases without imposing a priori spatial structures. In this context the goal should be to estimate the optimal level at which the correlation among spatial (cross–sectional) units rapidly decreases over time. When adding time information, a time–varying γ_t furthermore provides information on spatial degree variation over time, showing the evolution of the "unknown spatial radius" over which interactions become smaller. The meaning of dynamic spatial autocorrelation structures is that, generally speaking, interactions among spatial units in the realm of economics (e.g. economic agents, country–specific variables) but also in environmental and geophysical sciences (e.g. air pollution data), may change simultaneously over space and time. However, the radius within which we have to consider higher–order neighborhoods, especially if we consider regional data, or the significant part of the spatial correlation structure is generally unknown to the researcher.

Understanding the evolution of γ_t is also useful to check spatial model misspecification problems over time, due to the typically necessary condition of pre–specifying weighting schemes. This aim can be addressed simply by observing that in periods with reasonably low γ_t (i.e. $\gamma_t \to 0$), the commonly used first–order spatial weighting matrices do not adequately capture the intensity of the true spatial autoregressive process, with the result that spatial spillover effects may go further. In other words, periods with reasonably low values of γ_t favor the use of dense spatial weighting matrices, and vice versa, as a more appropriate weighting scheme to represent spatial autocorrelation processes for such periods.

The role of γ_t is to highlight the degree of importance of the spatial connections through the spatial

[7]Note that a k–nn approach can produce a weighting matrix similar to a first–order contiguity matrix as long as $k << N$, such that the resulting spatial weighting matrix is reasonably sparse as a first–order neighbourhood criterion. The only difference is the constant number k of neighbors for each spatial unit in this case.

weighting matrices in different periods of time. However, γ_t itself does not provide an immediately applicable index of connections between spatial units since its domain is unbounded. Here we show a simple index defined in the range $(0, 1)$ starting from the sequence of filtered correlation matrices $\{R_t, t > 0\}$, where $R_t = R_t(\gamma_t)$ is a function of γ_t. Specifically, we consider the highest eigenvalue of the filtered correlation matrix at time t as a function of γ_t, $\lambda(R_t(\gamma_t))$, and compare it with the case of no spatial association $\lambda(R_t(\infty))$. Thus we can define our measure of association as:

$$\varpi_t = 1 - \frac{\lambda(R_t(\infty))}{\lambda(R_t(\gamma_t))} \tag{19}$$

such that $\varpi_t \in (0, 1)$. Clearly, when γ_t diverges, the spatial units become uncorrelated and $\lambda(R_t(\infty)) = 1$. As detailed in Subsection 4.3, smaller values of γ_t indicates stronger connections among the spatial units, and thus higher values of ϖ_t. When $\varpi_t \to 1$, a dense weighting matrix is more appropriate to model the spatial association of the data, whereas when $\varpi_t \to 0$, a sparse weighting matrix is sufficient to properly account for the spatial spillover effects. This can be shown in different periods of time.

5. Statistical properties of the model

For notational simplicity, let us define the derivative of the score as

$$\widetilde{s}'_t(\widetilde{\gamma}) := \frac{\partial \widetilde{\mathcal{I}}_t\left(\widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_t\right)^{-1/2}}{\partial \widetilde{\gamma}} \times \frac{\partial \widetilde{\nabla}_t\left(\mathbf{y}_t, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_t\right)}{\partial \widetilde{\gamma}} \ ,$$

and the initial value for the filtered score as

$$\widetilde{s}_1(\widetilde{\gamma}) := \widetilde{\mathcal{I}}_1\left(\widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right)^{-1/2} \widetilde{\nabla}_1\left(\mathbf{y}_1, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right) \ ,$$

for some fixed initial value $\widetilde{\gamma} \in \mathbb{R}$ of the filter, where

$$\widetilde{\nabla}_1\left(\mathbf{y}_1, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right) = \frac{\partial \log p\left(\mathbf{y}_1; \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right)}{\partial \widetilde{\gamma}} \quad \text{and} \quad \widetilde{\mathcal{I}}_1\left(\widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right) = \mathbb{E}_0\left[\widetilde{\nabla}_t\left(\mathbf{y}_1, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right)^2\right].$$

Additionally, let us define the score process as written in terms of innovations $\boldsymbol{\varepsilon}_t$, after substitution of $\mathbf{y}_t$ for $\mathbf{A}_t^{-1}\mathbf{X}_t\boldsymbol{\beta} + \mathbf{A}_t^{-1}\boldsymbol{\varepsilon}_t$,

$$\widetilde{\nabla}_t^{\boldsymbol{\varepsilon}}\left(\boldsymbol{\varepsilon}_t, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_t\right) := \widetilde{\nabla}_t\left(\mathbf{A}_t^{-1}\mathbf{X}_t\boldsymbol{\beta} + \mathbf{A}_t^{-1}\boldsymbol{\varepsilon}_t, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_t\right) \ ,$$

as well as the derivatives of the data generating score process,

$$\widetilde{s}_t^{\boldsymbol{\varepsilon}\prime}(\widetilde{\gamma}) := \frac{\partial \widetilde{\mathcal{I}}_t\left(\widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_t\right)^{-1/2}}{\partial \widetilde{\gamma}} \times \frac{\partial \widetilde{\nabla}_t^{\boldsymbol{\varepsilon}}\left(\boldsymbol{\varepsilon}_t, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_t\right)}{\partial \widetilde{\gamma}} \ ,$$

and

$$\widetilde{s}_1^{\boldsymbol{\varepsilon}}(\widetilde{\gamma}) := \widetilde{\mathcal{I}}_1\left(\widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right)^{-1/2} \widetilde{\nabla}_1^{\boldsymbol{\varepsilon}}\left(\boldsymbol{\varepsilon}_1, \widetilde{\gamma}, \boldsymbol{\eta}, \mathbf{X}_1\right) \ ,$$

respectively.

The following proposition establishes the strict stationarity, ergodicity and bounded moments of the data sequence $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ generated by the dynamic spatial model in Equation (1). Additionally, it highlights that the sequence of normalized time–varying spatial weight matrices $\{\mathsf{W}_t^*\}_{t\in\mathbb{Z}}$ is also strictly stationary and ergodic (SE). The proof is obtained mainly as an application of Theorem 3.1 in Bougerol (1993) and the c_n–inequality in Dufour (2013).

Proposition 5.1. (Stationarity, Ergodicity and Moments) *Suppose that*

(i) The time–varying parameter is initialized with a logarithmic moment $\mathbb{E}\log^+|\widetilde{s}_1^{\boldsymbol{\varepsilon}}| < \infty$ *for some fixed* $\tilde{\gamma}$*;*

(ii) The recursion of the time-varying parameter is contracting: $\mathbb{E}\log\sup_{\tilde{\gamma}}|\alpha\widetilde{s}_t^{\boldsymbol{\varepsilon}\prime} + \xi| < 1$*;*

(iii) The spatial weights $w_{ij,t}$ *are defined as* $w_{ij,t} = f(\gamma_t, d_{ij})$ *where* f *is a measurable function satisfying* $0 \leq f(\gamma, d) \leq 1 \ \forall \ (\gamma, d)$*.*

(iv) The sequence of vectors of exogenous regressors $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ *is strictly stationary and ergodic with* n *bounded moments* $\mathbb{E}|\mathbf{X}_t|^n < \infty$*;*[8]

(v) The sequence of vectors of innovations $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ *is iid with* n *bounded moments* $\mathbb{E}|\boldsymbol{\varepsilon}_t|^n < \infty$*.*

Then the sequence of normalized spatial weighting matrices $\{\mathsf{W}_t^*\}_{t\in\mathbb{Z}}$ *is strictly stationary and ergodic, and it satisfies* $\mathbb{E}|w_{ij,t}^*|^m < \infty \ \forall \ m > 0$*. Furthermore, the data sequence* $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ *generated according the dynamic spatial model in Equation* (1) *is strictly stationary and ergodic, and it satisfies* $\mathbb{E}\|\mathbf{y}_t\|^n < \infty$*.*

Let $\{\hat{\gamma}_t\}_{t\in\mathbb{N}}$ be the filtered sequence initialized at some fixed $\tilde{\gamma} \in \mathbb{R}$, at time $t = 1$, and satisfying the recurrence equation,

$$\hat{\gamma}_{t+1} = (1 - \xi) + \alpha\hat{s}_t(\hat{\gamma}_t) + \xi\hat{\gamma}_t \ , \quad \text{for every} \ \ t \in \mathbb{N}.$$

Furthermore, let $\hat{\mathsf{W}}_t^*$ be the resulting normalized spatial weights matrix $\hat{\mathsf{W}}_t^* = g(\hat{\mathsf{W}}_t)$, whose elements are $\hat{w}_{ij,t}^*$ are measurable transformations of the weights $\hat{w}_{ij,t} = f(\hat{\gamma}_t, d_{ij})$. Below, we show that $\hat{\mathsf{W}}_t^*$

[8] For a random vector $\mathbf{Z}_t$ we let the bounded moment condition $\mathbb{E}|\mathbf{Z}_t|^n < \infty$ hold for all univariate elements of $\mathbf{Z}_t$. Additionally, the exogeneity condition means that the sequence of regressors $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ is independent of the sequence of innovations $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$.

converges to a unique strictly stationary and ergodic limit sequence with moments of arbitrary order. Unlike Proposition 5.1, which uses contraction condition that holds only at the true parameter, in Proposition 5.2 the contraction condition used for invertibility holds uniformly over the parameter space. Overall, the proof of invertibility follows steps that are similar to those of Proposition 1 in Blasques et al. (2016b) and Proposition 3.1 in Blasques et al. (2018), but it is written for a multivariate model with a univariate time–varying parameter, and it features added conditions to accommodate for exogenous variables.

Proposition 5.2. (Filter Invertibility) *Suppose that*

1. *The filter is initialized with a logarithmic moment* $\mathbb{E}\log^{+}\sup_{\boldsymbol{\theta}\in\Theta}|\widetilde{s}_1|<\infty$ *for some fixed* $\tilde{\gamma}$*;*
2. *The spatial weights* $w_{ij,t}$ *are defined as* $w_{ij,t}=f(\gamma_t,d_{ij})$*, where* f *is a measurable function satisfying* $\sup_{(\gamma,d)}\frac{\partial f(\gamma,d)}{\partial\gamma}<\infty$ *and* $0\leq f(\gamma,d)\leq 1\ \forall\ (\gamma,d)$*.*
3. *The filter is contracting:* $\mathbb{E}\log\sup_{\tilde{\gamma}}\sup_{\boldsymbol{\theta}\in\Theta}|\alpha\widetilde{s}_t'+\xi|<1$*;*
4. *The sequence of vectors* $\{(\mathbf{y}_t,\mathbf{X}_t)\}_{t\in\mathbb{Z}}$ *is strictly stationary and ergodic with* n *bounded moments* $\mathbb{E}|\mathbf{X}_t|^n<\infty$ *and* $\mathbb{E}|\mathbf{y}_t|^n<\infty$*.*

Then the sequence of normalized the filtered spatial weights $\{\hat{w}^*_{ij,t}\}_{t\in\mathbb{N}}$ *converge e.a.s. to a unique strictly stationary and ergodic sequence* $\{w^*_{ij,t}\}_{t\in\mathbb{Z}}$ *uniformly over the parameter space* Θ *and spatial units* (i,j)*,*

$$\iota^t\sup_{\boldsymbol{\theta}\in\Theta}\sup_{ij}|\hat{w}^*_{ij,t}-w^*_{ij,t}|\overset{as}{\to}0\quad as\quad t\to\infty\quad\ \ for\ some\ \iota>1,$$

where the limit sequence of the spatial weights satisfies $\sup_{\boldsymbol{\theta}\in\Theta}|w^*_{ij,t}|<\infty\ \forall\ (i,j)$*. This implies naturally that*

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\widehat{\mathsf{W}}^*_t-\mathsf{W}^*_t\|\overset{eas}{\to}0\quad as\quad t\to\infty\quad and\quad\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\mathsf{W}^*_t\|^n<\infty\ \forall\ n>0.$$

The invertibility result ensures that any error in the initialization of the filter is asymptotically negligible. This is a crucial element for the filter to be meaningful and reliable. Additionally, this proposition is useful in deriving asymptotic properties for the MLE as it helps us establish the asymptotic stationarity and bounded moments for the log likelihood function, which ultimately allow us to use laws of large numbers and central limit theorems.

5.1. Asymptotic Properties of the MLE

Below we establish the consistency and asymptotic normality of the MLE for the parameters of our dynamic spatial model. The MLE is defined as

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{T}\sum_{t=2}^{T} \log p(\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{X}_t, \hat{\mathsf{W}}_t) \ ,$$

where $\hat{\mathsf{W}}_t$ denotes the filtered spatial weights matrix whose elements are $\hat{w}_{ijt} = f(\hat{\gamma}_t, d_{ij})$ with $\hat{\gamma}_t$ initialized at a fixed $\hat{\gamma}_1 \in \mathbb{R}$. The consistency results are provided for both Gaussian and Student's t innovations $\boldsymbol{\varepsilon}_t$. The conditions for consistency differ in the number of bounded moments that are required for the data. The proof builds on the M–estimation theory (Theorem 3.4 in White (1994), or Theorem 3.3 in Gallant and White (1988)), and closely follows the results in Blasques et al. (2014). Again, the results differ from those in Blasques et al. (2014) and Blasques et al. (2018) as they apply to multivariate models with exogenous variables. They differ from those of the spatial model in Andree et al. (2017) as here we must deal with a filtered time-varying parameter. Furthermore, in contrast to Blasques et al. (2016b), our results apply to a model with a time-varying spatial weights matrix $\hat{\mathsf{W}}_t$. Additionally, while Blasques et al. (2016b) omit any proof for their general Theorem 1, here we provide more explicit conditions and offer proofs of stationarity for the DGP, invertibility for the filter and consistency and asymptotic normality for the MLE.

Showing identification can be challenging in complex nonlinear dynamic models. For this reason, following Potscher and Prucha (1997), we consider the set consistency of our MLE. Let $\Theta_0 \subseteq \Theta$ denote a set of observationally equivalent parameters which maximize the limit log likelihood function $\mathbb{E}\log p(\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{X}_t, \hat{\mathsf{W}}_t)$. Propositions 5.3 and 5.4 show that the MLE $\hat{\boldsymbol{\theta}}_T$ is set consistent to Θ_0 as $T \to \infty$, under a standard set-distance ρ_Θ. The results are based on Lemma 4.2 of Potscher and Prucha (1997) which holds without identification as long as the *level sets* are *regular*. Given the compactness of the parameter space Θ, the regularity of level sets is ensured here as the limit log likelihood is continuous (by uniform convergence and continuity of the sample log likelihood); see Definition 4.1 in Potscher and Prucha (1997) for additional details. In Corollary 5.1 we further note that consistency to a singleton $\boldsymbol{\theta}_0 \in \Theta$ is obtained when the parameter is identified; i.e. if $\boldsymbol{\theta}_0$ defines a unique probability measure that is distinct from any other parameter $\boldsymbol{\theta} \in \Theta$ such that $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

Proposition 5.3. *(MLE Consistency: Gaussian Model) Let the parameter space Θ be compact and such that the conditions of Proposition 1 and 2 hold with $n = 2$, $\forall\ \boldsymbol{\theta} \in \Theta$, Σ is positive definite*

$\forall\ \boldsymbol{\theta} \in \Theta$. *Then the MLE is set consistent* $\rho_\Theta(\hat{\theta}_T, \Theta_0) \xrightarrow{as} 0$ *as* $T \to \infty$.

Proposition 5.4. *(MLE Consistency: Student's t Model) Let the parameter space* Θ *be compact and such that the conditions of Proposition 1 and 2 hold with* $n > 0$ *uniformly for* $\boldsymbol{\theta} \in \Theta$*, both* $\nu > 2$ *and* Σ *is positive definite* $\forall\ \boldsymbol{\theta} \in \Theta$. *Then the MLE is set consistent* $\rho_\Theta(\hat{\theta}_T, \Theta_0) \xrightarrow{as} 0$ *as* $T \to \infty$.

Corollary 5.1. *Let the conditions of either Proposition 5.3 or 5.4 hold. Suppose further that the true parameter vector* $\boldsymbol{\theta}_0 \in \Theta$ *is identified. Then* $\hat{\boldsymbol{\theta}}_T \xrightarrow{as} \boldsymbol{\theta}_0 \in \Theta$.

Below, we let the first and second derivatives of the filtered spatial weights matrix $\widehat{\mathsf{W}}_t$ be denoted by $\dot{\widehat{\mathsf{W}}}_t := \partial \widehat{\mathsf{W}}_t / \partial \boldsymbol{\theta}$ and $\ddot{\widehat{\mathsf{W}}}_t := \partial^2 \widehat{\mathsf{W}}_t / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$. Additionally, we let $\{\mathsf{W}_t\}_{t\in\mathbb{Z}}$ and $\{\ddot{\mathsf{W}}_t\}_{t\in\mathbb{Z}}$ be the limit SE sequences. Finally, we let $\ell_t(\boldsymbol{\theta}) := \log p(\mathbf{y}_t | \boldsymbol{\theta}, \mathbf{X}_t, \mathsf{W}_t)$, and denote its first a second derivative w.r.t. $\boldsymbol{\theta}$ by $\ell_t'(\boldsymbol{\theta})$ and $\ell_t''(\boldsymbol{\theta})$, respectively. For the criterion depending on the filtered spatial weights we use the notation $\hat{\ell}_t(\boldsymbol{\theta}) := \log p(\mathbf{y}_t | \boldsymbol{\theta}, \mathbf{X}_t, \hat{\mathsf{W}}_t)$, and similarly, for the derivatives $\hat{\ell}_t'(\boldsymbol{\theta})$ and $\hat{\ell}_t''(\boldsymbol{\theta})$. Please note also that $\hat{\ell}_t'(\boldsymbol{\theta})$ is a function of $(\widehat{\mathsf{W}}_t, \dot{\widehat{\mathsf{W}}}_t)$ and $\hat{\ell}_t''(\boldsymbol{\theta})$ is a function of $(\widehat{\mathsf{W}}_t, \dot{\widehat{\mathsf{W}}}_t, \ddot{\widehat{\mathsf{W}}}_t)$.

Proposition 5.5. *(MLE Asymptotic Normality) Let the conditions of either Proposition 5.3 or 5.4 hold. Furthermore, let the parameter space* Θ *be such that (a)* $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$*; (b) the score* $\ell_t'(\boldsymbol{\theta}_0)$ *has two bounded moments; (c) the Hessian* $\ell_t'(\boldsymbol{\theta}_0)$ *has two bounded moments; and (d) the filter derivatives of the spatial weights matrix satisfy*

$$\|\dot{\widehat{\mathsf{W}}}_t(\boldsymbol{\theta}_0) - \dot{\mathsf{W}}_t(\boldsymbol{\theta}_0)\| \xrightarrow{eas} 0 \quad \textit{and} \quad \sup_{\boldsymbol{\theta}\in\Theta} \|\ddot{\widehat{\mathsf{W}}}_t - \ddot{\mathsf{W}}_t\| \xrightarrow{eas} 0 \quad \textit{as} \quad t \to \infty.$$

Finally, assume that the limit Hessian $\mathbb{E}\ell_t''(\boldsymbol{\theta})$ *is non-singular. Then,* $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{V})$ *as* $T \to \infty$.

6. Simulation study

In this section we report two simulation studies to investigate the finite sample properties of the maximum likelihood estimator (MLE) for the Dynamic Spatial Weighting Matrix (DSWM) model as well as the filtering ability of the proposed score updating mechanism for the spatial decay parameter γ_t. For all cases we employ the negative exponential scheme reported in (16).

T	ρ	α	β	κ	σ	ρ	α	β	κ	σ
	DGP1					DGP2				
500	−0.6996	0.0199	0.9615	0.6844	0.9998	−0.3004	0.0226	0.8804	0.6927	1.0001
	(*0.0046*)	(*0.0024*)	(*0.0184*)	(*0.1055*)	(*0.0058*)	(*0.0095*)	(*0.0213*)	(*0.2384*)	(*0.0863*)	(*0.0063*)
1000	−0.7000	0.0199	0.9660	0.6957	0.9997	−0.3000	0.0223	0.9320	0.6943	1.0001
	(*0.0034*)	(*0.0017*)	(*0.0101*)	(*0.0760*)	(*0.0040*)	(*0.0064*)	(*0.0125*)	(*0.1218*)	(*0.0604*)	(*0.0040*)
2000	−0.6999	0.0200	0.9675	0.6988	1.0001	−0.3001	0.0206	0.9568	0.6964	1.0003
	(*0.0025*)	(*0.0012*)	(*0.0071*)	(*0.0596*)	(*0.0022*)	(*0.0047*)	(*0.0091*)	(*0.0577*)	(*0.0426*)	(*0.0026*)
	DGP3					DGP4				
500	0.3004	0.0232	0.8914	0.6911	1.0002	0.7000	0.0198	0.9613	0.6940	1.0000
	(*0.0087*)	(*0.0200*)	(*0.1985*)	(*0.0967*)	(*0.0059*)	(*0.0044*)	(*0.0024*)	(*0.0195*)	(*0.1087*)	(*0.0057*)
1000	0.3002	0.0222	0.9318	0.6949	0.9999	0.6999	0.0197	0.9662	0.6914	0.9998
	(*0.0061*)	(*0.0133*)	(*0.1231*)	(*0.0543*)	(*0.0040*)	(*0.0031*)	(*0.0017*)	(*0.0107*)	(*0.0777*)	(*0.0039*)
2000	0.2999	0.0208	0.9554	0.6948	0.9999	0.6998	0.0199	0.9675	0.6985	1.0000
	(*0.0043*)	(*0.0089*)	(*0.0556*)	(*0.0407*)	(*0.0026*)	(*0.0022*)	(*0.0012*)	(*0.0068*)	(*0.0578*)	(*0.0027*)

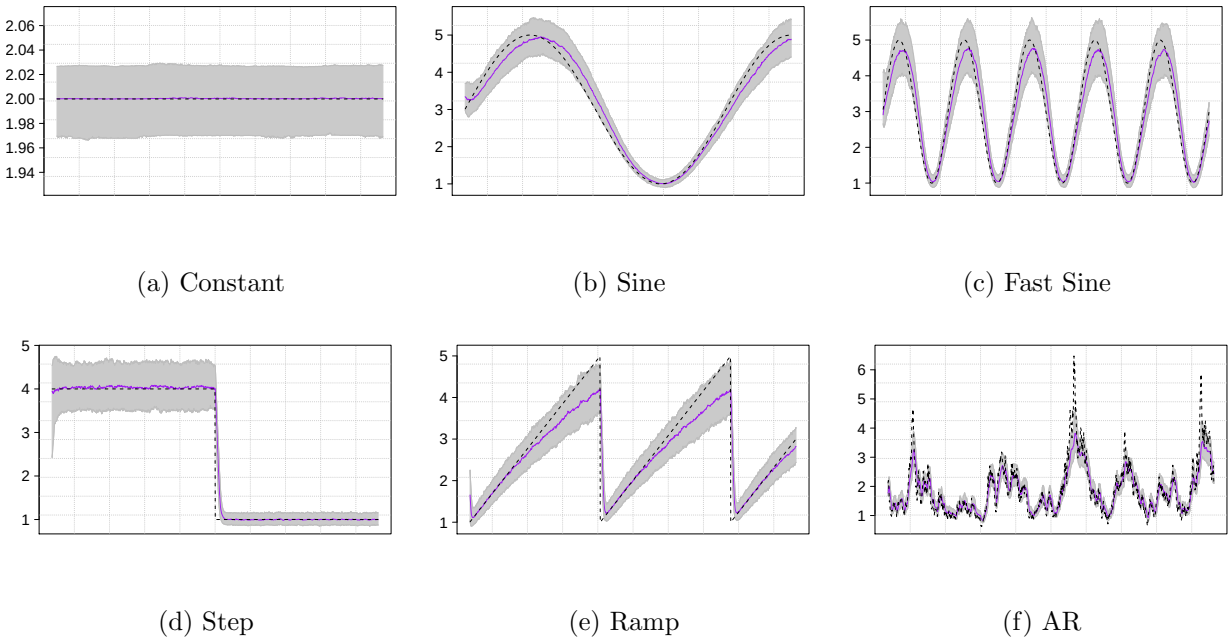
Table 1: Simulation results for the ML estimator for $N = 50$. The table reports the average estimates across replications along with their root mean squared error in brackets. True parameters are set to $\alpha = 0.02$, $\beta = 0.97$, $\kappa = \log(2) \approx 0.6931$, and $\sigma = 1$. True values for ρ are: $\rho = -0.7$ for DGP1, $\rho = -0.3$ for DGP2, $\rho = 0.3$ for DGP3, and $\rho = 0.7$ for DGP4.

6.1. Finite sample properties of the MLE

In order to investigate the finite sample properties of the MLE for the DSWM model we simulate $B = 500$ series of pseudo observations from the true model defined in Equations (4) and (8). We consider different values of the sample size T and the dependence parameter ρ in model (2), and subsequently we estimate the DSWM model on the simulated data. The number of spatial units is equal to $N = 50$, and the distance vector $\mathbf{d}$ is simulated in a way to ensure a plausible spatial interpretation.[9] The considered sample sizes are $T = 500$, $T = 1000$, $T = 2000$. The coefficients for the γ_t recursion are fixed to $\alpha = 0.02$, $\beta = 0.97$, $\kappa = \log 2 \approx 0.6931$. The homoscedastic variance is fixed to $\Sigma = \sigma \mathbb{I}_N$ with $\sigma = 1.0$, whereas we consider four values for ρ: the cases with strong and moderate negative (SN and MN, respectively) dependence ($\rho = -0.7$, $\rho = -0.3$), and the cases of strong and moderate positive (SP and MP, respectively) dependence ($\rho = 0.7$, $\rho = 0.3$). Table 1 reports the average estimates across replicates along with their root mean squared error.

We find that the MLE in finite samples reports remarkably good results, even when the sample

[9] Results for $N = 100$ are qualitatively the same and are reported in the supplementary material along with additional simulation results.



(a) Constant (b) Sine (c) Fast Sine

(d) Step (e) Ramp (f) AR

Figure 1: Filtering properties of the DSWM model of Equation 1 for six artificial patterns of γ_t. Black dashed lines represent the true value of γ_t, purple solid lines represent the median value of γ_t computed over $B = 1000$ estimates. Gray bands indicate the $(10\% - 90\%)$ quantiles evaluated at each point in time using the 1000 estimates.

size is relatively small. Furthermore, the estimated coefficients are unbiased with decreasing root mean squared error when the sample size increases. It is worth remarking that these results hold for different sample sizes and levels of spatial dependence. Additional simulations indicate that changing N leads to qualitatively analogous results.

6.2. Filtering properties of the model

We now investigate the flexibility of the score updating mechanism assumed for the γ_t recursion. Similar to the experiment reported by Engle (2002) in a different context, we assume several artificial deterministic and stochastic patterns for the distance decay parameter γ_t, and we study the ability of the proposed model to recover the true values. Specifically we assume six different patterns: *i*) Constant, $\gamma_t = 2$, *ii*) Sine, $\gamma_t = 3 + 2\sin(2\pi t/800)$, *iii*) Fast Sine, $\gamma_t = 3 + 2\sin(2\pi t/200)$, *iv*) Step, $\gamma_t = 4 - 3\,(t > 500)$, *v*) Ramp, $\gamma_t = \mod(t/400)/100 + 1$, and *vi*) AR, $\gamma_t = \exp(\widetilde{\gamma}_t)$, where $\widetilde{\gamma}_t = 0.015 + 0.98\widetilde{\gamma}_{t-1} + 0.1\eta_t, \quad \eta_t \sim \mathcal{N}(0, 1)$.

The considered artificial patterns for γ_t allow for breaks as well as for a slow and fast evolution of the process. To assess the flexibility of the DSWM model, for each of the considered artificial

dynamics, we simulate $B = 500$ series of length $T = 1000$ using a spatial dimension of $N = 25$. The other parameters of the model are fixed to $\sigma = 1$ and $\rho = 0.7$ and are not estimated for this experiment. Figure 1 reports the median across the B estimates at each point in time $t = 1, \ldots, T$ for the six artificial γ_t processes along with the empirical $(5\% - 95\%)$ confidence bands. We can observe that our DSWM model displays very good filtering ability for all the considered artificial dynamics.

7. Empirical applications

In this section we report two different empirical applications. The first one is related to the evolution in the European perceived sovereign credit risk from 28 July 2008 until 2 July 2018, whereas the second application is related to the analysis of the spatio–temporal dynamics of real house prices within the UK economy during the period 1974–2016. The two applications are described in greater detail in Subsections 7.1 and 7.2, respectively. For both applications, we consider a series of specifications nested in DSWM model defined after setting constraints on the distance decay parameter γ_t, the heteroscedastic disturbances, and the way in which the $\mathsf{W}(\gamma_t, \mathbf{d})$ matrix is parameterized and normalized.

Let us first consider the DSWM model of Equation (2) with the extension of time–varying shock volatilities detailed in Section 3.1 and labelled as "(γ_t, Σ_t)". We can now obtain a class of dynamic/static spatial–nested models according to the type of constraints that we set:

1. DSWM model of Equation (2) labelled as "(γ_t, Σ)": if $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ for all $t = 1, \ldots, T$, which is achieved by setting $\alpha_\sigma = \xi_\sigma = 0$ in Equation (A.2).
2. Static–γ model of Halleck Vega and Elhorst (2015) labelled as "(γ, Σ)": if $\gamma_t = \gamma$ and $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ for all $t = 1, \ldots, T$, which is achieved by setting $\alpha = \xi = 0$ and $\alpha_\sigma = \xi_\sigma = 0$ in Equations (2) and (A.2), respectively.

The above models can be further differentiated if we consider the two ways of parametrization of $\mathsf{W}(\gamma_t)$ as in Subsection 4.1, the type of normalization rule as in Subsection 4.2, and the type of error distribution, i.e. Gaussian and Student's t. In the following two empirical applications we found that the models with the negative exponential parametrization and the spectral–normalization are to be preferred. By combining the three specifications (γ_t, Σ_t), (γ_t, Σ), and (γ, Σ) with the two distributional assumptions we obtain a total of six model specifications. In the following, we report our results based on this configuration.

A further degree of flexibility is given by the specification of the variance of the cross–sectional innovations. Catania and Billé (2017) discuss the different modeling approaches in terms of time and cross homo/heterosccedasticity in details. In our context, the model specification detailed in (2) exhibits time homoscedasticity since the variances of the innovations are constant, and cross heteroscedasticity since the elements in the diagonal matrix $\mathbf{\Sigma}$ in Equation (2) are different among them. The model extension detailed in Section 3.1 allows for both time and cross heteroscedasticity.

Unreported analysis indicated that cross homoscedasticity is more suitable for our data sets and thus results are presented using this specification. Cross homoscedasticity implies that $\Sigma = \sigma\mathbb{I}$ for the model specification of Equation (2) and $\kappa_{\sigma,i} = \kappa_{\sigma,j} = \kappa_\sigma$ for all i, j for the model extension discussed in Section 3.1.[10]

7.1. Association between European countries' perceived risk

In this analysis, we evaluate the evolution of perceived sovereign credit risk over a period that includes the Eurozone sovereign debt crisis. Sovereign credit spreads in Europe have been recently analyzed by important contributions in a spatial context, see e.g. Eder and Keiler (2015) and Blasques et al. (2016b). Eder and Keiler (2015) model the contagion risk amongst financial institutions by using credit default swap (CDS) spreads, which reflect the probability of default of the underlying reference entity. Their aim is to address, through a static spatial econometric model, the question of how the CDS spread of a financial institution i depends on the CDS spreads of all other institutions within the financial system. An extension of the static spatial model to time–varying spillover effects is given in Blasques et al. (2016b). Their time–varying spatial coefficient provides a measure of changes in systemic risk and the market's perception of contagion within the euro area over time.

In our analysis, we study the weekly logarithmic changes in percentage of the CDS of eight European countries. A similar data set is used by Blasques et al. (2016b) for a different period of time. Differently from Blasques et al. (2016b), our measure of distance is based on the spearman correlation coefficient among the series of debt to GDP ratio of the European countries downloaded from OECD (2018) as in Catania and Billé (2017). Specifically, we define $d_{ij} = \sqrt{2\left(1 - \rho^s_{ij}\right)}$ as the

[10]Note that, cross homoscedasticity for the model extension of Section 3.1 implies that the long run level of the volatility of the innovations is the same, but it does not imply that the conditional volatilities are equal at particular points in time.

metric among pairs of spatial units, where ρ^s_{ij} is the spearman correlation coefficient among country i and j.

7.1.1. Data

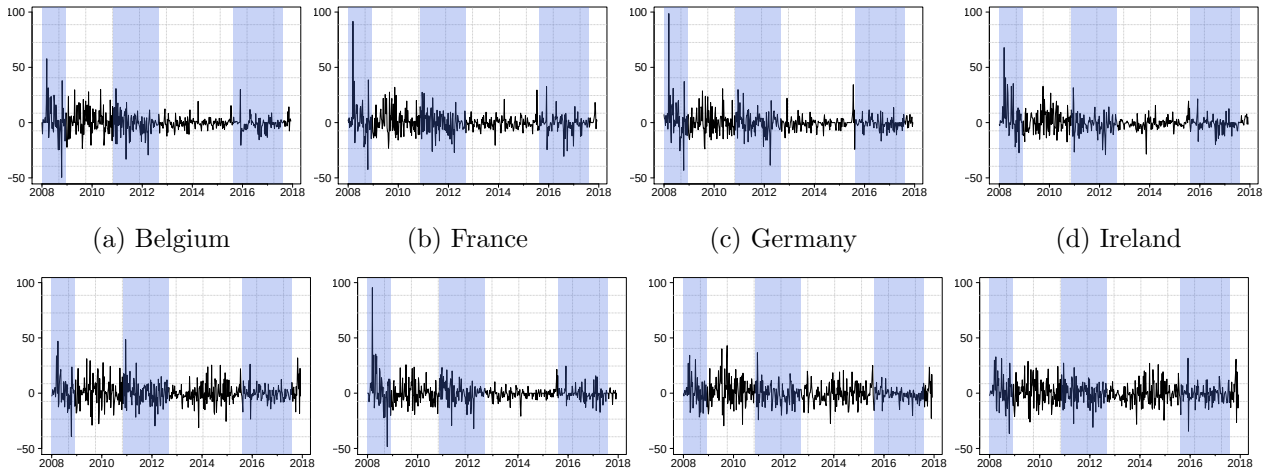
Data are obtained from Datastream and are related to the credit default swap (CDS) spreads from 28 July 2008 until 2 July 2018 (519 weekly observations) for the eight European countries: Belgium, France, Germany, Ireland, Italy, the Netherlands, Portugal, and Spain. Figure 2 shows the logarithmic changes in percentage of the credit default swap (CDS) spreads for the above mentioned European countries. Descriptive statistics are reported in the supplementary material. The time series reveal common patters such as volatility clustering and the presence of extreme values. Overall, Italy, Spain, and Portugal seems to be more volatile. As stressed by Blasques et al. (2016b), the evolution of the Ireland credit spread is roughly in line with that of the other countries before mid 2010 and after mid 2012, but departing during the height of the European sovereign debt crisis. This motivates Blasques et al. (2016b) to account for time–varying spillover effects among CDS spreads, whereas in our paper we aim at identifying the degree of these spatial associations over time. We consider the explanatory variables as in Blasques et al. (2016b), obtained from Datastream.

In particular, we use a constant, the lagged values of the CDS changes, and the change in the volatility index (VStoxx). In addition, we use the country–specific price equity indices listed in Table 1 of their paper, which are the result of (log) returns of the main stock index and the absolute changes in the interest rate spreads between government bonds with one year and ten years maturities. As they stressed, "*local stock market returns are a measure of the well–being of the local economy and in this way an indirect measure of the ability of governments to pay off debt in the long run through tax collection*". All the variables are included with a lag of one period.

7.1.2. Main results

Table 2 reports the estimated coefficients for all the nested specification detailed at the begin of this section. We note that, all the estimated coefficients are statistically significant and most of them at 1%. The static spatial autocorrelation coefficient ρ^* is positive and about 0.7 for all the considered static/dynamic model specifications, which means that the spillover effects play a crucial role in European CDS spreads.

The coefficients associated with the explanatory variables are statistically and economically



(a) Belgium (b) France (c) Germany (d) Ireland

(e) Italy (f) the Netherlands (g) Portugal (h) Spain

Figure 2: Weekly percentage logarithmic changes of the credit default swap (CDS) spread in yields for eight European countries. The series span from 28 July 2008 to 2 July 2018 for a total of 519 observations. Data are obtained from Datastream. The blue shaded bars indicate periods of European recession according to the OECD based Recession Indicators for Euro Area from the Period following the Peak through the Trough (EURORECD), see Federal Reserve Bank of St. Louis (2018a).

significant. Furthermore, they are robust with respect to the different model specifications we consider. The coefficient associated with the implied volatility of the stocks (β_2) is negative and significantly different from zero at the usual confidence levels, suggesting that when the stocks' volatility increases, the CDS spreads decrease. As stressed by Blasques et al. (2016b) this is consistent with the phenomenon of "flight to quality" from stocks to bonds when the price of risk increases in stock markets. In the same way, the coefficient associated with the countries' equity index (β_3) is negative and statistically different from zero, suggesting that local stock market upturns have a dampening effect on sovereign credit spreads. Notably, β_3 is much higher (in absolute value) than β_2, indicating that rather than looking at the stock market implied volatility, investors are pricing the ability of country's specific firms to generate positive future cash flows that translate into higher GDP levels and thus lower Debt-to-GPD ratios. Looking at the coefficients associated with γ_t, we see that the results change between the Gaussian and Student's t specifications. This result is not surprising since, as detailed in Section 3, the distributional assumption has important implications for the filter. Indeed, while the intercept (κ) and the autoregressive coefficient (ξ) are somehow similar across all the specifications, the score coefficient α is between 0.09 and 0.06 in the Gaussian case and

	Gaussian			Student's t		
	$(\gamma \quad \Sigma)$	$(\gamma_t \quad \Sigma)$	$(\gamma_t \quad \Sigma_t)$	$(\gamma \quad \Sigma)$	$(\gamma_t \quad \Sigma)$	$(\gamma_t \quad \Sigma_t)$
ρ^*	0.70 (0.0004)	0.70 (0.0004)	0.70 (0.0004)	0.68 (0.0005)	0.68 (0.0005)	0.68 (0.0005)
β_1	−0.02 (0.0036)	−0.02 (0.0033)	−0.05 (0.0034)	−0.09 (0.0028)	−0.09 (0.0029)	−0.09 (0.0027)
$\beta_2 \times 100$	−0.34 (0.0037)	−0.34 (0.0037)	−0.16 (0.0035)	−0.02 (0.0031)	−0.02 (0.0024)	−0.03 (0.0016)
β_3	−0.47 (0.0015)	−0.47 (0.0016)	−0.44 (0.0015)	−0.36 (0.0014)	−0.36 (0.0014)	−0.36 (0.0012)
$\beta_4 \times 100$	−0.01 (0.0040)	−1.60 (0.0041)	−0.01 (0.0047)	0.71 (0.0038)	0.71 (0.0038)	0.62 (0.0037)
κ	−0.37 (0.0058)	−0.42 (0.0085)	−0.12 (0.0049)	−0.47 (0.0076)	−0.62 (0.0158)	−0.47 (0.0143)
α		0.09 (0.0010)	0.06 (0.0017)		0.43 (0.0222)	0.43 (0.0093)
ξ		0.90 (0.0031)	0.83 (0.0037)		0.89 (0.0053)	0.91 (0.0013)
κ_σ	3.34 (0.0011)	3.35 (0.0011)	3.29 (0.0023)	3.72 (0.0062)	3.72 (0.0062)	3.72 (0.0065)
α_σ			5.00 (0.0004)			0.16 (0.0275)
ξ_σ			0.99 (0.0001)			0.92 (0.0206)
ν				3.05 (0.0105)	3.05 (0.0108)	3.05 (0.0113)

Table 2: Estimated coefficients for European credit default swap (CDS) spreads for different model specifications. Standard deviations based on the assumed asymptotic Gaussian distribution are reported in parentheses. The coefficient β_1 is associated to the intercept, β_2 to the change in the volatility index (VStoxx), β_3 to country–specific price equity indices, and β_4 to the lagged values of the CDS changes.

0.43 in the Student's t case. Since the score innovation is a unit variance martingale difference in both cases, this result suggests that the signal delivered by the Student's t model is more informative than that delivered using a Gaussian distribution. Notably, when we look at those specifications with time heteroscedastic errors ($\gamma_t, \quad \Sigma_t$), the estimates associated with the Student's t model are more reasonable than those delivered by the Gaussian model. This result is not surprising since, as in usual volatility score driven models, the filter implied by the Gaussian assumption is not robust to extreme observations, which leads to unsatisfactory filtered conditional volatilities.

Table 3 reports the Akaike (AIC) and Bayesian (BIC) information criteria for all the model

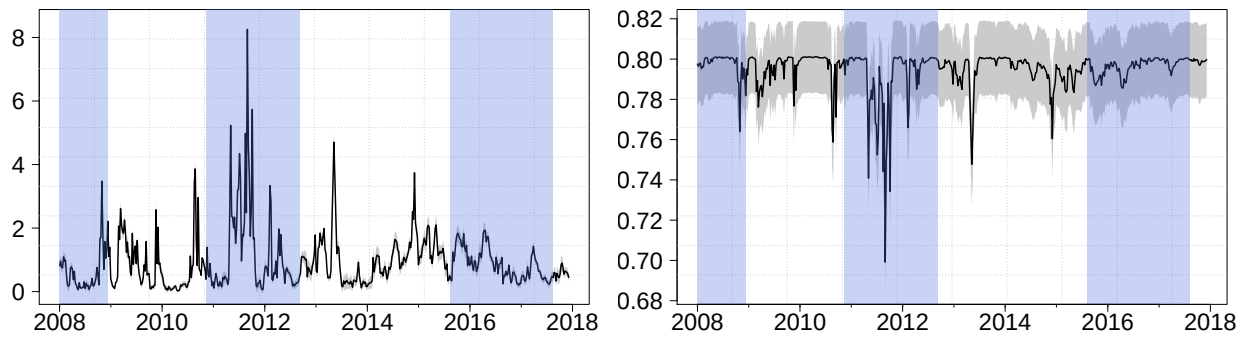
	Gaussian			Student's t		
	AIC	BIC	NP	AIC	BIC	NP
$(\gamma \quad \Sigma)$	26340.91	26400.44	14	25498.21	25561.98	15
$(\gamma_t \quad \Sigma)$	26335.04	26403.07	16	25487.34	25559.63	17
$(\gamma_t \quad \Sigma_t)$	26096.41	26172.95	18	25497.60	25578.38	19

Table 3: This table reports the AIC and BIC evaluated using the likelihood computed at its maximum value for different models using CDS data. The last column labelled "NP" reports the number of estimated parameters. The gray cells indicate the selected model according to AIC and BIC.

specifications. We find that models with Student's t distributed errors are generally preferred according to both information criteria. Looking at the different dynamic specifications, we find that the models with time–varying spatial dependence and homoscedastic disturbances, $(\gamma_t, \quad \Sigma)$, report lower AIC and BIC, suggesting that the flexibility induced by the dynamic specification is justified from a likelihood perspective. The model with heteroscedastic disturbances, $(\gamma_t, \quad \Sigma_t)$, is not selected in the Student's t case, while it is preferred in the Gaussian case. Overall, the model selected by AIC and BIC is the $(\gamma_t, \quad \Sigma)$ specification with Student's t distributed errors. The following results are reported according to this specification.

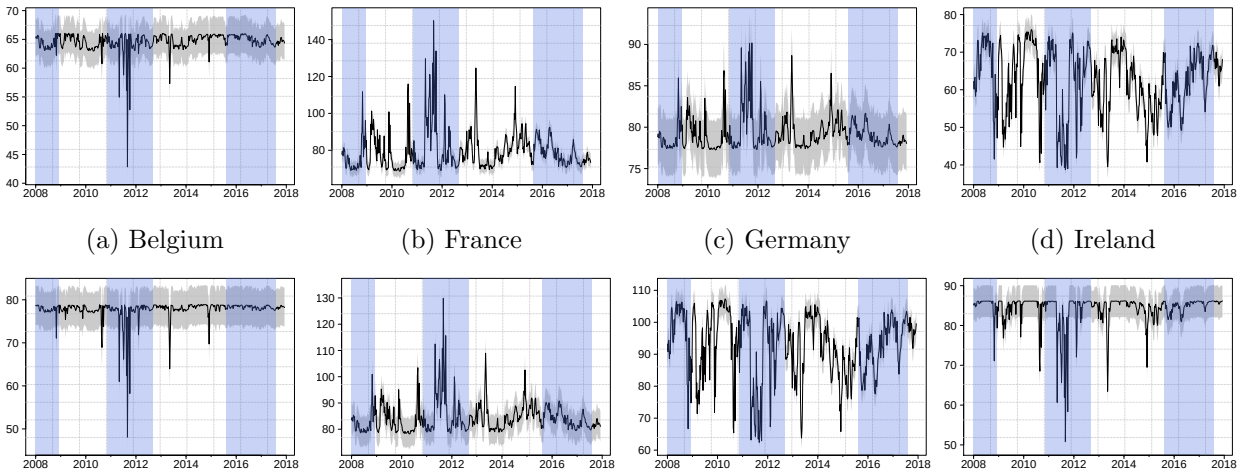
Figures 3a and 3b show the evolution of γ_t and the spatial indicator ϖ_t detailed in Section 4.3, respectively. We find that the level of γ_t ranges between approximately 0.5 and 4.0 over the sample period, while the spatial association index ranges between 0.75 and 0.8. These results indicate a relatively strong spatial connection between the perceived risk of European countries during the whole sample. Interestingly, we find that during the peak of the European sovereign debt crisis of 2008-2012 the spatial connection between the European countries diminished and reached its minimum of 75%, nonetheless still quite high.

Looking at the filtered conditional variances of the CDS changes in Figure 4 (the diagonal elements of $\widetilde{\Sigma}_t$ in Equation (4)), we note that the CDS variance reaction to changes in the spatial connections is heterogeneous among European countries. For example, the countries that are paying more, in terms of increasing uncertainty around their CDS changes, are France, Germany, and the Netherlands. On the contrary, Ireland, Italy, Portugal, and Spain experience a reduction in their conditional variance. This result suggests that European countries react differently to different levels of spatial



(a) γ_t (b) Spatial Association Index

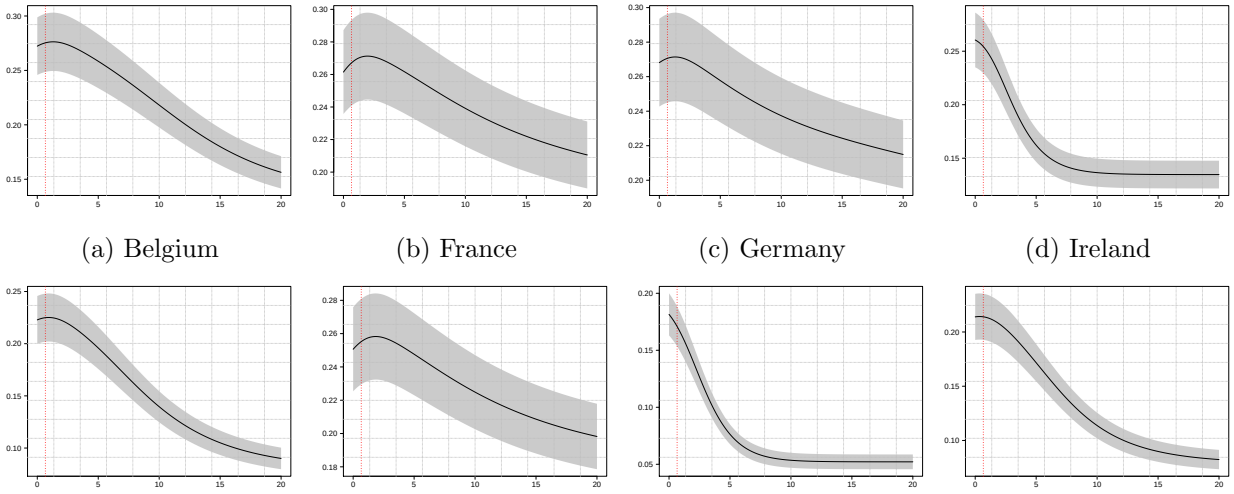
Figure 3: Filtered γ_t (panel a) and indicator of spatial association as in Equation (19) (panel b) for CDS spreads and $(10\% - 90\%)$ in–sample simulation–based confidence bands (gray bands) computed as in Blasques et al. (2016a). The estimation period spans from 28 July 2008 to 2 July 2018. The blue shaded bars indicate periods of European recession according to the OECD based Recession Indicators for Euro Area from the Period following the Peak through the Trough (EURORECD), see Federal Reserve Bank of St. Louis (2018a).



(a) Belgium (b) France (c) Germany (d) Ireland

(e) Italy (f) The Netherlands (g) Portugal (h) Spain

Figure 4: Credit default swap (CDS) spread filtered conditional variance for eight European countries and $(10\% - 90\%)$ in–sample simulation–based confidence bands (gray bands) computed as in Blasques et al. (2016a). The estimation period spans from 28 July 2008 to 2 July 2018. The blue shaded bars indicate periods of European recession according to the OECD based Recession Indicators for Euro Area from the Period following the Peak through the Trough (EURORECD), see Federal Reserve Bank of St. Louis (2018a).



(a) Belgium (b) France (c) Germany (d) Ireland

(e) Italy (f) The Netherlands (g) Portugal (h) Spain

Figure 5: Estimated long run objective function, $\ell_{i,t}$, for the individual European countries as a function of γ_t. Uncertainty is represented by the $(10\% - 90\%)$ in–sample simulation–based confidence bands (gray bands) computed as in Blasques et al. (2016a). The orange dashed vertical bar indicates the estimated long run value of γ_t.

dependence. Generally, the European countries want to reduce their CDS spreads and also control for the uncertainty around their evolution. According to this reasoning we might assume that the European country i at time t wants to maximize the following objective function:

$$\ell_{i,t} = -\frac{\widetilde{\mu}_{i,t}}{\widetilde{\sigma}_{i,t}}, \tag{20}$$

where $\widetilde{\mu}_{i,t}$ and $\widetilde{\sigma}_{i,t}$ are the mean and variance of $Y_{i,t}|\mathcal{F}_{t-1}$. Maximization of $\ell_{i,t}$ is achieved for decreasing values of $\widetilde{\mu}_{i,t}$ and $\widetilde{\sigma}_{i,t}$. Clearly, $\ell_{i,t} = \ell_{i,t}(\gamma_t, \mathbf{d})$ for all i, such that the objective function of each European country is affected by the level of spatial connection and the distance from other countries. While γ_t is endogenous and cannot be directly controlled by a single country, the "distance" with other countries, $\mathbf{d}$, can be modified by policy interventions.

In Figure 5 we report the long run level of $\ell_{i,t}$ for different values of γ_t for all the European countries. Interestingly, we find that, for high values of γ_t (low spatial connection) the objective function is lower than for low values of γ_t (high spatial connection) indicating that spatial connection among European countries increases each individual objective function. However, we note that the optimum value of γ_t is different among European countries. Indeed, while Ireland, Portugal, and Spain want to maximize their connection with other European countries ($\gamma_t \to 0$), Belgium ($\gamma_t = 1.27$), France ($\gamma_t = 1.92$), Germany ($\gamma_t = 1.26$), Italy ($\gamma_t = 0.92$), and the Netherlands ($\gamma_t = 1.80$) find

their optimal level of connection for values of γ_t in the range $(0.9, 2)$. At the end of the estimation period the value of γ_t is approximately 0.47, which indicates that, along with Ireland and Portugal, also Spain would benefit from a decrease in the level of connection with other countries in the future. Overall, our findings suggest that, at the end of the estimation period, economically strong countries like Germany, the Netherlands, and France have their perceived risk increased due to their spatial connection with economically weaker countries like Ireland, Portugal, and Spain. Finally, while it is reasonable that governments benefit from the maximization of equation (20), they also might pursue other maximization strategies which would lead to different results in terms of optimal spatial association.

7.2. House price dynamics in the UK

In this section we analyze the evolution of house price dynamics in UK. The spatio–temporal analysis of housing markets has recently received a growing attention in the literature, see Shi and Lee (2017b), Billé et al. (2017), Bailey et al. (2016), Brady (2011) and Holly et al. (2010). The aim of this section is to analyze, through the evolution of a distance decay parameter γ_t, the radius at which the spatial effects tend to diminish rapidly. In this way, we aim at identifying periods of time in which the evaluated time–series are more inter–connected, suggesting a common behavior among them. Periods of reasonably lower γ_t, i.e. $\gamma_t \to 0$, can guide the researcher in properly selecting the most appropriate weighting scheme between a *sparse* and a *dense* matrix in order to avoid, in a cross—sectional or panel data framework, potential model misspecification due to wrongly assumed weight matrices.

Although geographical locations/regions are time–invariant, the strength of spatial dependence may also depend on economic variables that are time–varying. The use of economic information to define a metric as an alternative to geographical distances has been accounted for by some authors, see e.g. Holly et al. (2011). In this paper, we consider geographical distances based on Euclidean distances between centroids of the regions (i.e. areal statistical units of interest).[11] Data are available from the UK Office of National Statistics website.

[11] In an unreported analysis we also employed distances based on an economic metric using information coming from the regional Gross Value Added (GVA) indicators. Although the level of the estimated γ_t turns out to be different from the case reported in this paper, its evolution over time looks very similar and thus does not change our results.

7.2.1. Data

Data are quality adjusted regional house price series available at the Nationwide Building Society website[12] and cover the period from Q1 1974 to Q2 2018. These quality adjusted house prices are referred to as "mixed adjusted", i.e. corrected for price variations due to location and physical characteristics of the housing stock. The definition of regions used by the Nationwide differs in significant ways from the regional definitions used by the Office of National Statistics that are based on the Nomenclature of Territorial Units for Statistics (NUTS) of the European Union. Starting from NUTS data, Nationwide regions are defined by aggregating NUTS data following a pre–specified neighbor criterion (see Holly et al., 2011, Tables 1–2). The Nationwide regions consist of 12 regions: East Anglia (EA), Outer South East (OSE), East Midlands (EM), Scotland (S), London (L), South West (SW), North (N), Wales (W), North West (NW), West Midlands (WM), Outer Metropolitan (OM), Yorkshire & Humberside (YH). We exclude Northern Ireland from our analysis.

[12] http://www.nationwide.co.uk/about/house-price-index/download-data

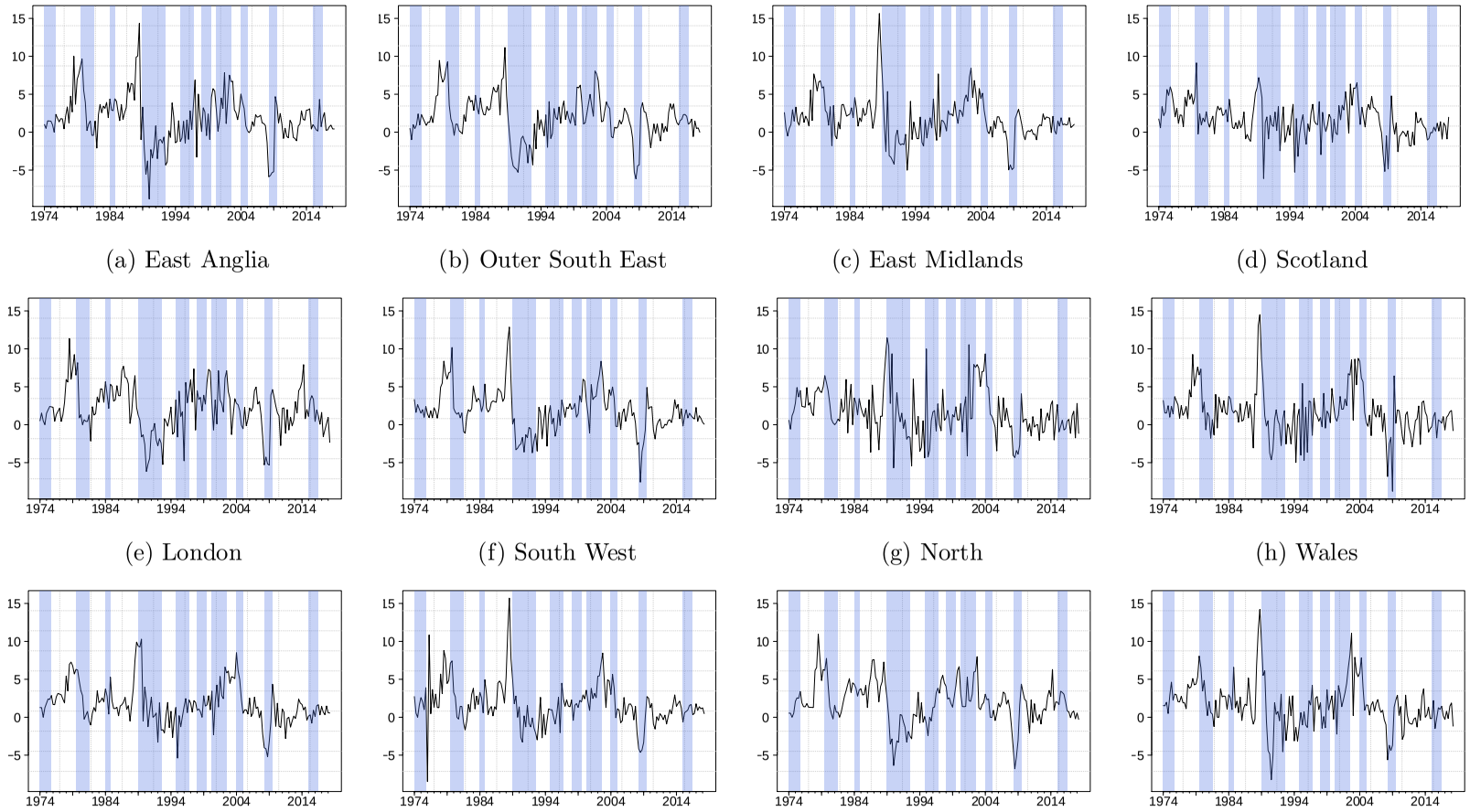


Figure 6: Quarterly percentage logarithmic changes of the United Kingdom house prices computed in different regions. Series span from Q1 1974 to Q2 2018 for a total of 178 observations. Data are obtained from the UK Nationwide Building Society at `https://www.nationwide.co.uk/`. Blue shaded bars indicate periods of UK recession according to the OECD based Recession Indicators for the United Kingdom from the Peak through the Trough (GBRRECDM), see Federal Reserve Bank of St. Louis (2018b).

	Gaussian			Student's t		
	AIC	BIC	NP	AIC	BIC	NP
$(\gamma \quad \Sigma)$	8619.03	8654.03	11	8263.70	8301.88	12
$(\gamma_t \quad \Sigma)$	8577.42	8618.79	13	8237.10	8281.65	14
$(\gamma_t \quad \Sigma_t)$	8228.49	8276.22	15	8115.52	8201.43	27

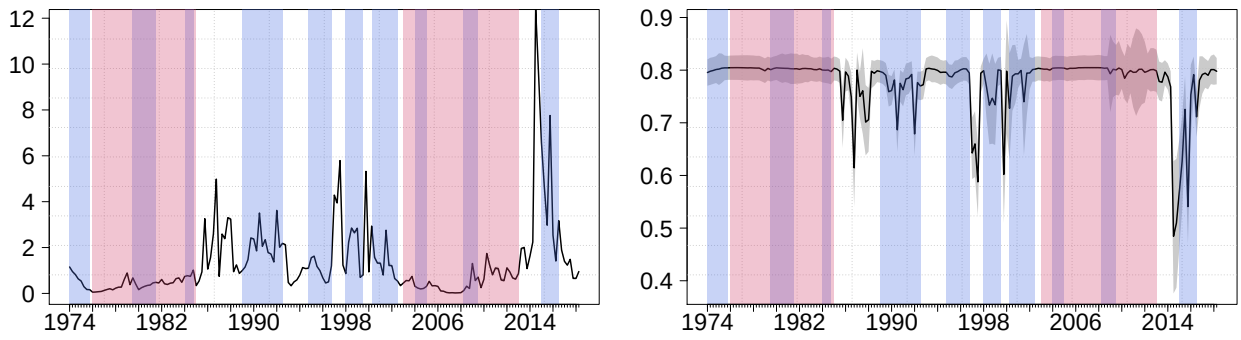
Table 4: This table reports the AIC and BIC evaluated using the likelihood computed at its maximum value for different models using the UK house price changes. The last column labelled "NP" reports the number of estimated parameters. Gray cells indicate the selected model according to AIC and BIC.

Figure 6 reports the house price series for each region defined by the Nationwide in the UK. Descriptive statistics are reported in the supplementary material. The house price time series reveal the highest peak close to the year 1989, just before the recession experienced by UK in early 1990s, with the exception of London, the Outer Metropolitan Area, and Scotland, and an important slump in the period 1990-1992. Most of the series also shows another significant house price slump during the 2007-2008 financial crisis called "the Great Recession". To explain UK house price dynamics, we use the series of explanatory variables as in Brady (2011). In particular, we consider the lag of house prices, the unemployment index, the industrial production index, the population level, and the interest rate. All variables are considered in their first difference. The unemployment rate, industrial production, and population level are provided by the Office for National Statistics, whereas the interest rate stems from the website of the Federal Reserve Bank of St. Louis.

7.2.2. Main results

Starting from the nationwide definition of regions, we parametrize our dynamic spatial weighting matrix W_t by defining the centroids and calculating Euclidean distances between them, following the parametrization function in Equation (16). Table 4 shows the results related to different nested–model specifications. The dynamic version with negative exponential function, spectral–normalization, and Student's t distribution is to be preferred.

The evolution of γ_t is reported in Figure 7a. The unconditional mean is 1.323. In particular, we can observe two decades of reasonably low values of γ_t: (i) 1976–1985 and (ii) 2004–2012. During both these periods, the evolution reveals an important role played by the spatial process. Specifically,



(a) γ_t (b) Spatial Association Index

Figure 7: Filtered γ_t (panel a) and indicator of spatial association as in Equation (19) (panel b) for UK house price dynamics and $(10\% - 90\%)$ in–sample simulation–based confidence bands (gray bands) computed as in Blasques et al. (2016a). The estimated period spans from Q1 1974 to Q2 2018 for a total of 178 observations. The blue shaded bars indicate periods of UK recession according to the OECD based Recession Indicators for the United Kingdom from the Peak through the Trough (GBRRECDM), see Federal Reserve Bank of St. Louis (2018b).

the spatial impact does not rapidly decrease after a first–order neighborhood, so that the house price time–series related to more distant regions appear to be highly inter–connected. Therefore, the evolution of house prices in the majority of the nationwide regions of the UK reveals a common behavior during the above–mentioned periods.

The estimated ρ^* coefficient is approximately 0.63 with a standard deviation of 0.01, indicating that the spatial process behavior is also not inhibitory.[13] From a statistical point of view, the evolution of γ_t in those periods highlights a potential model misspecification problem for researchers who make use of static cross–sectional spatial autoregressive models with sparse or first–order weighting matrices. A dense weighting matrix, defined by any smooth function, can be a more appropriate choice instead. Figure 7b shows the evolution of the indicator of spatial association ϖ_t in Equation (19) for UK house prices. According to the evolution of γ_t, we identify two periods (in red), i.e. (a) 1976 – 1985 and (b) 2004 – 2012, in which $\varpi_t \to 1$. In these two cases the spatial effects go beyond a first–order neighborhood, and a dense matrix is presumably the correct one.

Figure 8 displays the higher–order effects of the spatial process for 4 different periods of time as

[13] We do not report all estimated coefficients since the main goal of this analysis is to show periods when a dense/sparse matrix is more appropriate for UK house prices.

(a) Period 1976-1985 (b) Period 1986-2003 (c) Period 2004-2012 (d) Period 2013-2016

Figure 8: Higher–order effects of W_t with respect to London for 4 different periods of time. The intra–periods average values of γ_t are: (a) $\gamma_1 = 0.415$, (b) $\gamma_2 = 1.734$, (c) $\gamma_3 = 0.492$, (d) $\gamma_4 = 4.054$. The spatial units (i.e. regions) are defined according to the Office of National Statistics.

suggested by the evolution of γ_t in Figure 7a. These effects are evaluated with respect to London, which has been indicated by Holly et al. (2011) to be the dominant region in the house price markets in UK. Periods are: 1976 – 1985 and 2004 – 2012, when γ_t is smaller than 1 and equal to 0.42 and 0.49, respectively, and 1986 – 2003 and 2013 – 2016, when γ_t is larger than 1 and equal to 1.73 and 4.05, respectively. By looking at the figure, we observe that the magnitude of London's dominance over the other regions as described by Holly et al. (2011) is remarkably different across time. Indeed, during the periods 1976 – 1985 and 2004 – 2012 house prices in London have important spatial effects on the house prices of other regions, whereas during the periods 1986 – 2003 and 2013 – 2016 the spatial effects are much smaller.

8. Conclusion

In this paper we propose a new flexible spatio–temporal dynamic model named Dynamic Spatial Weighting Matrix (DSWM). We account for a time–varying spatial weighting matrix as well as time–varying error heteroscedasticity. We allow the time–varying model parameters to be updated using the scaled score of the spatial conditional distribution, relying on the recently proposed SD updating mechanism, see e.g. Creal et al. (2013) and Harvey (2013). Our specification generalizes the static SAR(1) model allowing for a time–varying distance decay parameter γ_t, i.e. by considering a time-varying spatial weighting matrix W_t. We also consider different parametrization and normalization of W_t as well as both Gaussian and Student's t distributions, defining several nested model specifications. The model is a novel contribution to the recent spatial literature that considers time–varying weight

matrices. We detail the model characteristics and assess the finite-sample properties of the maximum likelihood estimator for the DSWM model. The flexibility of the proposed model is also investigated in a simulation study. Specifically, we found that the DSWM model has very good filtering abilities and is able to replicate many artificial patterns for the spatial decay parameter. For instance, we found that our model is able to adequately approximate a SAR model with distance decay parameter following a first–order autoregressive process.

The paper also contributes from an empirical perspective. In this respect, we illustrate the usefulness of the DSWM model for two different empirical applications related to: (i) European credit default swap (CDS) spreads, (ii) UK house price dynamics. We find that the dynamic specification has to be preferred in both empirical applications. Specifically as regards the CDS spreads we find that there is a strong spatial connection between the risk perceived by European countries. Moreover, the CDS variance reaction to changes in the spatial connections is quite heterogeneous among the European countries: depending on different values of γ_t, some countries always benefit from their spatial connection while other countries do not. Pertaining to UK house price dynamics, we find an interesting evolution of the γ_t parameter. We identify two periods in which all the series appear to be highly inter–connected, i.e. the evolution of house prices in the regions of nationwide UK reveals a common behavior over a first–order neighborhood. Hence, the evolution of γ_t could furthermore avoid potential model misspecification due to a wrongly assumed weighting matrices.

Straightforward extensions of our DSWM model at the cost of additional calculation could include more time–varying spatial weighting matrices, spatially lagged regressors leading to a time–varying spatial Durbin specification of the model, or autocorrelated shocks leading to a time–varying spatial autoregressive model with autoregressive (and eventually heteroscedastic) disturbances, as well as individual and/or time fixed effects to account for specific form of unobserved heterogeneity.

Appendix A. Including time varying shock volatilities

The way γ_t enters the conditional variance of $\mathbf{y}_t|\mathcal{F}_{t-1}$ implies that spatial units exhibit heteroscedasticity over time. However, the implied form of heteroscedasticity can be quite limited and does not reflect the characteristics of the data. An even more problematic issue is that if the true data generating process incorporates heteroscedasticity in the individual shocks, $\boldsymbol{\varepsilon}_t$, the resulting filtered values of γ_t will be affected by this form of model misspecification leading to wrong conclusions about the dependence structure of the spatial units. To prevent this scenario, if there is evidence of time–varying shock volatility effects, the model presented in Section 3 can be extended as follows:

$$\varepsilon_t|\mathcal{F}_{t-1} \sim \mathcal{D}(\mathbf{0}, \boldsymbol{\Sigma}_t, \psi), \tag{A.1}$$

where $\boldsymbol{\Sigma}_t = \text{diag}\left(\sigma^2_{i,t}; i = 1, \ldots, n\right)$. Time–variation in the $\sigma^2_{i,t}$ is introduced relying on the score driven methodology as for γ_t and employing an exponential link function

$$\sigma^2_{i,t} = \exp(f_{i,t}) \tag{A.2}$$

$$f_{i,t} = (1 - \xi_\sigma)\kappa_{\sigma,i} + \alpha_\sigma u_{i,t-1} + \xi_\sigma f_{i,t-1}, \tag{A.3}$$

where $u_{i,t}$ is the scaled score of the conditional distribution $\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}$ evaluated with respect to $f_{i,t}$. Formulas to compute $u_{i,t}$ for the Gaussian and Student's t cases are reported in Appendix B. This solution only requires estimation of two additional parameters: α_σ and ξ_σ. A similar approach has been followed by Blasques et al. (2016b) and Catania and Billé (2017).

Appendix B. Score and information quantity of $\boldsymbol{\gamma_t}$

For the purpose of this appendix let us write the conditional distribution of $\mathbf{y}_t$ given $\mathcal{F}_{t-1}$ as

$$\mathbf{y}_t|\mathcal{F}_{t-1} \sim \mathcal{D}\left(\widetilde{\boldsymbol{\mu}}_t(\gamma_t), \widetilde{\boldsymbol{\Sigma}}_t(\gamma_t), \psi\right), \tag{B.1}$$

where:

$$\widetilde{\boldsymbol{\mu}}_t(\gamma_t) = \left(\mathbb{I} - \rho^* \mathsf{W}^*_t(\gamma_t)\right)^{-1} \mathbf{X}_t\boldsymbol{\beta} \tag{B.2}$$

$$\widetilde{\boldsymbol{\Sigma}}_t(\gamma_t) = \left(\mathbb{I} - \rho^* \mathsf{W}^*_t(\gamma_t)\right)^{-1} \boldsymbol{\Sigma} \left(\mathbb{I} - \rho^* \mathsf{W}^*_t(\gamma_t)\right)^{-1'}. \tag{B.3}$$

The following quantities are required:

$$\frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t} = \rho^* \mathbf{L}_t \mathbf{X}_t \boldsymbol{\beta} \tag{B.4}$$

$$\frac{\partial \widetilde{\boldsymbol{\Sigma}}_t}{\partial \gamma_t} = \mathbf{L}_t' \boldsymbol{\Sigma} \mathbf{A}_t^{-1} + \left(\mathbf{A}_t^{-1}\right)' \boldsymbol{\Sigma} \mathbf{L}_t, \tag{B.5}$$

with $\mathbf{L}_t = \mathbf{A}_t^{-1} \dot{\mathsf{W}}_t^* \mathbf{A}_t^{-1}$.

Theorem Appendix B.1. ***Score and information quantity of γ_t: Normal case.***

Let $\mathbf{y}_t|\mathcal{F}_{t-1}$ be distributed according to Equation (4). The log–pdf of the multivariate Gaussian distribution is proportional to:

$$\log p\left(\mathbf{y}_t|\gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right) \propto -\frac{1}{2}\log|\widetilde{\boldsymbol{\Sigma}}_t(\gamma_t)| - \frac{1}{2}\left(\mathbf{y}_t - \widetilde{\mu}_t(\gamma_t)\right)' \widetilde{\boldsymbol{\Sigma}}_t(\gamma_t)\left(\mathbf{y}_t - \widetilde{\mu}_t(\gamma_t)\right), \tag{B.6}$$

(a) The score of γ_t is given by:

$$\nabla_t\left(\mathbf{y}_t, \gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right) = \frac{\partial \log p\left(\mathbf{y}_t; \gamma, \boldsymbol{\eta}, \mathbf{X}_t\right)}{\partial \gamma}\Big|_{\gamma=\gamma_t} \tag{B.7}$$

$$= -\rho^*\left(\mathrm{Tr}\left[\mathbf{A}_t^{-1}\dot{\mathsf{W}}_t^*\right] - \mathbf{B}_t'\boldsymbol{\Sigma}_t^{-1}\dot{\mathsf{W}}_t^*\mathbf{y}_t\right), \tag{B.8}$$

where

$$\mathbf{B}_t = \mathbf{A}_t\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta} \tag{B.9}$$

and $\dot{\mathsf{W}}_t^*$ is an $N \times N$ matrix with (i,j)–th element equals to

$$\dot{\omega}_{ij,t}^* = \frac{\partial g_{ij}\left(\gamma, \mathbf{d}\right)}{\partial \gamma}\Big|_{\gamma=\gamma_t} \tag{B.10}$$

(b) The information quantity of γ_t is given by:

$$\mathcal{I}_t\left(\gamma_t, \boldsymbol{\eta}, \mathbf{X}_t\right) = \frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t}' \widetilde{\boldsymbol{\Sigma}}_t^{-1} \frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t} + \frac{1}{2}\mathrm{Tr}\left[\widetilde{\boldsymbol{\Sigma}}_t^{-1}\frac{\partial \widetilde{\boldsymbol{\Sigma}}_t}{\partial \gamma_t}\widetilde{\boldsymbol{\Sigma}}_t^{-1}\frac{\partial \widetilde{\boldsymbol{\Sigma}}_t}{\partial \gamma_t}\right] \tag{B.11}$$

where:

$$\widetilde{\boldsymbol{\mu}}_t(\gamma_t) = \left(\mathbb{I} - \rho^*\mathsf{W}_t^*(\gamma_t)\right)^{-1}\mathbf{X}_t\boldsymbol{\beta} \tag{B.12}$$

$$\widetilde{\boldsymbol{\Sigma}}_t(\gamma_t) = \left(\mathbb{I} - \rho^*\mathsf{W}_t^*(\gamma_t)\right)^{-1}\boldsymbol{\Sigma}\left(\mathbb{I} - \rho^*\mathsf{W}_t^*(\gamma_t)\right)^{-1'}. \tag{B.13}$$

and $\frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t}$ and $\frac{\partial \widetilde{\boldsymbol{\Sigma}}_t}{\partial \gamma_t}$ are defined in (B.4) and (B.5), respectively.

Theorem Appendix B.2. ***Score and information quantity of*** γ_t***: Student's*** t ***case.***

Recall that the log–pdf of the multivariate Student's t *distribution parametrized in terms of its covariance matrix is proportional to:*

$$p(\mathbf{y}_t|\gamma_t, \widetilde{\boldsymbol{\mu}}_t, \widetilde{\boldsymbol{\Sigma}}_t, \nu) \propto -\frac{\nu+N}{2}\left[1+\frac{(\mathbf{y}_t-\widetilde{\mu}_t(\gamma_t))'\,\widetilde{\boldsymbol{\Sigma}}_t(\gamma_t)\,(\mathbf{y}_t-\widetilde{\mu}_t(\gamma_t))}{\nu-2}\right]-\frac{1}{2}\log|\widetilde{\boldsymbol{\Sigma}}_t(\gamma_t)| \tag{B.14}$$

(a) The score with respect to γ_t *is given by:*

$$\frac{\partial \log p(\mathbf{y}_t|\cdot)}{\partial \gamma_t} = \left(\frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t}\right)' \frac{\partial \log p(\mathbf{y}_t|\cdot)}{\partial \widetilde{\boldsymbol{\mu}}_t} + \left(\frac{\partial vec\left(\widetilde{\boldsymbol{\Sigma}}_t\right)}{\partial \gamma_t}\right)' \frac{\partial \log p(\mathbf{y}_t|\cdot)}{\partial vec\left(\widetilde{\boldsymbol{\Sigma}}_t\right)}, \tag{B.15}$$

where:

$$\frac{\partial \log p(\mathbf{y}_t|\cdot)}{\partial \widetilde{\boldsymbol{\mu}}_t} = \frac{(\nu+N)\,\widetilde{\boldsymbol{\Sigma}}_t^{-1}\,(\mathbf{y}_t-\widetilde{\boldsymbol{\mu}}_t)}{w_t\,(\nu-2)} \tag{B.16}$$

$$\frac{\partial \log p(\mathbf{y}_t|\cdot)}{\partial vec\left(\widetilde{\boldsymbol{\Sigma}}_t\right)} = \frac{1}{2}\mathbf{D}_N'\,(\mathbf{J}_t\otimes\mathbf{J}_t)'\,[w_t\mathbf{J}_t\,(\mathbf{y}_t-\widetilde{\boldsymbol{\mu}}_t)-vec(\mathbf{I})], \tag{B.17}$$

and $\mathbf{J}_t$ *is an* $N\times N$ *matrix implicitly defined by* $\widetilde{\boldsymbol{\Sigma}}_t^{-1}=\mathbf{J}_t'\mathbf{J}_t$*. The matrix* $\mathbf{D}_N$ *is the duplication matrix such that* $\mathbf{D}_N vech(Q)=vec(Q)$ *where* $vech(\cdot)$ *is the half–vectorization operator. The quantities* $\frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t}$ *and* $\frac{\partial vec(\widetilde{\boldsymbol{\Sigma}}_t)}{\partial \gamma_t}=vec\left(\frac{\partial \widetilde{\boldsymbol{\Sigma}}_t}{\partial \gamma_t}\right)$ *are defined in* (B.4) *and* (B.5)*, respectively. The quantity* w_t *is equal to:*

$$w_t=\frac{\nu+N}{\nu-2+(\mathbf{y}_t-\widetilde{\boldsymbol{\mu}}_t)'\,\widetilde{\boldsymbol{\Sigma}}_t^{-1}\,(\mathbf{y}_t-\widetilde{\boldsymbol{\mu}}_t)}. \tag{B.18}$$

(b) The information quantity of γ_t *is given by*

$$\mathcal{I}(\gamma_t)=\begin{pmatrix}\frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t}' & \frac{\partial vec(\widetilde{\boldsymbol{\Sigma}}_t)}{\partial \gamma_t}'\end{pmatrix}\begin{pmatrix}\mathcal{I}_{\widetilde{\boldsymbol{\mu}}\widetilde{\boldsymbol{\mu}}} & \mathbf{0}\\ \mathbf{0} & \mathcal{I}_{vec(\widetilde{\boldsymbol{\Sigma}})vec(\widetilde{\boldsymbol{\Sigma}})}\end{pmatrix}\begin{pmatrix}\frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \gamma_t}\\ \frac{\partial vec(\widetilde{\boldsymbol{\Sigma}}_t)}{\partial \gamma_t}\end{pmatrix}, \tag{B.19}$$

where

$$\mathcal{I}_{\widetilde{\boldsymbol{\mu}}\widetilde{\boldsymbol{\mu}}}=\frac{\nu+N}{\nu+N+2}\frac{\nu}{\nu-2}\widetilde{\boldsymbol{\Sigma}}_t^{-1}, \tag{B.20}$$

and

$$\mathcal{I}_{vec(\tilde{\mathbf{\Sigma}})vec(\tilde{\mathbf{\Sigma}})} = \frac{1}{4} D_N'(\mathbf{J}_t \otimes \mathbf{J}_t)' \left[\frac{\nu + N}{\nu + 2 + N} \mathbf{G} - vec(I) vec(I)' \right] D_N(\mathbf{J}_t \otimes \mathbf{J}_t), \tag{B.21}$$

and $\mathbf{G}$ *is a* $N \times N$ *matrix with* $\mathbf{G}[(i-1)k + l, (j-1)k + m] = \delta_{ij}\delta_{lm} + \delta_{il}\delta_{jm} + \delta_{im}\delta_{jl}$ *for* $i, j, l, m = 1, \ldots, N$ *and* $\delta_{hk} = 1$ *if* $h = k$ *and* $\delta_{hk} = 0$ *otherwise, see Creal et al. (2011).*

Theorem Appendix B.3. ***Score and information quantity of*** $\sigma_{i,t}$**.**

In order to compute $u_{i,t}$ *in Equation* (A.2) *we first compute:*

$$\frac{\partial vec(\tilde{\mathbf{\Sigma}}_t)}{\partial \sigma_{i,t}} = \mathbf{E} vec\left(\mathbf{A}_t^{-1} \mathbf{U}_i \mathbf{A}_t^{-1\prime} \right), \tag{B.22}$$

where $\mathbf{E}$ *is the elimination matrix such that* $\mathbf{E}vec(\mathbf{M}) = vech(\mathbf{M})$ *and* $\mathbf{U}_i = \mathbf{e}_i \mathbf{e}_i'$ *where* $\mathbf{e}_i$ *is a vector of size* N *of zeros with 1 at its* i*-th position.*

The score and the information quantity for $\sigma_{i,t}$ *for the Student's t case are then computed as:*

$$\frac{\partial \log p(\mathbf{y}_t|\cdot}{\partial \sigma_{i,t}} = \frac{\partial vec(\tilde{\mathbf{\Sigma}}_t)}{\partial \sigma_{i,t}}' \frac{\partial \log p(\mathbf{y}_t|\cdot)}{\partial vec\left(\tilde{\mathbf{\Sigma}}_t\right)}, \tag{B.23}$$

and

$$\mathcal{I}(\sigma_{i,t}) = \frac{\partial vec(\tilde{\mathbf{\Sigma}}_t)}{\partial \sigma_{i,t}}' \mathcal{I}_{vec(\tilde{\mathbf{\Sigma}})vec(\tilde{\mathbf{\Sigma}})} \frac{\partial vec(\tilde{\mathbf{\Sigma}}_t)}{\partial \sigma_{i,t}}, \tag{B.24}$$

respectively. The Gaussian case is obtained by letting $\nu \to \infty$ *in* $\frac{\partial \log p(\mathbf{y}_t|\cdot)}{\partial vec(\tilde{\mathbf{\Sigma}}_t)}$ *and* $\mathcal{I}_{vec(\tilde{\mathbf{\Sigma}})vec(\tilde{\mathbf{\Sigma}})}$*.*

Appendix B.1. Derivatives of $g(\cdot)$

For the case

$$g\left(\mathsf{W}_t\right) = \frac{\mathsf{W}_t}{\lambda\left(\mathsf{W}_t\right)}, \tag{B.25}$$

the quantities $\dot{\omega}^*_{ij,t}$ needed for the evaluation of $\nabla_t\left(\mathbf{y}_t, \widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right)$ and $\mathcal{I}_t\left(\widetilde{\gamma}_t, \boldsymbol{\eta}, \mathbf{X}_t\right)$ are given by

$$\dot{\omega}^*_{ij,t} = -\omega^*_{ij,t} \left(1 + \frac{1}{\lambda\left(\mathsf{W}_t\right)} \frac{d\lambda\left(\mathsf{W}_t\right)}{d\gamma_t} \right), \tag{B.26}$$

and

$$\dot{\omega}^*_{ij,t} = -\omega^*_{ij,t} \left(\log d_{ij} + \frac{1}{\lambda\left(\mathsf{W}_t\right)} \frac{d\lambda\left(\mathsf{W}_t\right)}{d\gamma_t} \right), \tag{B.27}$$

for the inverse distance and negative exponential decay functions, respectively.[14]

Under row–normalization we have that $g(\mathsf{W}_t) = \mathsf{W}_t^*$ such that the i, j element of W_t^* is given by:

$$w_{ij}^* = \frac{f(\gamma_t, d_{ij})}{\sum_{h=1}^N f(\gamma_t, d_{ih})}. \tag{B.28}$$

The quantities $\dot{\omega}_{ij,t}^*$ for all $i \neq j$, have the form

$$\dot{\omega}_{ij,t}^* = -\frac{\omega_{ij,t}\left[-\log(d_{ij})\sum_{l=1}^N \omega_{il,t} + \sum_{l=1}^N \omega_{il,t}\log(d_{il})\right]}{\left[\sum_{l=1}^N \omega_{il,t}\right]^2}, \tag{B.29}$$

and

$$\dot{\omega}_{ij,t}^* = -\frac{\omega_{ij,t}\left[-d_{ij}\sum_{l=1}^N \omega_{il,t} + \sum_{l=1}^N \omega_{il,t}d_{il}\right]}{\left[\sum_{l=1}^N \omega_{il,t}\right]^2}, \tag{B.30}$$

for the inverse distance and negative exponential decay functions, respectively.

Appendix C. Proofs

Appendix C.1. Proof of Proposition 1

Proof: The stationarity and ergodicity of the limit sequence $\{\tilde{\gamma}_t\}_{t\in\mathbb{Z}}$,

$$\tilde{\gamma}_{t+1} = (1-\xi)\kappa + \alpha\tilde{s}_t^{\varepsilon} + \xi\tilde{\gamma}_t,$$

is obtained by the application of the Theorem 3.1 in Bougerol (1993) under conditions (i), (ii), (iv) and (v) in Proposition 5.1. The stationarity and ergodicity of the resulting sequence of the spatial weights

$$\{w_{ij,t}\}_{t\in\mathbb{Z}} = \{f(\gamma_t, d_{ij})\}_{t\in\mathbb{Z}} = \{f(\exp(\tilde{\gamma}_t), d_{ij})\}_{t\in\mathbb{Z}},$$

follows by the application of Proposition 4.3 in Krengel (1985) as f is assumed to be a measurable function (condition (iii) in Proposition 5.1). The stationarity and ergodicity of the normalized $\{\mathsf{W}_t^*\}_{t\in\mathbb{Z}}$ also follows, by Proposition 4.3 in Krengel (1985), from the assumed measurability of the normalization function $g(\cdot)$.

The stationarity and ergodicity of the data sequence $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$

$$\mathbf{y}_t = \mathbf{A}_t^{-1}\mathbf{X}_t\boldsymbol{\beta} + \mathbf{A}_t^{-1}\boldsymbol{\varepsilon}_t \quad \text{where} \quad \mathbf{A}_t = \mathbb{I}_N - \rho^*\mathsf{W}_t^*,$$

[14] The quantity $\frac{d\lambda(\mathsf{W}_t)}{d\gamma_t}$ is the derivative of the spectral radius of W_t with respect to γ_t and can be numerically evaluated by slightly increasing the computational cost.

is then naturally inherited from the stationarity and ergodicity of the sequence of the spatial weights matrices $\{\mathsf{W}_t\}_{t\in\mathbb{Z}}$, the sequence of exogenous regressors $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ (condition (iv)) and the sequence of innovations $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ (condition (v)).

Finally, the n bounded moments for the data follow by a simple application of the c_n-inequality in Dufour (2015), which ensures that $\exists\ 0 < c < \infty$ such that

$$\begin{aligned}\mathbb{E}\|\mathbf{y}_t\|^n &\leq c\mathbb{E}\|\mathbf{A}_t^{-1}\mathbf{X}_t\boldsymbol{\beta}\|^n + c\mathbb{E}\|\mathbf{A}_t^{-1}\boldsymbol{\varepsilon}_t\|^n \\ &\leq c\mathbb{E}\|\mathbf{A}_t^{-1}\|^n\mathbb{E}\|\mathbf{X}_t\boldsymbol{\beta}\|^n + c\mathbb{E}\|\mathbf{A}_t^{-1}\|^n\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^n < \infty\end{aligned}$$

with $\mathbb{E}\|\mathbf{A}_t^{-1}\|^n \leq A$ for some $A < \infty$, since that all the elements $w_{ij,t}$ of W_t are uniformly bounded by condition (iii). The bounded moments for the normalized weights $\mathbb{E}|w^*_{ij,t}|^m < \infty\ \forall\ m > 0$ follow trivially from the assumption that $g(\cdot)$ is differentiable, hence continuous on the compact $N \times N$ dimensional cube $[0,1]^{N\times N}$, and thus bounded by the BolzanoWeierstrass extreme value theorem. □

Appendix C.2. Proof of Proposition 2

Proof: The invertibility of the data sequence $\{\hat{\gamma}_t\}_{t\in\mathbb{N}}$ is obtained by the application of the Theorem 3.1 in Bougerol (1993) under conditions (i), (ii), (iv) and (v) in Proposition 5.1. The convergence of the sequence of the spatial weights is obtained as follows

$$\begin{aligned}\iota^t \sup_{\boldsymbol{\theta}\in\Theta}\sup_{ij}|\hat{w}_{ij,t} - w_{ij,t}| &= \iota^t \sup_{\boldsymbol{\theta}\in\Theta}\sup_{ij}|f(\hat{\gamma}_t, d_{ij}) - f(\gamma_t, d_{ij})| \\ &\leq \sup_{(\gamma,d)}\left|\frac{\partial f(\gamma,d)}{\partial\gamma}\right|\iota^t \sup_{\boldsymbol{\theta}\in\Theta}|\hat{\gamma}_t - \gamma_t| \xrightarrow{as} 0 \quad \text{as} \quad t\to\infty \quad \text{for some } \iota > 1,\end{aligned}$$

where $\iota^t \sup_{\boldsymbol{\theta}\in\Theta}|\hat{\gamma}_t - \gamma_t| \xrightarrow{as} 0$ by invertibility of $\{\hat{\gamma}_t\}_{t\in\mathbb{N}}$ and $\sup_{(\gamma,d)}\left|\frac{\partial f(\gamma,d)}{\partial\gamma}\right| < \infty$ by condition (iii). Furthermore, $\sup_{\boldsymbol{\theta}\in\Theta}|w_{ij,t}| < \infty$ is immediately implied by the uniform bound on condition (iii) in Proposition 5.1. Finally, we note that the e.a.s. convergence of the normalized weights $\hat{w}^*_{ij,t}$ follows immediately by measurability and Lipschitz continuity of the normalization function $g(\cdot)$,

$$\iota^t \sup_{\boldsymbol{\theta}\in\Theta}\sup_{ij}|\hat{w}^*_{ij,t} - w^*_{ij,t}| \leq \iota^t \sup_{\boldsymbol{\theta}\in\Theta}\sup_{ij} C\cdot|\hat{w}_{ij,t} - w_{ij,t}| \xrightarrow{as} 0 \quad \text{as} \quad t\to\infty \quad \text{for some } \iota > 1,$$

with $\sup_{\boldsymbol{\theta}\in\Theta}|w_{ij,t}| < \infty$ holding by continuity on the compact $N\times N$ dimensional cube $[0,1]^{N\times N}$, and thus boundedness by the BolzanoWeierstrass extreme value theorem. □

Appendix C.3. Proof of Proposition 5.3

Proof: We obtain the desired result by showing that the sample likelihood converges uniformly over the parameter space

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\hat{\mathsf{W}}_t) - \mathbb{E}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)\right| \xrightarrow{a.s.} 0\ \forall\ f_1\in\mathcal{F}\ \text{ as }\ T\to\infty \tag{C.1}$$

see e.g. Theorem 3.4 in White (1994) or Theorem 3.3 in Gallant and White (1988) for classical consistency theorems requiring identifiable uniqueness of $\boldsymbol{\theta}_0$, and Lemma 4.2 of Potscher and Prucha (1997) for a set consistency result derived only through the uniform convergence of the criterion function and the *regularity* of the *level sets*. Given the

compactness of the parameter space Θ, the regularity of level sets is ensured here as the limit log likelihood is continuous (by uniform convergence and continuity of the sample log likelihood); see Definition 4.1 in Potscher and Prucha (1997).

The uniform convergence in (C.1) follows by noting that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\hat{\mathsf{W}}_t)-\mathbb{E}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)\Big| \\
&\quad\leq \sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\hat{\mathsf{W}}_t)-\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)\Big| \\
&\qquad+\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)-\mathbb{E}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)\Big|.
\end{aligned} \tag{C.2}$$

The second term in the last inequality of (C.2) converges through the application of the ergodic theorem for separable Banach spaces in Rao (1962), as in Straumann and Mikosch (2006, Theorem 2.7). The stationarity and ergodicity of $\{\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)\}_{T\in\mathbb{Z}}$ holds by the stationarity and ergodicity of $\{(\mathbf{y}_t,\mathbf{X}_t)\}_{t\in\mathbb{Z}}$ and the invertibility of W_t established in Proposition 2, and Proposition 4.3 in Krengel (1985). The moment bound needed for the law of large numbers holds since that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)|=\frac{1}{2}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\log\det\boldsymbol{\Sigma}|+\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\log\det\mathbf{A}_t|+\frac{1}{2}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|(\mathbf{A}_t\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{A}_t\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})|<\infty$$

and $\Sigma_-<\det\boldsymbol{\Sigma}<\Sigma^+$ for some $\Sigma_->0$ and some $\Sigma^+<\infty$, $A_-<\det\mathbf{A}_t<A^+$ for some $A_->0$ and some $A^+<\infty$, and

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|(\mathbf{A}_t\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{A}_t\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})|\leq\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{V})|+\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{\mu}^\top\boldsymbol{\Sigma}\boldsymbol{\mu}|<\infty\ ,$$

where $\boldsymbol{\mu}$ and $\mathbf{V}$ denote the mean vector and covariance matrix, respectively, of the random vector $\mathbf{A}_t\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta}$, which has two bounded moments by Proposition 5.1.

The first term in the last inequality of (C.2) vanishes by the invertibility of the spatial weights matrix

$$\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\hat{\mathsf{W}}_t)-\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)\Big|\leq\frac{1}{T}\sum_{t=2}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t^*)}{\partial\mathsf{W}}\right\|\|\hat{\mathsf{W}}_t-\mathsf{W}_t\|\overset{as}{\to}0$$

where the mean value theorem is obtained equation by equation, with W_t^* denoting the implied set of mean–value points, under some abuse of notation. The a.s. convergence is achieved since that $\sup_{\boldsymbol{\theta}\in\Theta}\|\hat{\mathsf{W}}_t-\mathsf{W}_t\|\overset{eas}{\to}0$ holds by the invertibility of the filter proved in Proposition 5.2.

□

Appendix C.4. Proof of Theorem 5.4

Proof: The proof follows the same steps as those in Proposition 5.3. This time however, the moment bound needed for the law of large numbers holds by noting that

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)|=&\Big|\log\Gamma\left(\frac{\nu+n}{2}\right)\Big|+\Big|\log\Gamma\left(\frac{\nu}{2}\right)\Big|+\Big|\frac{n}{2}\log(\nu-2)\Big| \\
&+\Big|\log\det\mathbf{A}_t\Big|+\Big|\frac{1}{2}\log\det\boldsymbol{\Sigma}\Big|+\Big|\frac{\nu+n}{2}\Big|\,\mathbb{E}\Big|\log\left(1+\frac{\mathbf{z}_t'\mathbf{z}_t}{\nu-2}\right)\Big|<\infty
\end{aligned}$$

since that the restrictions on the parameter space Θ are such that $\nu>2$, ensuring that $\left|\frac{n}{2}\log(\nu-2)\right|<\infty$, $\Sigma_-<\det\boldsymbol{\Sigma}<\Sigma^+$ for some $\Sigma_->0$ and some $\Sigma^+<\infty$, $\left|\frac{1}{2}\log\det\boldsymbol{\Sigma}\right|<\infty$, and $A_-<\det A_t<A^+$ for some

$A_- > 0$ and some $A^+ < \infty$, $\left|\log\det \mathbf{A}_t\right| < \infty$. Finally, the logarithmic moment for $\mathbf{z}_t'\mathbf{z}_t$ is guarantee by the fact that Proposition 5.1 and 5.2 hold with $n > 0$, and hence $\mathbb{E}|\mathbf{z}_t'\mathbf{z}_t|^m < 0$ for some $m > 0$. □

Appendix C.5. Proof of Corollary 5.1

Proof: We obtain the consistency to a singleton $\boldsymbol{\theta}_0$ through the uniform convergence of the criterion function established in the proof of propositions 5.3 and 5.4, the identifiable uniqueness shown below, and by appealing to e.g. Theorem 3.4 in White (1994) or Theorem 3.3 in Gallant and White (1988).

Identifiable uniqueness of $\boldsymbol{\theta}_0$ is the condition that

$$\sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\epsilon} \mathbb{E}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t) < \mathbb{E}\log p(\mathbf{y}_t|\boldsymbol{\theta}_0,\mathbf{X}_t,\mathsf{W}_t)\ \forall\ \epsilon > 0; \tag{C.3}$$

The fact that $\boldsymbol{\theta}_0$ is the unique maximizer of the limit log likelihood follows easily under the assumption of identifiability, by application of Gibbs' inequality. Finally, (C.3) is implied by the uniqueness of $\boldsymbol{\theta}_0$, the compactness of the parameter space Θ, and the continuity of the limit log likelihood $\mathbb{E}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)$ in $\boldsymbol{\theta} \in \Theta$, see e.g. Potscher and Prucha (1986). The continuity of the limit criterion follows from the continuity of $\frac{1}{T}\sum_{t=2}^{T}\log p(\mathbf{y}_t|\boldsymbol{\theta},\mathbf{X}_t,\mathsf{W}_t)$ in $\boldsymbol{\theta} \in \Theta\ \forall\ T \in \mathbb{N}$ and the uniform convergence in (C.1), see also Potscher and Prucha (1986).

Appendix C.6. Proof of Proposition 5.5

Proof: We obtain the asymptotic Gaussianity of the maximum likelihood estimator by employing the usual quadratic expansion of the log likelihood found e.g. in White (1994, Theorem 6.2) and relying on (i) the strong consistency of $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$; (ii) the a.s. twice continuous differentiability of the log likelihood $\frac{1}{T}\sum_{t=2}^{T}\ell_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$; (iii) the asymptotic normality of the score

$$\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\hat{\ell}_t'(\boldsymbol{\theta}_0) \overset{d}{\to} \mathcal{N}(0,\mathcal{J}(\boldsymbol{\theta}_0)\big), \qquad \mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}\big(\ell_t'(\boldsymbol{\theta}_0)\ell_t'(\boldsymbol{\theta}_0)^\top\big); \tag{C.4}$$

(iv) the uniform convergence of the likelihood's second derivative,

$$\sup_{\boldsymbol{\theta}\in\Theta}\big\|\frac{1}{T}\sum_{t=2}^{T}\hat{\ell}_T''(\boldsymbol{\theta}) - \frac{1}{T}\sum_{t=2}^{T}\hat{\ell}_\infty''(\boldsymbol{\theta})\big\| \overset{a.s.}{\to} 0; \tag{C.5}$$

and finally, (v) the non–singularity of the limit $\ell_\infty''(\boldsymbol{\theta}) = \mathbb{E}\ell_t''(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$.

The consistency of the MLE $\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ in (i) follows by Proposition 5.3 or 5.4 and the additional assumption that $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$. The differentiability required by (ii) is trivially satisfied for either the Gaussian or Student's-t densities used in Proposition 5.3 and 5.4. The asymptotic normality of the score in (C.7) follows by first noting that

$$\begin{aligned}\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\hat{\ell}_t'(\boldsymbol{\theta}_0) &= \sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\Big[\hat{\ell}_t'(\boldsymbol{\theta}_0) - \ell_t'(\boldsymbol{\theta}_0)) + \ell_t'(\boldsymbol{\theta})\Big] \\ &= \frac{1}{T}\sum_{t=2}^{T}\sqrt{T}\Big(\hat{\ell}_t'(\boldsymbol{\theta}_0) - \ell_t'(\boldsymbol{\theta}_0)\Big) + \sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\ell_t'(\boldsymbol{\theta}_0).\end{aligned} \tag{C.6}$$

Asymptotic normality for the second term in (C.6) holds by an application of the CLT for SE martingales in Billingsley (1961), to obtain

$$\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\ell_t'(\boldsymbol{\theta}_0) \overset{d}{\to} \mathcal{N}(0,\mathcal{J}(\boldsymbol{\theta}_0)\big) \ \text{ as } \ T\to\infty, \tag{C.7}$$

where $\mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}(\ell_t'(\boldsymbol{\theta}_0)\ell_t'(\boldsymbol{\theta}_0)^\top) < \infty$. In addition, we have that

$$\sqrt{T}\Big[\frac{1}{T}\sum_{t=2}^{T}\Big(\hat{\ell}_t'(\boldsymbol{\theta}_0) - \ell_t'(\boldsymbol{\theta}_0)\Big)\Big] \xrightarrow{p} 0$$

as $T \to \infty$, by noting that

$$\sqrt{T}\Big\|\hat{\ell}_t'(\boldsymbol{\theta}_0) - \ell_t'(\boldsymbol{\theta}_0)\Big\| \leq \Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)\Big\| \times \sqrt{T}\max\Big\{\|\widehat{\mathsf{W}}_t - \mathsf{W}_t\| \ , \ \|\widehat{\dot{\mathsf{W}}}_t - \dot{\mathsf{W}}_t\|\Big\} \ ,$$

where $\Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)\Big\|$ denotes the derivative of $\ell_t'(\boldsymbol{\theta}_0)$ w.r.t. $(\mathsf{W}_t, \dot{\mathsf{W}}_t)$ and evaluated, equation by equation, at a point $(\mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)$ between $(\mathsf{W}_t, \dot{\mathsf{W}}_t)$ and $(\hat{\mathsf{W}}_t, \widehat{\dot{\mathsf{W}}}_t)$ as typically applied on the multivariate mean–value–theorem.[15] Now $\sqrt{T}\max\Big\{\|\widehat{\mathsf{W}}_t - \mathsf{W}_t\| \ , \ \|\widehat{\dot{\mathsf{W}}}_t - \dot{\mathsf{W}}_t\|\Big\}$ will vanish in probability by the e.a.s. assumption, and $\Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)\Big\|$ can be shown to be bounded in probability since that, for t large enough, we have $\max\Big\{\|\widehat{\mathsf{W}}_t - \mathsf{W}_t\| \ , \ \|\widehat{\dot{\mathsf{W}}}_t - \dot{\mathsf{W}}_t\|\Big\} < 1$ a.s., and hence

$$\begin{aligned}
&\lim_{t,c\to\infty} P\Big(\Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)\Big\| > c\Big) \\
&\qquad \leq \lim_{t,c\to\infty} P\Big(\sup_{(W,W'):\max\left\{\|\mathsf{W}_t - W\| \ , \ \|\dot{\mathsf{W}}_t - W'\|\right\}<1} \Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t + W, \dot{\mathsf{W}}_t + W')\Big\| > c\Big)
\end{aligned}$$

and since the sequence $\Big\{\sup_{(W,W'):\max\left\{\|\mathsf{W}_t - W\| \ , \ \|\dot{\mathsf{W}}_t - W'\|\right\}<1} \Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t + W, \dot{\mathsf{W}}_t + W')\Big\|\Big\}$ is SE by continuity of the supremum operator, we have,

$$\lim_{t,c\to\infty} P\Big(\sup_{(W,W'):\max\left\{\|\mathsf{W}_t - W\| \ , \ \|\dot{\mathsf{W}}_t - W'\|\right\}<1} \Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t + W, \dot{\mathsf{W}}_t + W')\Big\| > c\Big) = 0$$

which implies

$$\Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)\Big\| \times \sqrt{T}\max\Big\{\|\widehat{\mathsf{W}}_t - \mathsf{W}_t\| \ , \ \|\widehat{\dot{\mathsf{W}}}_t - \dot{\mathsf{W}}_t\|\Big\}.$$

The uniform convergence in (iv) is obtained in a similar fashion by noting that, for every $\boldsymbol{\theta} \in \Theta$, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T}\hat{\ell}_t''(\boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=2}^{T}\Big[\hat{\ell}_t''(\boldsymbol{\theta}) - \ell_t''(\boldsymbol{\theta})) + \ell_t'(\boldsymbol{\theta})\Big] \\
&= \frac{1}{T}\sum_{t=2}^{T}\Big(\hat{\ell}_t''(\boldsymbol{\theta}) - \ell_t''(\boldsymbol{\theta})\Big) + \sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\ell_t''(\boldsymbol{\theta}).
\end{aligned} \tag{C.8}$$

Uniform convergence for the second term in (C.8) holds under the moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\ell_t''(\boldsymbol{\theta})\| < \infty$ and by the SE nature of $\{\ell_T''\}_{t\in\mathbb{Z}}$. In addition, we have that

$$\sup_{\boldsymbol{\theta}\in\Theta}\Big\|\frac{1}{T}\sum_{t=2}^{T}\Big(\hat{\ell}_t''(\boldsymbol{\theta}_0) - \ell_t''(\boldsymbol{\theta}_0)\Big)\Big\| \xrightarrow{p} 0$$

as $T \to \infty$, by noting that

$$\Big\|\hat{\ell}_t''(\boldsymbol{\theta}_0) - \ell_t''(\boldsymbol{\theta}_0)\Big\| \leq \Big\|\nabla_{\mathsf{W}}\ell_t''(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*, \ddot{\mathsf{W}}_t^*)\Big\| \times \sqrt{T}\max\Big\{\|\widehat{\mathsf{W}}_t - \mathsf{W}_t\| \ , \ \|\widehat{\dot{\mathsf{W}}}_t - \dot{\mathsf{W}}_t\| \ , \ \|\widehat{\ddot{\mathsf{W}}}_t - \ddot{\mathsf{W}}_t\|\Big\} \ ,$$

where $\Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*, \ddot{\mathsf{W}}_t^*)\Big\|$ denotes the derivative of $\ell_t''(\boldsymbol{\theta}_0)$ w.r.t. $(\mathsf{W}_t, \dot{\mathsf{W}}_t, \ddot{\mathsf{W}}_t)$ and evaluated, equation by equation, at a point $(\mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)$ between $(\mathsf{W}_t, \dot{\mathsf{W}}_t, \ddot{\mathsf{W}}_t)$ and $(\hat{\mathsf{W}}_t, \widehat{\dot{\mathsf{W}}}_t, \widehat{\ddot{\mathsf{W}}}_t)$. Now $\sqrt{T}\max\Big\{\|\widehat{\mathsf{W}}_t - \mathsf{W}_t\| \ , \ \|\widehat{\dot{\mathsf{W}}}_t - \dot{\mathsf{W}}_t\|, \ \|\widehat{\ddot{\mathsf{W}}}_t - \ddot{\mathsf{W}}_t\|\Big\}$ will vanish in probability by the e.a.s. assumption, and $\Big\|\nabla_{\mathsf{W}}\ell_t''(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*, \ddot{\mathsf{W}}_t^*)\Big\|$ can be shown to be bounded in probability by a similar argument as the one above for $\Big\|\nabla_{\mathsf{W}}\ell_t'(\boldsymbol{\theta}_0, \mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)\Big\|$. □

[15] The usual abuse of notation is applied here as a different point $(\mathsf{W}_t^*, \dot{\mathsf{W}}_t^*)$ may be selected equation by equation.

References

Ahrens, A. and Bhattacharjee, A. (2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics*, 3(1):128–155.

Andree, B., Blasques, F., and Koomen, E. (2017). Smooth transition spatial autoregressive models. Technical report, Tinbergen Institute Discussion Paper.

Bailey, N., Holly, S., and Pesaran, M. H. (2016). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, 31(1):249–280.

Bao, Y. and Ullah, A. (2007). Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics*, 137(2):396–413.

Benjanuvatra, S., Burridge, P., et al. (2015). Qml estimation of the spatial weight matrix in the mr-sar model. *DERS University of York Working Paper*.

Bhattacharjee, A. and Jensen-Butler, C. (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics*, 43(4):617–634.

Billé, A. G., Benedetti, R., and Postiglione, P. (2017). A two-step approach to account for unobserved spatial heterogeneity. *Spatial Economic Analysis*.

Billé, A. G. and Leorato, S. (2020). Partial ml estimation for spatial autoregressive nonlinear probit models with autoregressive disturbances. *Econometric Reviews*, 39(5):437–475.

Billingsley, P. (1961). The Lindeberg-Levy Theorem for Martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12,(1):1019–1052.

Blasques, F., Koopman, S., and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343.

Blasques, F., Koopman, S. J., Łasak, K., and Lucas, A. (2016a). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting*, 32(3):875 – 887.

Blasques, F., Koopman, S. J., and Lucas, A. (2014). Maximum likelihood estimation for generalized autoregressive score models. Technical report, Tinbergen Institute.

Blasques, F., Koopman, S. J., Lucas, A., and Schaumburg, J. (2016b). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics*, 195(2):211–223.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization*, 31(4):942–959.

Brady, R. R. (2011). Measuring the diffusion of housing prices across space and over time. *Journal of Applied Econometrics*, 26(2):213–231.

Catania, L. and Billé, A. G. (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics*, pages 1–27.

Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities

and correlations. *Journal of Business & Economic Statistics*, 29(4):552–563.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Dufour, J.-M. (2013). Properties of moments of random variables.

Eder, A. and Keiler, S. (2015). Cds spreads and contagion amongst systemically important financial institutions–a spatial econometric approach. *International Journal of Finance & Economics*, 20(4):291–309.

Engle, R. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Federal Reserve Bank of St. Louis (2018a). Oecd based recession indicators for euro area from the period following the peak through the trough [eurorecd], retrieved from fred, federal reserve bank of st. louis; `https://fred.stlouisfed.org/series/EURORECD`. Technical report.

Federal Reserve Bank of St. Louis (2018b). Oecd based recession indicators for the united kingdom from the peak through the trough [gbrrecdm], retrieved from fred, federal reserve bank of st. louis; `https://fred.stlouisfed.org/series/GBRRECDM`. Technical report.

Fischer, M. M., Scherngell, T., and Reismann, M. (2009). Knowledge spillovers and total factor productivity: evidence using a spatial panel data model. *Geographical Analysis*, 41(2):204–220.

Gallant, R. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Cambridge University Press.

Halleck Vega, S. and Elhorst, J. P. (2015). The slx model. *Journal of Regional Science*, 55(3):339–363.

Han, X. and Lee, L.-F. (2016). Bayesian analysis of spatial panel autoregressive models with time-varying endogenous spatial weight matrices, common factors, and random coefficients. *Journal of Business & Economic Statistics*, 34(4):642–660.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press.

Hepple, L. W. (2004). Bayesian model choice in spatial econometrics. *Advances of Econometrics*, 18:101–126.

Holly, S., Pesaran, M. H., and Yamagata, T. (2010). A spatio-temporal model of house prices in the usa. *Journal of Econometrics*, 158(1):160–173.

Holly, S., Pesaran, M. H., and Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the uk. *Journal of urban economics*, 69(1):2–23.

Kelejian, H. H. and Piras, G. (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, 46:140–149.

Kelejian, H. H. and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1):53–67.

Kostov, P. (2010). Model boosting for spatial weighting matrix selection in spatial lag models. *Environment and Planning B: Planning and Design*, 37(3):533–549.

Krengel, U. (1985). *Ergodic theorems*. De Gruyter studies in Mathematics, Berlin.

Lee, L.-f. and Yu, J. (2012). Qml estimation of spatial dynamic panel data models with time varying spatial weights matrices. *Spatial Economic Analysis*, 7(1):31–74.

LeSage, J. P. and Pace, R. K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, 140(1):190–214.

OECD (2018). General government debt (indicator). doi: 10.1787/a0528cc2-en (accessed on 18 july 2018). Technical report.

Otto, P. and Steinert, R. (2018). Estimation of the spatial weighting matrix for spatiotemporal data under the presence of structural breaks. *arXiv preprint arXiv:1810.06940*.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Porcu, E., Bevilacqua, M., and Genton, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association*, 111(514):888–898.

Potscher, B. M. and Prucha, I. R. (1986). *Consistency in nonlinear econometrics: A generic uniform law of large numbers and some comments on recent results*. Working paper, Department of Economics, University of Maryland.

Qu, X. and Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2):209–232.

Qu, X., Lee, L.-f., and Yu, J. (2017). Qml estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics*, 197(2):173–201.

Qu, X., Wang, X., and Lee, L.-f. (2016). Instrumental variable estimation of a spatial dynamic panel model with endogenous spatial weights when t is small. *The Econometrics Journal*, 19(3):261–290.

Rao, R. R. (1962). Relations between Weak and Uniform Convergence of Measures with Applications. *The Annals of Mathematical Statistics*, 33(2):659–680.

Shi, W. and Lee, L.-f. (2017a). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics*.

Shi, W. and Lee, L.-f. (2017b). A spatial panel data model with time varying endogenous weights matrices and common factors. *Regional Science and Urban Economics (to appear)*.

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

Wang, W. and Yu, J. (2015). Estimation of spatial panel data models with time varying spatial weights matrices. *Economics Letters*, 128:95–99.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge Books. Cambridge University Press.