Spillover Dynamics for Systemic Risk Measurement Using Spatial Financial Time Series Models[1]

Francisco Blasques[a], Siem Jan Koopman[a,b], Andre Lucas[a], Julia Schaumburg[a]

[a] VU University Amsterdam and Tinbergen Institute, The Netherlands

[b] CREATES, Aarhus University, Denmark

September 2014

Abstract

We extend the well-known static spatial lag model by introducing a new model with time-varying spatial dependence. We show that the updating steps in this model have information theoretic optimality properties. All parameters of the model can be conveniently estimated by maximum likelihood. We establish the theoretical properties of the model and show that the maximum likelihood estimators of the static parameters are consistent and asymptotically normal. Using spatial weights based on cross-border lending data and European sovereign CDS spread data over the period 2009–2014, we find high, time-varying spatial spillovers in the perceived credit riskiness of European sovereigns during the sovereign debt crisis. We find a particular downturn in spatial dependence in the first half of 2012 after the policy measures taken by the European Central Bank. The findings are robust to a wide range of alternative model specifications.

Keywords: Spatial correlation, time-varying parameters, systemic risk, European debt crisis, generalized autoregressive scores.

1 Introduction

We propose a new parsimonious model to measure the time-varying cross-sectional dependence in European sovereign credit spread changes. The model builds on the well-known spatial lag model for panel data. The strength of contemporaneous spillover effects is summarized in a single time-varying parameter: the spatial dependence parameter. We argue that this parameter may be interpreted as a measure of sovereign systemic risk that relates to the connectedness of the system.

Our paper contributes to two strands of literature. First, we contribute to the applied spatial econometrics literature. Spatial models have been widely used in applied geographic and regional science studies, and have recently also been applied in empirical finance; see Fernandez (2011) for a CAPM model augmented by spatial dependencies, Wied (2013), Arnold et al. (2013), and Asgharian et al. (2013) for analyses of spatial dependencies in stock markets, Denbee et al. (2013) for a network approach to assess interbank liquidity, and Saldias (2013) for a spatial error model to identify sector risk determinants. Keiler and Eder (2013) and Tonzer (2013) both use spatial lag models, to model CDS spreads of financial institutions and banking sector risks, respectively.

All of the above models, however, treat the spatial dependence parameter as static. To the best of our knowledge, explicitly endowing the spatial dependence parameter in the spatial lag model with time series dynamics is a new development. Allowing for such dynamics may be important empirically; see for example our financial systemic stability application in Section 6. We model the dynamics using the score driven framework proposed by Creal et al. (2011, 2013) and Harvey (2013). Given the nonlinear impact of the time-varying parameter on the model, the theoretical properties of this model and the asymptotic properties of the maximum likelihood estimator (MLE) for the remaining static parameters are challenging and have not been established so far. We show under what conditions the filtered spatial dependence parameters are well behaved, such that the model is invertible. Invertibility is a key property for establishing consistency and asymptotic normality of the MLE; see for example Wintenberger (2013). We derive new conditions for the asymptotic properties of the MLE compared to Blasques et al. (2014b), allowing for exogenous regressors to be part of the specification. We also discuss the information theoretic optimality of the model and illustrate in a simulation study that the model is able to track a range of different patterns for the time-varying spatial dependence parameter.

Second, we contribute to the literature that studies the dynamics of financial systemic risk in

the context of a network of sovereigns or financial firms. Since the beginning of the European sovereing debt crisis in 2009, the sharp increases and comovements of sovereign credit spreads have been the subject of a growing number of empirical studies in finance. For instance, by employing an asset pricing model, Ang and Longstaff (2013) investigate the differences between U.S. and European credit default swap (CDS) spreads as a reflection of systemic risk. Lucas et al. (2014) and Kalbaska and Gatkowski (2012) use multivariate time series models to model comovements in European sovereign CDS spreads. Ait-Sahalia et al. (2014) model sovereign credit default intensities using multivariate jump processes. De Santis (2012) and Arezki et al. (2011) study credit risk spillover effects that are induced by rating events, such as downgrades of Greek government bonds. Leschinski and Bertram (2013) find contagion effects in European sovereign bond spreads using the simultaneous equations approach of Pesaran and Pick (2007). Caporin et al. (2013), on the other hand, employ Bayesian quantile regressions, and conclude that comovements in European credit spreads during the debt crisis are only due to increased volatities, but not contagion.

Our approach differs from the studies above since we introduce cross-sectional correlation not only through contemporaneous error correlations, but also through spillovers induced by shocks to the regressors, such as stock market crashes or interbank lending rates. Furthermore, we explicitly offer financial sector linkages as the source of sovereign credit risk comovements. This view is supported by the results of Korte and Steffen (2013), Kallestrup et al. (2013), Gorea and Radev (2013), and Beetsma et al. (2012), in which cross-border exposures between international financial sectors are relevant drivers of sovereign credit spreads. By exploiting these debt interconnections as economic distances between sovereigns in our spatial model, we obtain a scalar time-varying (spatial) dependence coefficient. We interpret this parameter in the systemic context as the overall tendency for shock spillovers. As such, it provides a measure of systemic risk that is easy to monitor over time. Also, in contrast to earlier empirical literature, we allow the spatial dependence parameter to vary over time, implying that sovereign interconnectedness and therefore systemic risk, may be time-varying as well.

We organize the remainder of this paper as follows. Section 2 introduces our spatial score model with time-varying parameters and states that the updating mechanism possesses information theoretic optimality properties. We examine the theoretical properties of the model in Section 3. In Section 4, we provide Monte Carlo evidence of the model's ability to track different dynamic patterns in spatial dependence over time. Section 5 describes the data for our study on European sovereign CDS spread dynamics. Section 6 provides the results, and Section 7 concludes.

2 Spatial models with dynamic spatial dependence

2.1 Static spatial lag model for panel data

The spatial lag model for panel data is given by

$$y_t = \rho W y_t + X_t \beta + e_t, \qquad e_t \sim p_e(e_t; \Sigma, \lambda), \qquad t = 1, \dots, T, \tag{1}$$

where $y_t = (y_{1t}, \dots, y_{nt})'$ denotes a vector of n cross-sectional observations at time t, ρ is the spatial dependence coefficient, W is an $n \times n$ matrix of exogenous spatial weights, X_t is an $n \times k$ matrix of exogenous regressors, β is a $k \times 1$ vector of unknown coefficients, including an intercept, and e_t is an $n \times 1$ disturbance vector with multivariate density $p_e(e_t, \Sigma; \lambda)$, mean zero, unknown $k \times k$ covariance (or scale) matrix Σ, and other parameters describing the shape of the distribution as collected in the parameter vector λ. For example, if p_e is a Student's t distribution, λ denotes the degrees of freedom parameter.

Model (1) implies that each entry y_{it}, for $i = 1, \dots, n$, of the vector y_t depends on k individual-specific regressors x_{it} as well as on other entries y_{jt} for $j \neq i$. For a moderately large n, we cannot estimate such a system of contemporaneous dependencies without imposing further restrictions. The idea of a spatial dependence model is to specify the spatial weight matrix W as a function of geographic or economic distances, and in this way exogenously define a neighborhood structure between the cross-sectional units. It is standard practice to use a row-normalized weight matrix W such that $\sum_{j=1}^{n} w_{ij} = 1$ for $i = 1, \dots, n$, where w_{ij} is the (i, j)th element of W. The impact of the (spatially weighted) contemporaneous dependent variables $W y_t$ on y_t is captured by a scalar spatial dependence parameter ρ. For shocks to die out over space, we require $\rho \in (1/\omega_{min}, 1)$ where ω_{min} is the smallest eigenvalue of W; see for example Lee (2004).

The basic form of the spatial lag model (1) can capture nonlinear feedback effects across units. This can be shown by rewriting the model as

$$y_t = Z X_t \beta + Z e_t, \tag{2}$$

where we assume that the inverse matrix $Z = (\mathrm{I}_n - \rho W)^{-1}$ exists, with I_n denoting the $n \times n$ identity matrix. Using an infinite power series expansion as in LeSage and Pace (2008), we obtain

$$y_t = X_t \beta + \rho W X_t \beta + \rho^2 W^2 X_t \beta + \cdots + e_t + \rho W e_t + \rho^2 W^2 e_t + \cdots. \tag{3}$$

Equation (3) reveals that e_{it} and $x'_{it}\beta$ for unit i spill over to other units $j \neq i$. The extent of spillover depends on the relative proximity of j to i via the weight matrix W and the spatial dependence parameter ρ. At the same time, there are possible feedback effects back to unit i itself, for example if w_{ij} and w_{ji} are both non-zero, such that i and j are mutual neighbors, and i is a 'second-order neighbor' to itself.

The simultaneous equations structure of (1) leads to an endogeneity problem and causes the least squares estimator in (1) to be inconsistent, As an alternative solution, we can estimate the parameters by the method of Maximum Likelihood (ML) or Quasi-ML (QML) where the latter is typically based on the normal distribution.[1] The ML Estimator (MLE) for spatial models with static dependence parameter was first studied in Ord (1975) in the context of cross-sectional data sets. Lee (2004) derives asymptotic properties of the QML Estimator (QMLE) for $n \to \infty$, and Hillier and Martellosio (2013) investigate its finite sample distribution. Large n and large T asymptotics for the QMLE of the spatial model with static dependence parameter are studied in Yu et al. (2008). For further textbook treatments, we refer to Anselin (1988) and LeSage and Pace (2008). For a survey on the panel data spatial lag model and parameter estimation, see Lee and Yu (2010).

2.2   Score dynamics for the spatial dependence parameter

We can interpret the spatial dependence parameter ρ in (1) as a measure of the strength of cross-sectional spillovers. In many empirical applications involving panel data, it is unrealistic to assume that ρ is constant over the entire sample period. We therefore introduce a time-varying spatial dependence parameter ρ_t in the model, that is

$$y_t = \rho_t W y_t + X_t \beta + e_t, \qquad e_t \sim p_e(e_t; \Sigma, \lambda), \qquad t = 1, \ldots, T, \tag{4}$$

where $\rho_t = h(f_t)$ is a monotonic transformation of a time-varying parameter f_t. We choose the link function h such that $\rho_t \in (-1, 1)$. To describe the dynamics of f_t, we adopt the autoregressive score framework of Creal et al. (2011, 2013) and Harvey (2013). The score framework for time-varying parameters has been adopted successfully in a range of different empirical settings, including the multivariate volatility model of Creal et al. (2011), the systemic risk model of Oh and Patton (2013) and Lucas et al. (2014), the credit risk dynamic factor model of Creal et al. (2014), and the location and scale models with fat tails of Harvey and Luati (2014).[2]

---

[1] Alternatively, we can use GMM as in, for example, Kelejian and Prucha (2010).

[2] See www.gasmodel.com for a more complete compilation of papers.

The score framework centers around the use of the scaled score of the conditional density p_e to drive the time-variation in f_t. The updating equation for f_t is given by

$$f_{t+1} = \omega + A s_t + B f_t, \tag{5}$$

where ω, A, and B are fixed unknown parameters, and $s_t = S_t \nabla_t$ is the scaled score function. The scaled score function is defined as the first derivative of the predictive loglikelihood function at time t with respect to f_t, possibly multiplied by some local scaling factor S_t. In our case, the score function is given by $\nabla_t = \partial \ell_t / \partial f_t$ where

$$\ell_t = \log p_e \left(y_t - h(f_t) W y_t - X_t \beta, \Sigma; \lambda \right) + \log \left| (\mathrm{I}_n - h(f_t) W) \right|. \tag{6}$$

Throughout this paper, we use unit scaling, that is $S_t \equiv 1$ such that $s_t = \nabla_t$. Other scaling choices are also feasible; see Creal et al. (2013).[3] Equation (6) differs from the likelihood of a simple linear regression model by the term $\log \left| (\mathrm{I}_n - h(f_t) W) \right|$. This term accounts for the nonlinearity of the model in ρ_t as shown in equation (2). We define the vector of static parameters $\boldsymbol{\theta} = (\omega, A, B, \beta, \lambda)'$ and estimate $\boldsymbol{\theta}$ via the numerical maximization of the likelihood function

$$\mathcal{L}_T = \sum_{t=1}^{T} \ell_t. \tag{7}$$

We consider two specifications for the disturbance density p_e, namely the multivariate normal distribution and the multivariate Student's t distribution. The latter is particularly relevant for our empirical study because changes in credit default swap (CDS) spreads may be fat-tailed. Also, Creal et al. (2011) and Harvey and Luati (2014) argue that the Student's t distribution can render the dynamics more robust to incidental influential observations and outliers.

Using the standard expression for the multivariate normal density, we obtain the time t contribution to the loglikelihood function as

$$
\begin{aligned}
\ell_t &= \log \left| \mathrm{I} - h(f_t) W \right| - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log \left| \Sigma \right| \\
&\quad - \frac{1}{2} (y_t - h(f_t) W y_t - X_t \beta)' \Sigma^{-1} (y_t - h(f_t) W y_t - X_t \beta),
\end{aligned}
$$

---

[3] In a simulation (not reported here) we show that different choices of scaling, such as scaling by the inverse information matrix scaling or by its square root, did not have much impact on our empirical results.

and the resulting score

$$\nabla_t = \left(y_t'W'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta) - \text{tr}(Z(f_t)W)\right) \cdot \dot{h}(f_t), \tag{8}$$

where $\text{tr}(\cdot)$ is the trace operator, $Z(f_t) = (I_n - h(f_t)W)^{-1}$, and $\dot{h}(f_t)$ is the first derivative of the transformation function h with respect to f_t. For instance, if $h(f_t) = \gamma \tanh(f_t)$ with $\gamma \in (0,1)$, then $\dot{h}(f_t) = \gamma(1 - \tanh^2(f_t))$. When the density of the disturbance vector e_t is a multivariate Student's t distribution with λ degrees of freedom, we obtain

$$
\begin{aligned}
\ell_t &= \log|Z(f_t)^{-1}| + \log\left(\frac{\Gamma\left(\frac{\lambda+n}{2}\right)}{|\Sigma|^{1/2}(\lambda\pi)^{n/2}\Gamma\left(\frac{\lambda}{2}\right)}\right) \\
&\quad - \left(\frac{\lambda+n}{2}\right)\log\left(1 + \frac{(y_t - h(f_t)Wy_t - X_t\beta)'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta)}{\lambda}\right),
\end{aligned}
$$

with the corresponding score function

$$
\begin{aligned}
\nabla_t &= \left(\tilde{w}_t \cdot y_t'W'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta) - \text{tr}(Z(f_t)W)\right) \cdot \dot{h}(f_t), \tag{9} \\
\tilde{w}_t &= (1 + \lambda^{-1}n)/\left(1 + \lambda^{-1}(y_t - h(f_t)Wy_t - X_t\beta)'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta)\right).
\end{aligned}
$$

Is is easy to verify that for $\lambda \to \infty$ we obtain $\tilde{w}_t \to 1$. The score expression in (9) in that case collapses to the one in (8). The weight $\tilde{w}_t$ is small if the residuals $y_t - h(f_t)Wy_t - X_t\beta$ are 'large' in a multivariate sense. The implication of a small weight $\tilde{w}_t$ is that the observation has a smaller impact on the updates of f_t. This provides a robustness feature to the dynamics of f_t if we assume a fat-tailed distribution such as the Student's t; see also the discussion in Creal et al. (2011, 2013) and Harvey (2013). The intuition is straightforward: a large residual may be attributable to the fat-tailedness of the Student's t distribution rather than to a recent increase in the spatial correlation parameter $\rho_t = h(f_t)$.

The score expressions in (8) and (9) also depart from the expressions for the standard linear regression model. In particular, the additional correction term $-\text{tr}(Z(f_t)W)$ accounts for the simultaneity bias in the standard least squares estimator and follows from the presence of the term $\log|Z(f_t)^{-1}|$ in the likelihood at time t. Economically, this term accounts for the fact that there may be feedback effects from unit i to unit j and then back to unit i. Hence the spatial autoregressive score model integrates time-varying direct and indirect effects; both are used to determine the appropriate transition dynamics for ρ_t.

2.3 Optimality of score updating in the time-varying spatial model

The score-driven framework may provide an intuitively and statistically appealing way to update the time-varying spatial dependence parameter ρ_t. But possibly more importantly, the score based updates have also optimal properties in an information theoretic sense under very mild regularity conditions. This was proven in a generic setting by Blasques et al. (2014a). To understand the issue for our particular time-varying spatial dependence model, we repeat the main argument of Blasques et al. (2014a) for our specific setting.

Let $p_t := p(\,\cdot\,|X_t)$ denote the true unknown conditional density of y_t. Similarly, let $\tilde{p}_t := \tilde{p}(\,\cdot\,|\tilde{f}_t, X_t)$ denote the conditional density implied by the score model given the filtered time-varying parameter $\tilde{f}_t$, the regressors X_t, the postulated innovation density p_e, and the static parameter vector $\boldsymbol{\theta}$. Ideally, whenever a new observation y_t becomes available, we want the filtered value $\tilde{f}_{t+1}$ to be such that the new conditional density implied by the model $\tilde{p}_{t+1} := \tilde{p}(\cdot|\tilde{f}_{t+1}, X_t)$ is as close as possible to the true unknown conditional density p_t from which y_t was drawn.

Following Blasques et al. (2014a), we focus on the notion of Kullback-Leibler divergence to measure the distance between the two densities

$$\mathcal{D}_{\mathrm{KL}}\big(p_t\,,\,\tilde{p}_{t+1}\big) = \int_Y p(y|X_t) \log \frac{p(y|X_t)}{\tilde{p}(y|\tilde{f}_{t+1}, X_t; \boldsymbol{\theta})}\, \mathrm{d}y, \tag{10}$$

where $Y \subseteq \mathbb{R}$ is the set over which the divergence is evaluated locally. In particular, we would like an update $\tilde{f}_{t+1}$ for which the divergence $\mathcal{D}_{\mathrm{KL}}(p(\,\cdot\,|f_t, X_t)\,,\,\tilde{p}(\,\cdot\,|\tilde{f}_{t+1}, X_t))$ is smaller than the previous divergence $\mathcal{D}_{\mathrm{KL}}(p(\,\cdot\,|f_t, X_t)\,,\,\tilde{p}(\,\cdot\,|\tilde{f}_t, X_t))$, implying that the update from $\tilde{f}_t$ to $\tilde{f}_{t+1}$ reduces the Kullback-Leibler divergence to the true unknown conditional density.

We can show that only score updates are very special in the following sense.

PROPOSITION 2.1 (Proposition 2 in Blasques et al. (2014a)). *A smooth observation driven update from $\tilde{f}_t$ to $\tilde{f}_{t+1}$ is optimal in the sense of $\mathcal{D}_{\mathrm{KL}}(p_t, \tilde{p}_{t+1}) < \mathcal{D}_{\mathrm{KL}}(p_t, \tilde{p}_t)$ for every $(y_t, \tilde{f}_t, \tilde{f}_t)$ if and only if the update is score equivalent.*

It follows that only score (equivalent) updates have the property that they always locally reduce the Kullback-Leibler divergence and thus provide a local improvement to the statistical model given the data. In particular, the spatial model structure and Student's t specification in Section 2.2 are sufficiently smooth for all local optimality results to apply. Moreover, the score driven time-varying spatial correlation model is sufficiently regular to also obtain non-local regions where the score steps ensure Kullback-Leibler improvements. We refer to Blasques et al. (2014a) for more details, optimality results, and proofs.

3 Statistical properties of the model

In this section, we establish the existence, strong consistency and asymptotic normality of the MLE of the static parameters $\boldsymbol{\theta}$ that define the stochastic properties of the spatial score model from Section 2. We first present the results in a more general setting than the spatial score model, thus extending the results in Blasques et al. (2014b) to allow for the presence of exogenous regressors. We then particularize the results to the MLE for the spatial score model in Corollary 1. All proofs of the results stated in this section can be found in the supplemental appendix.

3.1 Stochastic properties of the filtered spatial dependence parameter

To establish the consistency and asymptotic normality of the MLE, we first study the stochastic properties of the filtered parameter f_t defined through equations (5), (8), and (9). The filtered f_ts directly determine the time-varying spatial parameter $\rho_t = h(f_t)$. Understanding the properties of the filtered parameters is key to understanding the stochastic properties of the likelihood function over the parameter space $\boldsymbol{\Theta}$.

We first introduce some additional notation. Let the T-period sequences $\{y_t(\omega)\}_{t=1}^T$ and $\{X_t(\omega)\}_{t=1}^T$ be subsets of the realized path of n and k-variate stochastic sequences $\boldsymbol{y}(\omega) := \{y_t(\omega)\}_{t\in\mathbb{Z}}$ and $\boldsymbol{X}(\omega) := \{X_t(\omega)\}_{t\in\mathbb{Z}}$, for some ω in the event space Ω. In particular,[4] we let $y_t(\omega) \in \mathcal{Y} \subseteq \mathbb{R}^n$ and $X_t(\omega) \in \mathcal{X} \subseteq \mathbb{R}^k$ for all $(\omega, t) \in \Omega \times \mathbb{Z}$. For every $\omega \in \Omega$, the stochastic sequences $\boldsymbol{y}(\omega)$ and $\boldsymbol{X}(\omega)$ thus live on the spaces $(\mathcal{Y}_\infty, \mathfrak{B}(\mathcal{Y}_\infty), \mathrm{P}_0^y)$ and $(\mathcal{X}_\infty, \mathfrak{B}(\mathcal{X}_\infty), \mathrm{P}_0^X)$ where the probability measures P_0^y are P_0^X are defined over the elements of the Borel σ-algebras $\mathfrak{B}(\mathcal{Y}_\infty)$ and $\mathfrak{B}(\mathcal{X}_\infty)$. We write the filtered time-varying parameter as $\tilde{f}_t$ to distinguish it from the true time-varying parameter f_t. More precisely, we write the filtered time-varying parameter as $\{\tilde{f}_t(y^{1:t-1}, X^{1:t-1}; \boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$, which depends naturally on the initialization $\bar{f}_1 \in \mathcal{F} \subseteq \mathbb{R}$, the past data $y^{1:t-1} = \{y_s\}_{s=1}^{t-1}$ and $X^{1:t-1} = \{X_s\}_{s=1}^{t-1}$, and the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. For notational simplicity we often omit the dependence on the data and write $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ instead.

We can now rewrite the score update in (5) as

$$\tilde{f}_{t+1}(\boldsymbol{\theta}, \bar{f}_1) = \omega + A\, s\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \beta, \lambda\big) + B\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) \quad \forall\, t \in \mathbb{N},$$

---

[4]The random sequences $\boldsymbol{y}$ and $\boldsymbol{X}$ are thus $\mathfrak{F}/\mathfrak{B}(\mathcal{Y}_\infty)$ and $\mathfrak{F}/\mathfrak{B}(\mathcal{X}_\infty)$-measurable mappings $\boldsymbol{y} : \Omega \to \mathcal{Y}_\infty \subseteq \mathbb{R}_\infty^n$ and $\boldsymbol{X} : \Omega \to \mathcal{X}_\infty \subseteq \mathbb{R}_\infty^k$ where $\mathbb{R}_\infty^n := \times_{t=-\infty}^{t=\infty} \mathbb{R}^n$ and $\mathbb{R}_\infty^k := \times_{t=-\infty}^{t=\infty} \mathbb{R}^k$ denote Cartesian products of infinite copies of $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively, and $\mathcal{Y}_\infty = \times_{t=-\infty}^{t=\infty} \mathcal{Y}$ and $\mathcal{X}_\infty = \times_{t=-\infty}^{t=\infty} \mathcal{X}$ with $\mathfrak{B}(\mathcal{Y}_\infty) \equiv \mathfrak{B}(\mathbb{R}_\infty^n) \cap \mathcal{Y}_\infty$ and $\mathfrak{B}(\mathcal{X}_\infty) \equiv \mathfrak{B}(\mathbb{R}_\infty^k) \cap \mathcal{X}_\infty$; see (Billingsley, 1995, p.159). Here, $\mathfrak{B}(\mathbb{R}_\infty^n)$ and $\mathfrak{B}(\mathbb{R}_\infty^k)$ denote the Borel σ-algebras generated by the finite dimensional product cylinders of $\mathbb{R}_\infty^n$ and $\mathbb{R}_\infty^k$ respectively, $\mathfrak{F}$ denotes a σ-field defined on the event space Ω, and together with the probability measure P_0 on $\mathfrak{F}$, the triplet $(\Omega, \mathfrak{F}, \mathrm{P}_0)$ denotes the common underlying complete probability space of interest.

where $s(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \beta, \lambda)$ denotes the unit scaled score function. To shorten the notation, we define the random function

$$
\begin{aligned}
\phi_t\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1); \boldsymbol{\theta}\big) \quad &:= \quad \phi\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \boldsymbol{\theta}\big) \\
&:= \quad \omega + A\, s(f_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \beta, \lambda) + B f_t(\boldsymbol{\theta}, \bar{f}_1),
\end{aligned}
$$

as well as the supremum of its derivative,

$$
\bar{\phi}'_t(\boldsymbol{\theta}) := \sup_{f \in \mathcal{F}} \left| A\, \frac{\partial s(f, y_t, X_t; \beta, \lambda)}{\partial f} + B \right|. \tag{11}
$$

Note that $\bar{\phi}_t(\boldsymbol{\theta})$ is also a random variable due to its dependence on (y_t, X_t).

The following theorem states sufficient conditions for the stochastic sequence $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ initialized at $\bar{f}_1 \in \mathcal{F}$ to converge almost surely, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and exponentially fast to a limit stationary and ergodic (SE) sequence $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ that has N_f bounded moments. We repeatedly make use of this notion of uniform exponentially fast almost sure convergence (e.a.s.), which means that there exists a $\gamma > 1$ such that

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \gamma^t \left| \tilde{f}_t\big(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1\big) - \tilde{f}_t\big(y^{t-1}, X^{t-1}, \boldsymbol{\theta}\big) \right| \overset{a.s.}{\to} 0 \quad \text{as} \quad t \to \infty;
$$

see Straumann and Mikosch (2006). Note that the limit sequence starts in the infinite past and hence depends on the infinite past data $y^{t-1} := \{y_s\}_{s=-\infty}^{t-1}$ and $X^{t-1} := \{X_s\}_{s=-\infty}^{t-1}$, i.e., $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}} \equiv \{\tilde{f}_t(y^{t-1}, X^{t-1}; \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$. We thus establish the convergence of the sequence of random functions $\{\tilde{f}_t(\cdot, \bar{f}_1)\}_{t\in\mathbb{N}}$ defined on $\boldsymbol{\Theta}$ with random elements taking values in the Banach space $(\mathbb{C}(\boldsymbol{\Theta}, \mathcal{F}), \|\cdot\|_{\boldsymbol{\Theta}})$ for every $t \in \mathbb{N}$, to an SE limit $\{\tilde{f}_t(\cdot)\}_{t\in\mathbb{Z}}$ with elements taking values in $(\mathbb{C}(\boldsymbol{\Theta}), \|\cdot\|_{\boldsymbol{\Theta}})$, where $\|\cdot\|_{\boldsymbol{\Theta}}$ denotes the supremum norm on $\boldsymbol{\Theta}$. We have the following result.

THEOREM 1. *Let $\mathcal{F}$ be convex, $\boldsymbol{\Theta}$ be compact, $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ be SE, $s \in \mathbb{C}(\mathcal{F} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \times \Lambda)$ and assume there exists a non-random $\bar{f}_1 \in \mathcal{F}$ such that*

(i) $\mathbb{E}\log^+ \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} |s(\bar{f}_1, y_t, X_t; \beta, \lambda)| < \infty$;

(ii) $\mathbb{E}\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\phi}'_1(\boldsymbol{\theta}) < 0$.

Then $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to the unique limit SE process $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$.

If furthermore $\exists\, N_f \geq 1$ such that

(iii) $\mathbb{E}\sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} |s(\bar{f}_1, y_t, X_t; \beta, \lambda)|^{N_f} < \infty$;

and either

(iv) $\sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} |s(f,y,X;\beta,\lambda) - s(f',X,f;\beta,\lambda)| < |f-f'| \ \forall \ (f,f',y,X) \in \mathcal{F}\times\mathcal{F}\times$
$\mathcal{Y}\times\mathcal{X}$;

or

(iv') $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\phi}_1'(\boldsymbol{\theta})^{N_f} < 1$ *and* $\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1) \perp \bar{\phi}_t'(\boldsymbol{\theta}) \ \forall \ (t,\bar{f}_1) \in \mathbb{N}\times\mathcal{F}$, *where* $\perp$ *denotes independence;*

then both $\{\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)\}_{t\in\mathbb{N}}$ *and the limit SE process* $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *have* N_f *bounded moments, i.e.,* $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)|^{N_f} < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tilde{f}_t(\boldsymbol{\theta})|^{N_f} < \infty$.

The first claim of Theorem 1 makes use of the conditions in Bougerol (1993). Condition *(i)* requires the existence of an arbitrarily small moment for the score, and condition *(ii)* requires the spatial score update to be contracting on average. The uniqueness of the SE limit follows from Straumann and Mikosch (2006). The second claim of Theorem 1 uses stricter moment conditions and contraction conditions to obtain bounded moments of higher order for the filtered sequence. This constitutes an extension of Proposition 1 in Blasques et al. (2014b) to the spatial score setting with exogenous random variables X_t as well as vector and matrix arguments. Remark 1 below highlights that in the special case where the score is uniformly bounded, then the filter has infinitely many bounded moments under simpler conditions.

REMARK 1. *Let* $|B| < 1$. *If* $\bar{s} := \sup_{(\beta,\lambda,f,y,X)\in\mathcal{B}\times\mathcal{F}\times\mathcal{Y}\times\mathcal{X}} |s(f,y,X;\beta,\lambda)| < \infty$, *then* $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)|^{N_f} < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tilde{f}_t(\boldsymbol{\theta})|^{N_f} < \infty$ *hold for very* $N_f \geq 1$.

The proof of this statement follows immediately by noting that $|\tilde{f}_{t+1}| \leq \sum_{j=0}^{t-2} |B|^j(|\omega| + |A|\,\bar{s}) + |B^{t-1}\bar{f}_1| < \infty$.

3.2 Asymptotic properties of the maximum likelihood estimator

The observation-driven structure of the time-varying spatial lag model allows us to perform maximum likelihood (ML) estimation in a straightforward way. Following equation (7), we define the ML estimator (MLE) of the static parameter vector $\boldsymbol{\theta}$ as an element of the $\arg\max$ set of the sample log likelihood function $\mathcal{L}_T(\boldsymbol{\theta},\bar{f}_1)$,

$$\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \in \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}_T(\boldsymbol{\theta},\bar{f}_1), \tag{12}$$

where

$$\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) = \frac{1}{T} \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}, \bar{f}_1)$$

$$= \frac{1}{T} \sum_{t=1}^{T} \log p_e\Big(y_t - h\big((\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1))Wy_t - X_t\beta \,;\, \lambda\big) - \log|Z(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1))|.$$

with $Z(f_t)$ defined below (8).

We can now use the stationarity, ergodicity, and moment results from Theorem 1 to establish existence, consistency and asymptotic normality of the MLE. For existence, we make the following assumption.

ASSUMPTION 1. $(\boldsymbol{\Theta}, \mathfrak{B}(\boldsymbol{\Theta}))$ *is a measurable space and* $\boldsymbol{\Theta}$ *is a compact set. Furthermore,* $h :$
$\mathcal{F} \to \mathcal{F} \subseteq \mathbb{R}$ *and* $p_e : \mathbb{R}^n \times \Lambda \to \mathbb{R}$ *are continuously differentiable in their arguments.*

In Section 2, we have opted for the unit scaling of the score in our model. We can easily generalize all results below to the case of a non-constant scaling function S as long as we assume $S : \mathcal{F} \to \mathbb{R}$ is sufficiently smooth. Theorem 2 below establishes the existence and measurability of the MLE.

THEOREM 2. (Existence) *Let Assumption 1 hold. Then there exists a.s. an* $\mathfrak{F}/\mathfrak{B}(\boldsymbol{\Theta})$-*measurable map* $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) : \Omega \to \boldsymbol{\Theta}$ *satisfying (12) for all* $T \in \mathbb{N}$ *and every initialization* $\bar{f}_1 \in \mathcal{F}$.

To obtain consistency of the MLE, we impose conditions that ensure that the likelihood function satisfies a uniform law of large numbers for SE processes. We first ensures that the filter $\tilde{f}(\boldsymbol{\theta}, \bar{f}_1)$ is SE and has N_f bounded moments by application of Theorem 1.

ASSUMPTION 2. $\exists (N_f, f) \in [1, \infty) \times \mathcal{F}$ *and a* $\boldsymbol{\Theta} \subset \mathbb{R}^{3+d_\lambda}$ *such that*

(i) $\sup_{(\beta,\lambda) \in \mathcal{B} \times \Lambda} \mathbb{E}|s(f, y_t, X_t; \beta, \lambda)|^{N_f} < \infty,$

and either

(ii) $\sup_{(f,y,X,\beta,\lambda) \in \mathbb{R} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \times \Lambda} |B + A \, \partial s(f, y, X; \beta, \lambda)/\partial f| < 1,$

or

(ii′) $\mathbb{E} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \bar{\phi}'_{t,N_f}(\boldsymbol{\theta}) = \mathbb{E} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sup_f |B + A \, \partial s(f, y_t, X_t; \beta, \lambda)/\partial f| < 1$
   and $\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}, \bar{f}_1) \perp \bar{\phi}'_{t+1,N_f}(\boldsymbol{\theta}) \,\forall\, (t, \bar{f}_1) \in \mathbb{N} \times \mathcal{F}.$

Next, we ensure a bounded expectation for the likelihood function. To do this, we use the notion of 'moment preserving map'; see Blasques et al. (2014b)for a detailed description of the moment preserving properties of a wide catalogue of functions. This allows us to derive the appropriate number of bounded moments of the likelihood function from the moments of its arguments

DEFINITION 1. (Moment Preserving Maps) *A function $H : \mathbb{R}^{k_1} \times \Theta \to \mathbb{R}^{k_2}$ is said to be n/m-moment preserving, denoted as $H \in \mathbb{M}_{\Theta}(n, m)$, if and only if $\mathbb{E} \sup_{\theta \in \Theta} |\mathbf{x}_t(\theta)|^n < \infty$ implies $\mathbb{E} \sup_{\theta \in \Theta} |H(\mathbf{x}_t(\theta); \theta)|^m < \infty.$*[5]

ASSUMPTION 3. *$N_{\ell} = \min\{N_{\log p_e}, N_{\log |Z|}\} \geq 1$, where $\log |Z| \in \mathbb{M}_{\Theta}(N_f, N_{\log |Z|})$ and $\log p_e \in \mathbb{M}_{\Theta}(N, N_{\log p_e})$, with $N = \min\{N_y, N_x\}$, where N_y and N_x denote the moments of y_t and X_t, respectively.*

The moment N_{ℓ} in Assumption 3 corresponds to the number of moments of the likelihood function. Rather than assuming $N_{\ell} \geq 1$ as a high-level assumption, we follow Blasques et al. (2014b) and define N_{ℓ} as a function of the score model constituents directly, thus obtaining a set of low-level conditions for strong consistency. The requirements imposed in Assumption 3 follow easily by application of a generalized Holder inequality to the likelihood expression below (12). Note that $N = \min\{N_y, N_x\}$ follows directly by the fact that the argument $(y_t - h(\tilde{f}_t(\theta, \bar{f}_1)W y_t - X_t \beta)$ of p_e is linear in both y_t and X_t, and $\sup_{f \in \mathcal{F}} |h(f)| \leq 1$. The current conditions extend those of Blasques et al. (2014b) by accounting for the presence of exogenous variables X_t in the model.

Theorem 3 now establishes the strong consistency of the MLE for the parameters of our time-varying spatial score model if the data are SE.

THEOREM 3. (Consistency) *Let $\{y_t\}_{t \in \mathbb{Z}}$ and $\{X_t\}_{t \in \mathbb{Z}}$ be SE sequences satisfying $\mathbb{E}|y_t|^{N_y} < \infty$ and $\mathbb{E}|X_t|^{N_x} < \infty$ for some $N_y > 0$ and $N_x > 0$ and let Assumptions 1, 2, and 3 hold. Furthermore, let $\theta_0 \in \Theta$ be the unique maximizer of $\mathcal{L}_{\infty}(\theta)$ on the parameter space Θ. Then the MLE satisfies $\hat{\theta}_T(\bar{f}_1) \overset{a.s.}{\to} \theta_0$ as $T \to \infty$ for every $\bar{f}_1 \in \mathcal{F}$.*

Remark 2 below highlights that if the score s is uniformly bounded, we can change Assumption 2 in line with Remark 1.

REMARK 2. *We can substitute Assumption 2 in Theorem 3 by*

(i) $\sup_{(\beta, \lambda, f, y, X) \in \mathcal{B} \times \Lambda \times \mathcal{F} \times \mathcal{Y} \times \mathcal{X}} |s(f, y, X; \beta, \lambda)| < \infty$;

(ii) $\mathbb{E} \log \sup_{\theta \in \Theta} \bar{\phi}'_{1,1}(\theta) < 0$ *and* $|B| < 1$.

Finally, we establish the asymptotic normality of the MLE. For this, we require the existence of a sufficient number of bounded moments for the likelihood function and its derivatives.

---

[5]The $(k_1 \times 1)$-vector $\mathbf{x}_t$ satisfies $\mathbb{E} \sup_{\theta \in \Theta} |\mathbf{x}_t(\theta)|^n < \infty$ if its elements $x_{i,t}(\theta)$ satisfy $\mathbb{E} \sup_{\theta \in \Theta} |x_{i,t}(\theta)|^n < \infty$, $i = 1, ..., k_1$. The same element-wise definition applies when $\mathbf{x}_t(\theta)$ is a matrix.

For notational simplicity, we define the function $q_t := q(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \beta, \lambda) := \log p_e(y_t - h(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)) W y_t - X_t \beta; \lambda)$, as well as the cross-derivatives

$$s^{(K_1,K_2,K_3)}(f, y, X; \beta, \lambda) := \frac{\partial^{K_1+K_2+K_3} s(f, y, X; \beta, \lambda)}{\partial f^{K_1} \partial \beta^{K_2} \partial \lambda^{K_3}}.$$

The (cross)-derivatives $q^{(K_1,K_2,K_3)}$ and $(\log|Z|)^{(K_1)}$ are defined similarly. Assumption 4 now imposes sufficient moment conditions for the asymptotic normality of the MLE.

ASSUMPTION 4.   *(i)* $s^{(\boldsymbol{K})} \in \mathbb{M}_{\boldsymbol{\Theta}}(\boldsymbol{N}, N_s^{(\boldsymbol{K})})$, $q^{(\boldsymbol{K}')} \in \mathbb{M}_{\boldsymbol{\Theta}}(N, N_q^{(\boldsymbol{K}')})$, $\boldsymbol{N} := (N_f, N_y, N_x)$,
with N as defined in Assumption 3;

(ii) $N_{\ell'} \geq 2$, $N_{\ell''} \geq 1$, $N_f^{(1)} > 0$, and $N_f^{(2)} > 0$, with

$$
\begin{aligned}
N_{\ell'} &= \min\left\{ N_q^{(0,1,0)}, N_q^{(0,0,1)}, \frac{N_{\log|Z|}^{(1)} N_f^{(1)}}{N_{\log|Z|}^{(1)} + N_f^{(1)}}, \frac{N_q^{(1,0,0)} N_f^{(1)}}{N_q^{(1,0,0)} + N_f^{(1)}} \right\}, \\
N_{\ell''} &= \min\left\{ N_q^{(0,2,0)}, N_q^{(0,0,2)}, N_q^{(0,1,1)}, \frac{N_q^{(1,1,0)} N_f^{(1)}}{N_q^{(1,1,0)} + N_f^{(1)}}, \frac{N_q^{(1,0,1)} N_f^{(1)}}{N_q^{(1,0,1)} + N_f^{(1)}}, \right. \\
&\qquad \left. \frac{N_q^{(2,0,0)} N_f^{(1)}}{2N_q^{(2,0,0)} + N_f^{(1)}}, \frac{N_q^{(1,0,0)} N_f^{(2)}}{N_q^{(1,0,0)} + N_f^{(2)}}, \frac{N_{\log|Z|}^{(1)} N_f^{(2)}}{N_{\log|Z|}^{(1)} + N_f^{(2)}}, \frac{N_{\log|Z|}^{(2)} N_f^{(1)}}{2N_{\log|Z|}^{(2)} + N_f^{(1)}} \right\}, \\
N_f^{(1)} &= \min\left\{ N_f, N_s, N_s^{(0,1,0)}, N_s^{(0,0,1)} \right\}, \\
N_f^{(2)} &= \min\left\{ N_f^{(1)}, N_s^{(0,1,0)}, N_s^{(0,0,1)}, N_s^{(0,2,0)}, N_s^{(0,0,2)}, N_s^{(0,1,1)}, \right. \\
&\qquad \left. \frac{N_s^{(1,0,0)} N_f^{(1)}}{N_s^{(1,0,0)} + N_f^{(1)}}, \frac{N_s^{(2,0,0)} N_f^{(1)}}{2N_s^{(2,0,0)} + N_f^{(1)}}, \frac{N_s^{(1,1,0)} N_f^{(1)}}{N_s^{(1,1,0)} + N_f^{(1)}}, \frac{N_s^{(1,0,1)} N_f^{(1)}}{N_s^{(1,0,1)} + N_f^{(1)}} \right\}.
\end{aligned}
$$

Rather than assuming $N_{\ell'} \geq 2$ and $N_{\ell''} \geq 1$ directly as a high-level condition, we define $N_{\ell'}$ and $N_{\ell''}$ explicitly in terms of their lower-level constituents. The moment conditions in Assumption 4 extend those of Blasques et al. (2014b) by allowing for exogenous regressors. The expressions may seem complicated at first, but we show below that their verification is often straightforward; see also Blasques et al. (2014b) for the verification of similar moment conditions in a wide range of observation-driven models.

The quantities $N_f^{(1)}$ and $N_f^{(2)}$ in Assumption 4 correspond to the moments of the first and second derivatives of the filter $\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)$ with respect to the parameter $\boldsymbol{\theta}$. Similarly, $N_{\ell'}$ and $N_{\ell''}$ denote the moments of the first and second derivatives of the likelihood function, respectively.

Theorem 4 now establishes the asymptotic normality of the MLE. Here, $\text{int}(\boldsymbol{\Theta})$ denotes the interior of $\boldsymbol{\Theta}$.

THEOREM 4. (Asymptotic Normality) *Let $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ be SE sequences that satisfy* $\mathbb{E}|y_t|^{N_y} < \infty$ *and* $\mathbb{E}|X_t|^{N_x} < \infty$ *for some* $N_y > 0$ *and* $N_x > 0$ *and let Assumptions 1–4 hold. Furthermore, let* $\boldsymbol{\theta}_0 \in \mathrm{int}(\boldsymbol{\Theta})$ *be the unique maximizer of* $\mathcal{L}_\infty(\boldsymbol{\theta})$ *on* $\boldsymbol{\Theta}$. *Then,*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\bar{f}_1) - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}\big(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\mathcal{J}(\boldsymbol{\theta}_0)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big) \ \ as \ \ T \to \infty,$$

where $\mathcal{J}(\boldsymbol{\theta}_0) := \mathbb{E}\tilde{\ell}'_t(\boldsymbol{\theta}_0)\tilde{\ell}'_t(\boldsymbol{\theta}_0)^\top$ *is the expected outer product of gradients and* $\mathcal{I}(\boldsymbol{\theta}_0) := \mathbb{E}\tilde{\ell}''_t(\boldsymbol{\theta}_0)$ *is the Fisher information matrix.*

Next we apply the theory developed above to consider the properties of the MLE for the spatial score model from Section 2.2. We consider the model in (4) with Student's t distributed innovations with $\lambda > 0$ degrees of freedom. Consider a transformation function h that is (a.s.) bounded away from minus one and one with uniformly bounded derivatives $h^{(i)}$,

$$-1 < \underline{\rho} \leq \rho_t = h(f_t) \leq \bar{\rho} < 1 \quad \text{a.s.}; \qquad \sup_{f\in\mathcal{F}} |h^{(i)}(f)| < \infty, \quad i = 1, 2. \tag{13}$$

For example, to restrict the correlation to the interval $(-\bar{\rho}, \bar{\rho})$, we can take $h(f_t) = \bar{\rho} \tanh(f_t)$, where $\bar{\rho}$ can be arbitrarily close to one. We have the following corollary.

COROLLARY 1. *Consider the spatial score model with link function (13). If $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ are SE with $\mathbb{E}|y_t| < \infty$ and $\mathbb{E}|X_t| < \infty$, then there exists a compact parameter space $\boldsymbol{\Theta}$ with $|B| < 1 \ \forall \ \boldsymbol{\theta} \in \boldsymbol{\Theta}$, such that the MLE exists (a.s.) and is strongly consistent for any initialization $\bar{f}_1 \in \mathcal{F}$. If $\mathbb{E}|y_t|^{2+\epsilon} < \infty$ and $\mathbb{E}|X_t|^{2+\epsilon} < \infty$ for some $\epsilon > 0$, then the MLE is asymptotically normal with covariance matrix given in Theorem 4.*

The corollary is a direct consequence of the previous theorems and useful for to the spatial score model that we apply in our empirical section later on. It shows that we can use the MLE both for estimation and inference.

4   Monte Carlo study

To study the performance of the time-varying spatial score model in filtering out different dynamic patterns for the spatial dependence parameter, we conduct a simulation study. In this study, we also investigate whether the MLE is well-behaved and approximately normally distributed in larger samples as shown in Section 3.

To limit the complexity of the experiment, we consider a spatial lag model without regressors. We set the sample size to realistic values given the empirical application in Section 6. The data
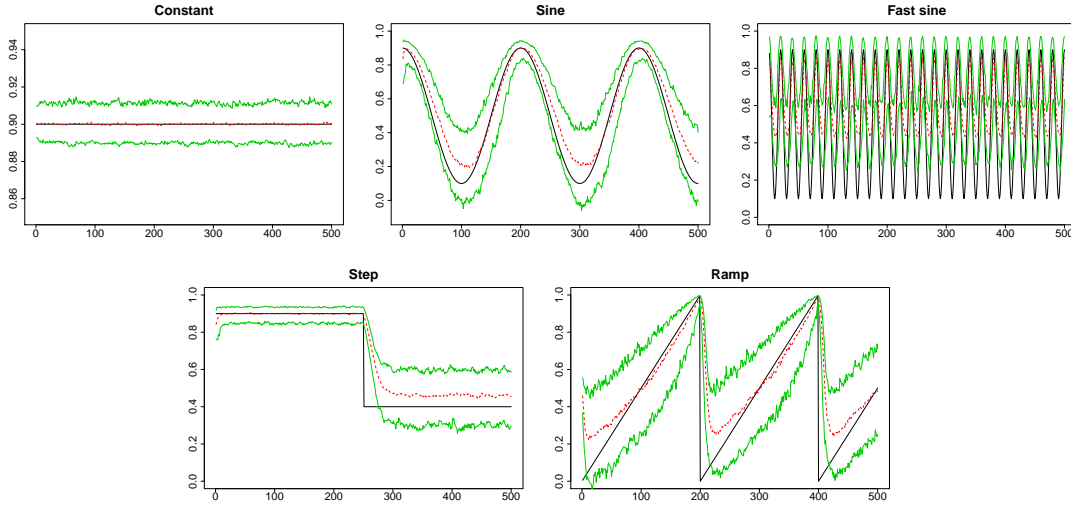
Figure 1: Simulated true spatial dependence process (black line), median filtered parameter (dashed red line) and 2.5% and 97.5% (green lines) quantiles of the filtered parameters. The figures are based on 250 replications.

generating process is

$$y_t = Z(f_t)e_t, \quad e_t \overset{i.i.d.}{\sim} \text{Student's } t(0, I_n; 5), \tag{14}$$

where $Z(f_t) = (I_n - \tanh(f_t)W)^{-1}$, $t = 1, ..., T$ and with cross-sectional dimension $n = 9$. The spatial weight matrix W is specified similar to the row-normalized cross-border exposures of the financial sectors of European countries as used in our empirical application. We simulate 250 data sets according to (14) using five processes with different dynamic patterns for the spatial dependence parameter. These patterns are similar to the ones in Engle (2002).[6]

Figure 1 shows that the filtered spatial dependence parameters are able to capture the patterns of the simulated processes quite accurately. At the low extremes of each path for ρ_t there is some over-smoothing compared to the high extremes, but this is intuitively plausible: the signal present in strongly cross-sectionally correlated data y_t is much more apparent than that in weakly correlated data.

In our second simulation study, we again use $n = 9$ cross-sectional units. We assume that the disturbances are normally distributed with common variance σ^2, and we include one regressor variable $X_t \sim N(0, I_9)$. The data-generating process is the Gaussian spatial score model laid out in Section 2. In contrast to our previous experiment, the model is now correctly specified. We simulate 500 paths y_t using the parameters $\omega = 0.05$, $A = 0.05$, $B = 0.8$, $\beta = 1.5$, and $\sigma^2 = 2$.

---

[6]In particular, we consider a constant ($\rho_t = 0.9$); sine ($\rho_t = 0.5 + 0.4\cos(2\pi t/200)$); fast sine ($\rho_t = 0.5 + 0.4\cos(2\pi t/20)$); step function ($\rho_t = 0.9 - 0.5 * I(t > T/2)$); and ramp ($\rho_t = \mod(t/200)$).
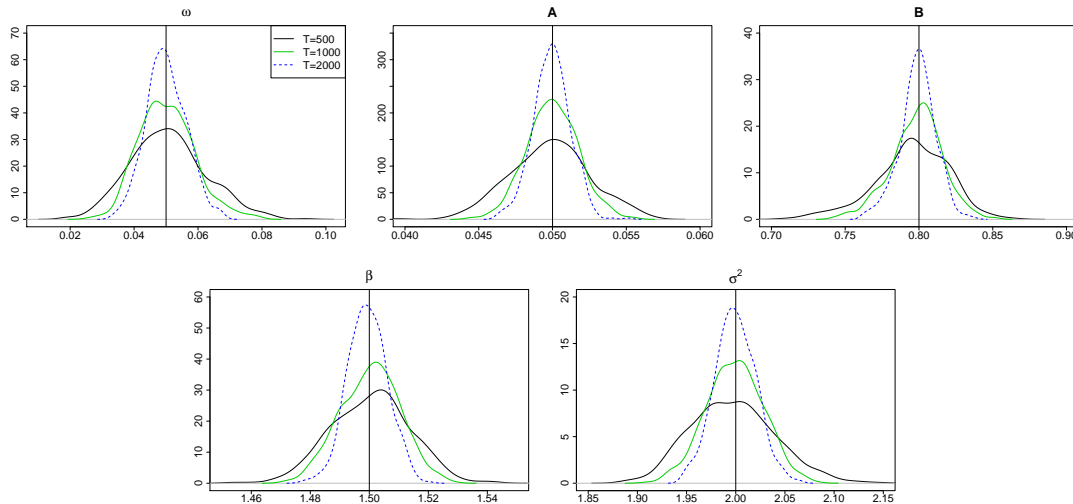
Figure 2: Kernel density estimates of estimated parameters from 500 simulations for 3 sample sizes ($T = 500, 1000, 2000$), vertical lines indicating the true parameter value

We plot the kernel density estimates of the distribution of the MLE for three different sample sizes, $T = \{ 500, 1000, 2000\}$, in Figure 2.

The figure clearly shows that for smaller sample sizes of around $T = 500$, the estimators are still not perfectly normal. For larger sample sizes, however, we see a clear convergence to the limiting result. In particular, for empirically relevant sample sizes of around $T = 2,000$ given our empirical application in the next section, all distributions look close to a normal centered around the true parameter values. We therefore apply the MLE and its associated standard errors in our empirical application below.

5 Data

In our empirical study we evaluate the evolution of perceived sovereign credit risk over a period that includes the Eurozone sovereign debt crisis. In particular, we investigate the time-varying features of the spatial dependence structure between the changes in sovereign credit default swap (CDS) spreads, particularly in relation to a number of the policy responses by regulators. Our spatial structure is directly linked to the bank sectors' cross-exposures to other sovereigns and financial sectors within the European Union.

5.1 Credit default spread data

Since EU countries have been affected by the crisis to different degrees, sovereign credit spreads in Europe are strongly cross-sectionally dependent. Figure 3 shows the credit default swap spreads
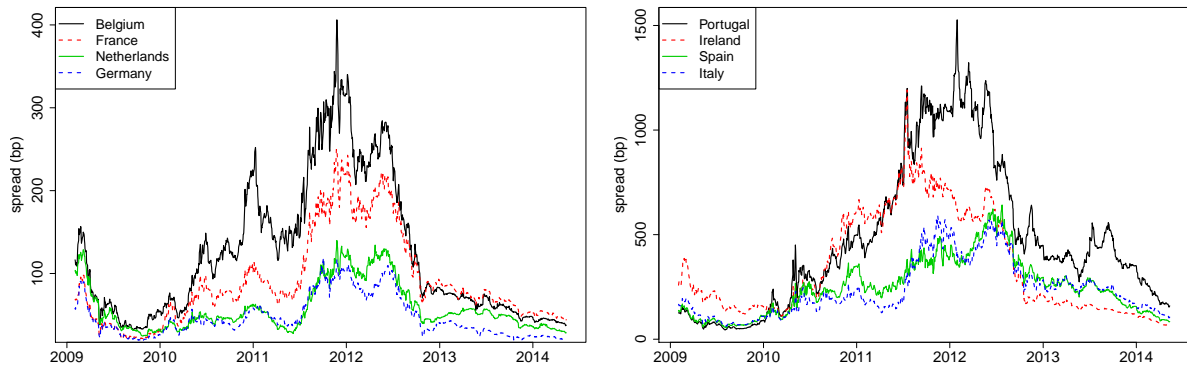
Figure 3: Credit default swap spreads of eight European sovereigns, Feb 2, 2009 – May 12, 2014. The different countries are split in two groups.

from February 2, 2009, until May 12, 2014 (1375 daily observations) for the eight euro area countries in our sample: Belgium, France, Germany, Ireland, Italy, the Netherlands, Portugal, and Spain. As in Acharya et al. (2013), we use relative changes (log differences multiplied by 100) of U.S. Dollar-denominated sovereign CDS spreads for each of these countries using data obtained from Bloomberg.

The time series reveal clear common patterns, particularly among the non-stressed Eurozone countries (Germany, France, Netherlands, Belgium, and to a lesser extend Spain and Italy). At the same time, there are temporary dissimilarities: for example, the evolution of the Ireland credit spread appears to be 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. The combination of commonalities with possible temporary changes in commonality warrants the use of the time-varying spatial score model proposed in this paper.

5.2 Other explanatory variables

Our empirical model contains two regressors that capture the state of European financial markets; see also Caporin et al. (2013). The first variable is the change in the volatility index VStoxx. The VStoxx is measured using the implied volatility of the EuroStoxx 50 and captures changes in risk appetite. Our second variable is the difference between the three month Euribor and the overnight rate EONIA. This measure captures financial sector stress and the perceived counterparty credit risk between banks.

We also incorporate two country-specific regressors, namely the (log) returns of the main stock

18

Table 1: List of country-specific stock indices included in the time-varying spatial score model as regressor variables.

Belgium	BEL 20 Price Index	France	CAC 40 Price Index
Germany	DAX 30 Price Index	Ireland	ISEQ 20 Price Index
Italy	FTSE MIB Price Index	Netherlands	AEX Price Index
Portugal	PSI 20 Price Index	Spain	IBEX 35 Price Index

index in each of the respective countries, and the absolute changes in 10-year government bond yields. We list the local stock indices in Table 1. 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. We expect a negative relation with credit spread changes. The changes in 10-year yields mainly reflect the long-term borrowing costs of governments, and we expect a positive relation with sovereign credit default swap spreads.

All variables are included in the model with a lag of one period. The data are obtained from Datastream. Augmented Dickey-Fuller unit root test statistics indicate that all time series are stationary. Table 2 presents the summary statistics.

5.3 Spatial weights matrix

The choice of the spatial weight matrix is a key ingredient of the spatial model, as it determines the structure of the 'economic distance' between the sovereign CDS spread changes and defines the channel for cross-sectional spillovers. Recently, domestic banks' cross-border exposures have been identified as relevant pricing factors for sovereign credit spreads, see for example Kallestrup et al. (2013), Korte and Steffen (2013), and Beetsma et al. (2012). A possible reason for this connection is outlined in Korte and Steffen (2013). They argue that until recently, risk management rules for banks implied a so-called 'zero risk weight channel': European banks were not required to hold capital buffers against EU member states' debt. This led to regulatory arbitrage incentives for banks to hold more government debt; see also Acharya and Steffen (2013). At the same time and due to the banks' willingness to take on government debt, governments were able to issue large amounts of debt, thus creating a potentially problematic feedback loop: if sovereign credit risk materialized, banks could become stressed, and due to possible bail-outs, governments in turn might become stressed as well.

To account for this type of possible feedback loop, we use a weight matrix that is constructed from cross-border debt data provided by the Bank for International Settlements (BIS).[7] We average

---

[7]The data can be found at http://www.bis.org/statistics/consstats.htm, Table 9B: International bank claims, consoli-

Table 2: Data summary. Stock index log returns are calculated from closing prices. All stock indices are quoted in domestic currency (Euro).

	mean	min.	25% quant.	median	75% quant.	max.
CDS spread changes (log changes*100)						
Belgium	-0.08	-19.34	-1.09	-0.07	1.78	17.04
France	-0.03	-19.44	-1.84	-0.07	1.56	19.82
Germany	-0.07	-26.71	-1.89	0.00	1.56	25.43
Ireland	-0.11	-32.69	-1.57	-0.03	1.32	26.81
Italy	-0.03	-43.73	-2.09	-0.10	1.76	20.27
Netherlands	-0.09	-22.20	-1.66	-0.03	1.39	14.92
Portugal	0.02	-47.38	-1.80	0.00	1.66	20.54
Spain	-0.04	-37.04	-2.02	0.00	1.99	25.17
local stock index returns (log returns*100)						
Belgium	0.04	-5.49	-0.59	0.03	0.69	8.96
France	0.03	-5.63	-0.68	0.02	0.80	9.22
Germany	0.06	-5.99	-0.57	0.07	0.75	5.90
Ireland	0.06	-6.79	-0.62	0.02	0.83	6.95
Italy	0.01	-7.04	-0.88	0.04	1.03	10.68
Netherlands	0.04	-5.34	-0.58	0.04	0.71	7.07
Portugal	0.01	-5.51	-0.69	0.02	0.77	10.20
Spain	0.02	-6.87	-0.82	0.01	0.87	13.48
local long-term bond yields (changes)						
Belgium	-0.16	-30.2	-2.6	-0.10	2.2	34.4
France	-0.14	-26.2	-2.6	-0.12	2.4	24.2
Germany	-0.13	-25.6	-2.8	-0.10	2.3	18.5
Ireland	-0.20	-102.8	-3.6	-0.25	2.8	75.0
Italy	-0.11	-78.0	-3.3	-0.09	3.1	50.9
Netherlands	-0.16	-22.4	-2.8	-0.05	2.1	15.6
Portugal	-0.07	-147.0	-5.1	-0.01	5.1	168.6
Spain	-0.11	-88.3	-3.6	0.00	3.5	37.3
Eurozone-wide variables						
VStoxx change	-0.02	-10.94	-0.86	-0.11	0.67	12.79
term spread	0.35	-0.37	0.14	0.34	0.52	1

the bilateral raw exposure data from 2007 Q4 - 2008 Q2. As the consolidated data are published on the BIS homepage with a lag of approximately two quarters, this avoids a possible source of endogeneity for W. The resulting weight matrix is denoted as W_{raw}.

Due to large differences in the sizes of the member countries' financial sectors, the weights implied by W_{raw} vary significantly. To mitigate the size of these differences, we form three discrete categories of mutual lending ('low', 'medium', and 'high'). The entries of the resulting

---

dated - immediate borrower basis. Last accessed on March 20, 2014.

matrix W_{cat} are constructed as

$$
W_{cat,ij} = \begin{cases}
1, & \text{if } 0 < W_{raw,ij} \leq Q_{0.33}(W_{raw}), \\
2, & \text{if } Q_{0.33}(W_{raw}) \leq W_{raw,ij} < Q_{0.67}(W_{raw}), \\
3, & \text{if } Q_{0.67}(W_{raw}) \leq W_{raw,ij},
\end{cases}
$$

where $Q_p(W_{raw})$ denotes the p-th quantile of the exposure data contained in W_{raw}. After constructing W_{cat}, we row-normalize it to obtain proper weights that sum to one. An advantage of the categorical matrix over the raw matrix is that the categories are almost time-invariant, so that using a constant W can be justified. In her spatial model for banking sector interconnections, Tonzer (2013) uses a similar data set, and averages the entries in W_{raw} over her sample period. Another alternative would be to normalize the exposure data by the GDP of the country in order to relate the size of multual debt to the size of the economy. We investigate this and other alternatives for constructing the weight matrix in our robustness checks in Section 6.2.

6 Results

6.1 Main results

Table 3 contains the estimation results for both the static spatial lag model and the time-varying spatial score model for normally and Student's t-distributed disturbances. For the benchmark models, we have a common, time-invariant variance. We relax this assumption in Section 6.2.

For the static model, we find strong evidence for spatial dependence, indicated by the high estimate and small standard error for ρ . Given that CDS spread changes are fat-tailed, it is not surprising to find that the model fit improves substantially for the Student's t vis-à-vis the normal distribution. The likelihood value increases by more than 1800 points upon adding a single parameter to the model, thus decreasing the AICc.

The dynamic spatial score model based on the normal distribution increases the likelihood by approximately 160 points compared to the static Gaussian model at the cost of adding two model parameters. The dynamics of the spatial dependence parameter are highly persistent with a value of B close to unity. The unconditional mean of f_t equals $\omega/(1 - B) \approx 0.8524$ with $\tanh(0.8524) \approx 0.6924$. Accounting for the fact that the expected value of $\tanh(f_t)$ is slightly larger than this due to Jensen's inequality, we see that the unconditional level for the Gaussian spatial score model is close to the static estimate of 0.7249. Similarly, the dynamic Student's t model increases the likelihood by approximately 68 points compared to its static counterpart. The

Table 3: Estimated parameters and their robust (sandwich) standard errors in parentheses, for the static spatial lag model and the time-varying spatial model, based on normally (N) and Student's t (t_λ) distributed disturbances. The maximized loglikelihood value (logL) and the Akaike information criterion, corrected for finite numbers of observations, (AICc) are also reported. Estimation period is February 2, 2009 – May 12, 2014.

	Static model		Time-varying model	
	N	t_λ	N	t_λ
ρ	0.7249	0.7146		
	(0.0071)	(0.0062)		
ω			0.0156	0.0181
			(0.0074)	(0.0192)
A			0.0144	0.0168
			(0.003)	(0.0085)
B			0.9817	0.9794
			(0.009)	(0.0219)
$\log(\sigma^2)$	1.8131	0.8392	1.8043	0.8426
	(0.0509)	(0.0444)	(0.0504)	(0.0446)
VStoxx	-0.0901	-0.0261	-0.0756	-0.0261
	(0.0473)	(0.0164)	(0.0326)	(0.0158)
term spread	0.0239	0.0320	0.1084	0.0818
	(0.1065)	(0.066)	(0.0998)	(0.0656)
local stocks	-0.2031	-0.1156	-0.1769	-0.1122
	(0.0426)	(0.0193)	(0.035)	(0.0187)
local 10Y yields	0.0256	0.0184	0.0258	0.0186
	(0.0041)	(0.0027)	(0.0039)	(0.0027)
const	-0.0137	-0.0341	-0.0660	-0.0578
	(0.0386)	(0.0216)	(0.0393)	(0.0244)
λ		2.5202		2.5649
		(0.1246)		(0.1288)
logLik	-26396.6	-24574.5	-26244.4	-24506.1
AICc	52807.3	49165.1	52507.0	49032.4

unconditional level of $\tanh(f_t)$ again lies close to its static estimate.

On the basis of the reported AICc values, the data clearly favors time variation in the spatial dependence parameter ρ_t using the Student's t distribution for both the disturbance e_t and the transition dynamics of ρ_t. The estimated degrees of freedom parameter λ for the Student's t models is around 2.5. Hence there is a substantial degree of fat-tailedness. A part of the unconditional fat-tailedness may also be due to the presence of volatility clustering. We discuss these robustness issues in more detail in Section 6.2.

The coefficients for the included regressors have the same signs throughout the four model specifications. Although the regression estimates vary somewhat, particularly between the normal and Student's t based models, the overall picture remains the same. A higher implied stock

volatility (VStoxx) correlates with lower CDS spreads. This is consistent with the phenomenon of 'flight to quality' from stocks to bonds when the price of risk increases in stock markets. A higher term spread on the interbank credit market implies a higher tendency to borrow overnight. This is correlated with higher CDS spread changes and may be a sign of a perceived bank-sovereign feedback loop: problems in the functioning of the interbank lending market may induce a fear of possible future bailouts and subsequent sovereign debt problems. Local stock market upturns have a dampening effect on sovereign credit spreads, while increases in long-term bond yields point to higher borrowing costs for governments and have a positive relation with sovereign CDS spread changes.

Figure 4 presents the evolution of the filtered spatial dependence parameter. We observe that the path of the spatial dependence coefficient corresponding to the Student's t spatial score model is more robust to outliers than its normal counterpart. This phenomenon is a common finding in the volatility literature; see for example Creal et al. (2013) and Harvey (2013). Comparing the score expressions in equations (8) and (9), it is clear that the time-varying spatial score model shares this feature. While the normal score is unbounded in the dependent variable and the regressors, the Student's t score contains a compensating effect in the denominator that leads to a down-weighting of large positive or negative observations; see the factor $\tilde{w}_t$ in (9). This leads to a different pattern between the two filtered spatial dependence series for the two distributions, particularly during mid 2010, the first half of 2012, and late 2013.

Throughout the sample period, systemic risk as captured by the spatial dependence coefficient is high, fluctuating around a value of 0.75 until the end of 2012. At that time, the level starts to decline towards a lower level of around 0.5 to 0.6. The pattern can be related to a number of important policy events during the European sovereign debt crisis.[8] Some events have a high visible impact. For example, the first Long Term Refinancing Operation (LTRO) at the end of 2012 caused a sudden and sharp drop in the spatial dependence parameter. The effect, however, was short-lived and the value of ρ_t bounced back soon after to similar levels as before. The second LTRO hardly has any visible effect on the spatial dependence parameter. It is only until Mario Draghi's speech at the Global Investment Conference in London in July[9] 2012 and the subsequent announcements and implementation of the Outright Monetary Transactions (OMT) and the European Stability Mechanism (ESM) in the months thereafter, that the fear of perceived spillover effects appears to be mitigated on a more permanent basis and ρ_t comes down to a lower

---

[8]A list of events can be found in Figure B.1 in the supplemental appendix. See also Table B.1 with a list of sources.

[9]Quote: "Within our mandate, the ECB is ready to do whatever it takes to preserve the euro. And believe me, it will be enough." Source: see Table B.1.
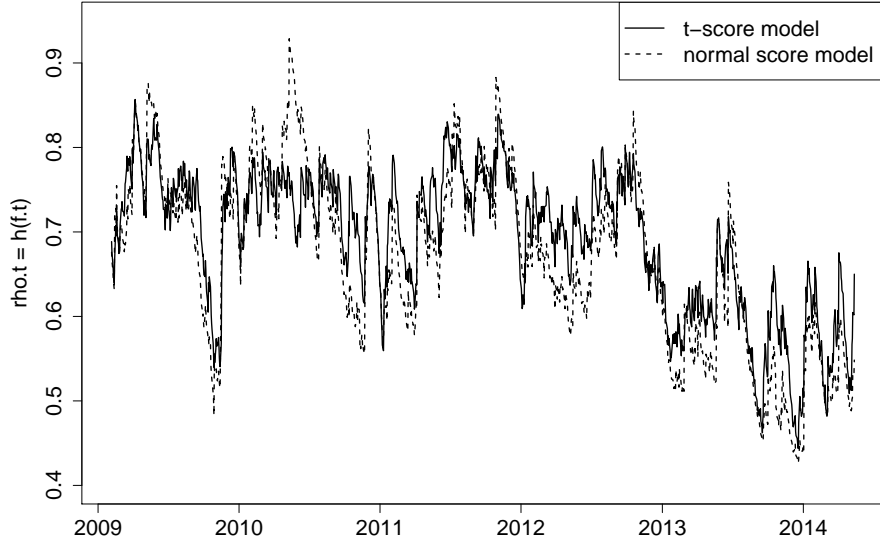
Figure 4: Filtered spatial dependence parameters obtained by imposing normally (dashed line) and Student's t (solid line) distributed disturbances.

level on a more permanent basis.

6.2 Extensions

In this section, we extend the time-varying spatial score model in different directions to investigate the robustness of our results. First, we allow for sovereign-specific volatility clustering. Second, we add an unobserved mean factor to try to distinguish common effects from spatial effects. Third, we re-estimate the models using different choices of spatial weight matrices.

Unobserved time-varying volatility factors

Given the patterns in the data, it is clearly unrealistic to assume a common, time-invariant variance for all sovereign CDS spread changes. We therefore extend the baseline model by adding a time-varying diagonal covariance matrix Σ_t for the disturbances in the spatial model,

$$y_t = h(f_t)Wy_t + X_t\beta + e_t \qquad e_t \sim p_e(0, \Sigma_t), \quad \text{with} \tag{15}$$

$$\Sigma_t := \Sigma(\boldsymbol{f}_t^\sigma) = \text{diag}\left(\sigma_1^2(f_{1,t}^\sigma), \ldots, \sigma_n^2(f_{n,t}^\sigma)\right) = \text{diag}\left(\exp(f_{1,t}^\sigma), \ldots, \exp(f_{n,t}^\sigma)\right), \tag{16}$$

where $\boldsymbol{f}_t^\sigma = (f_{1,t}^\sigma, \ldots, f_{n,t}^\sigma)'$ is a vector of sovereign-specific variance factors. As before, we endow the factors $\boldsymbol{f}_{j,t}^\sigma$ with score updating dynamics. To enforce parsimony, we allow for sovereign-

24

specific intercepts in the score updating equations for $\boldsymbol{f}_t^\sigma$, but impose common score sensitivity and persistence parameters A^σ and B^σ, so $f_{j,t+1}^\sigma = \omega_j^\sigma + A^\sigma \, s_{j,t}^\sigma + B^\sigma \, f_{j,t}^\sigma$; see Appendix A for further details. Although the covariance matrix of the disturbance vector Σ_t is diagonal, the reduced form covariance matrix of y_t is still a full (time-varying) matrix $\mathrm{Cov}(y_t) = Z(f_t)\Sigma_t Z(f_t)'$.

Unobserved time-varying mean factor

To distinguish commonalities from spatial spill-overs, we also extend the model with an additional unobserved time-varying mean factor. This factor is independent of the spatial lag structure,

$$y_t = h(f_t)Wy_t + X_t\beta + Z(f_t)^{-1}\boldsymbol{\lambda}f_t^\lambda + e_t, \qquad e_t \sim t_{\lambda_0}(0, \Sigma_t) \tag{17}$$

where λ_0 is the degrees of freedom parameter of the Student's t distribution, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$ is an $(n \times 1)$-vector of factor loadings, and $f_t^\lambda \in \mathbb{R}$ is an additional time-varying parameter endowed with score updating. Explicit formulas for the dynamics are given in Appendix A. Rewriting equation (17) in reduced form, we obtain

$$y_t = \boldsymbol{\lambda}f_t^\lambda + Z(f_t)X_t\beta + Z(f_t)e_t, \tag{18}$$

which allows for a direct comparison with the benchmark model without spatial lag structure,

$$y_t = X_t\beta + \boldsymbol{\lambda}f_t^\lambda + e_t. \tag{19}$$

Table 4 compares the goodness of fit of the seven empirical model specifications we consider in our analysis. Each extension improves the performance of the model. The model without any spatial structure performs worst, despite featuring an unobserved time-varying mean and time-varying volatilities. We therefore conclude that explicitly accounting for dynamic contemporaneous spillovers of shocks, as is done by the time-varying spatial score model, is an important feature when analyzing the dynamics of sovereign credit spread data.

The parameter estimates from the full model with spatial score updating, time-varying variances, and unobserved time-varying mean factor are given in Table 5. In contrast to the spatial factor, the variance factors and particularly the mean factor are less persistent, which is seen by the values of B^σ and B^λ, respectively. This is off-set by a larger impact of the scores in the transition equations; see the values of A^σ and A^λ.

None of the parameters λ_i, $i = 1, \dots, n$, corresponding to the mean factor are individually

Table 4: Comparison of goodness of fit of all empirical specifications considered. The largest loglikelihood value (logL) and smallest Akaike Information Criterion (AICc) are bolded.

$e_t \sim$	Static spatial		Time-varying spatial	
	$N(0, \sigma^2 \mathrm{I}_n)$	$t_\lambda(0, \sigma^2 \mathrm{I}_n)$	$N(0, \sigma^2 \mathrm{I}_n)$	$t_\lambda(0, \sigma^2 \mathrm{I}_n)$
logL	-26396.63	-24574.48	-26244.45	-24506.11
AICc	52807.35	49165.06	52507.03	49032.39

	Time-varying spatial-t		Benchmark-t
	(+tv. volas)	(+mean f.+tv.volas)	(+mean f.+tv.volas)
logL	-24175.70	**-24156.96**	-26936.15
AICc	48389.97	**48375.30**	53927.42

significantly different from zero. Jointly, however, these parameters improve the model fit, as is indicated by the AICc in Table 4. Also, the loading estimates have an economic interpretation: the non-stressed Eurozone countries have a negative coefficient λ_i, while the most stressed countries during part of the European sovereign debt crisis (Portugal, Ireland, Spain) have positive loadings.

With respect to the dynamic spatial dependence, the qualitative implications of the full model and the basic time-varying spatial t-model are very similar. This is shown in Figure 5. Omitting the additional variance and mean dynamics leads to a slight upward adjustment in the filtered spatial dependence parameter, but the overall pattern does not change.

Results from standard residual diagnostic tests are given in Table 6. The full model substantially reduces auto-correlations and ARCH effects for most individual series.Furthermore, cross-correlations are, on average, much lower for the model residuals than for the raw data. The full correlation matrices are provided in the supplemental appendix. The supplemental appendix also contains further robustness results using absolute instead of relative CDS spread changes as a dependent variable. Apart from an overall lower level of spatial dependence and a more clearly visible impact of the financial crisis at the beginning of the sample, the pattern for the spatial dependence parameter is similar to that obtained using log changes.

Choice of the spatial weight matrix

So far, all results reported were obtained using the categorical spatial weight matrix W_{cat} described in Section 5.3. As a final robustness check, we re-estimate the model using different choices for W: a matrix containing the averaged raw exposure data (W_{raw}), a model in which the matrix of exposure data is updated quarterly (W_{dyn}), and a binary matrix indicating the geographical

Table 5: Estimated parameters and their numerically approximated (sandwich-)standard errors in parentheses, for the full model featuring spatial score updating, time-varying sovereign-specific variances, an unobserved mean factor, and t-distributed disturbances. The maximized loglikelihood value (logL) and the Akaike information criterion (AICc) are also reported. Estimation period is February 2, 2009 - May 12, 2014.

ω^λ		-0.0012	ω_1^σ	Belgium	0.0426	ω	0.0307
		(0.0252)			(0.0125)		(0.0229)
A^λ		0.3494	ω_2^σ	France	0.0448	A	0.0190
		(0.8937)			(0.0142)		(0.0070)
B^λ		0.6891	ω_3^σ	Germany	0.0573	B	0.9636
		(0.1065)			(0.0155)		(0.0271)
λ_1	Belgium	-0.2776	ω_4^σ	Ireland	0.0301	const.	-0.0621
		(0.2308)			(0.0100)		(0.0240)
λ_2	France	-0.2846	ω_5^σ	Italy	0.0471	VStoxx	-0.0257
		(0.3137)			(0.0136)		(0.0157)
λ_3	Germany	-0.2029	ω_6^σ	Netherlands	0.0443	term sp.	0.0693
		(0.2811)			(0.0132)		(0.0705)
λ_4	Ireland	0.4050	ω_7^σ	Portugal	0.0524	stocks	-0.1020
		(0.6928)			(0.0153)		(0.0183)
λ_5	Italy	-0.1604	ω_8^σ	Spain	0.0591	yields	0.0173
		(0.2429)			(0.0160)		(0.0026)
λ_6	Netherlands	-0.1891	A^σ		0.1826	λ_0	3.1357
		(0.2519)			(0.0230)		(0.1977)
λ_7	Portugal	0.4614	B^σ		0.9479		
		(0.8334)			(0.0135)		
λ_8	Spain	0.0988				logLik	-24156.96
		(0.3635)				AICc	48375.30

Table 6: Diagnostic tests for the residuals of the full model featuring a spatial updating factor, volatilities, and an additional mean factor, all driven by dynamic score updating, compared to the raw CDS spread changes. LB refers to the Ljung-Box test for residual serial correlation, ARCH LM refers to the test for remaining auto-correlation in the squared residuals. The right-hand panel contains averages of pairwise cross-correlations.

sovereign	LB test stat.		ARCH LM test stat.		average cross-corr.	
	raw	residuals	raw	residuals	raw	residuals
Belgium	108.64	15.93	169.91	25.53	0.70	0.07
France	49.48	30.42	160.44	43.32	0.66	-0.01
Germany	62.61	19.49	142.70	53.78	0.63	-0.07
Ireland	129.89	17.53	302.23	87.11	0.64	-0.07
Italy	99.02	42.43	102.13	150.88	0.71	0.08
Netherlands	55.69	33.29	124.41	20.96	0.64	-0.05
Portugal	167.91	32.56	189.35	56.89	0.65	0.03
Spain	105.81	48.88	253.68	154.42	0.69	0.06

Figure 5: Filtered spatial dependence parameters obtained from the basic time-varying spatial score model with t-distributed disturbances (green) as well as with sovereign-specific, dynamic variances and an unobserved mean factor (red).
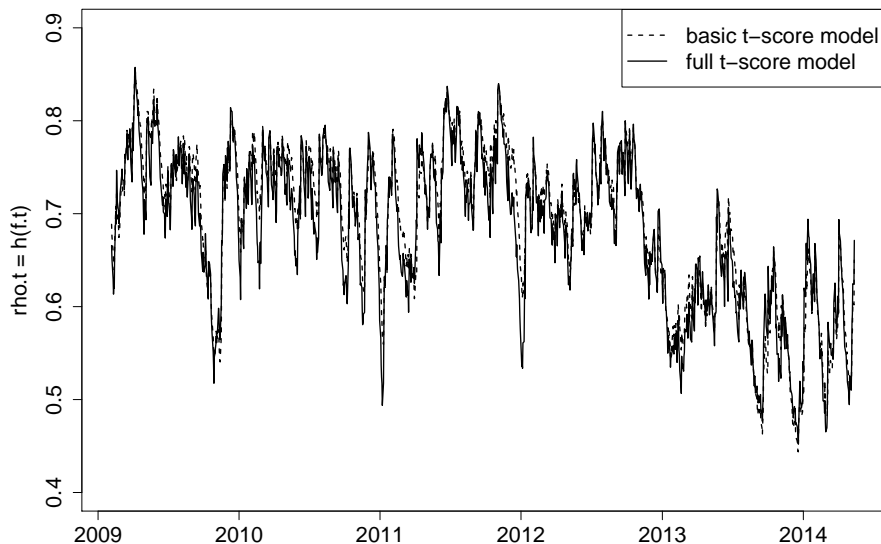

Table 7: Comparison of likelihood values for the time-varying spatial score model with Student's t disturnbances using different spatial weights matrices.

	W_{raw}	W_{dyn}	W_{cat}	W_{geo}
logL	-24745.56	-24679.44	**-24506.11**	-25556.85

neighborhood of the countries in our sample (W_{geo}). We also experimented with a weights matrix in which we weighted the raw exposures of the financial markets by the countries' respective GDP. However, this did not improve the model's fit. As the different models all have the same number of parameters, we can simply compare the likelihood values at the optimum.

Table 7 shows that the goodness of fit is quite different between the different specifications. The model with a categorical weights matrix provides the best fit. Despite the differences in fit, however, the parameter estimates and the dynamics of the spatial dependence parameter are very robust towards the specifiation of W, and none of the qualitative implications of our model change.

It is particularly interesting to see that the weight matrices based on economic distances as measured through financial cross-exposures (W_{raw}, W_{dyn}, and W_{cat}) provide a much better fit than a matrix based on geographic distances (W_{geo}). Some categorization is needed as well in order to make the sizes of cross-exposures comparable. However, as mentioned before, scaling

the exposures by the size of the economy (as measured by GDP) did not provide an improvement in terms of model fit.

7 Conclusion

In this paper, we propose a new model for time-varying spatial dependence in panel data sets. The model extends the widely used spatial lag model to a time-varying parameter framework by endowing the spacial dependence parameter with generalized autoregressive score dynamics and fat tails. Allowing for time-variation is particularly useful if we apply spatial models over longer time periods, where we can no longer be sure that the spatial dependence parameter is constant. The fat-tailed feature of our model is useful in a setting where we apply the model to financial data, which typically exhibit fatter tails than the normal.

We established the theoretical properties of our new model: the dynamics of the model are optimal in the sense that with each update step they locally reduce the Kullback-Leibler distance of the statistical model to the true unknown conditional density. Moreover, we established conditions for model invertibility and for consistency and asymptotically normality of the maximum likelihood estimator in this model.

In our empirical study based on our time-varying spatial score model, we showed that European sovereign CDS spread changes exhibit a strong, time-varying degree of spatial dependence. Cross-border debt linkages appear as a suitable transmission channel for the spatial spillovers. In our final model, we incorporated a time-varying common mean factor as well as time-varying volatilities into the specification. Using the filtered time-varying parameters of this final model, we found evidence for a break in spatial dependence towards the end of 2012. This illustrates that policies by regulators have at least been partly effective in breaking the high spill-over effects prevalent during the height of the European sovereign debt crisis.

References

Acharya, V., Drechsler, I., and Schnabl, P. (2013). A pyrrhic victory? bank bailouts and sovereign credit risk. *Journal of Finance, forthcoming*.

Acharya, V. V. and Steffen, S. (2013). The "greatest" carry trade ever? understanding eurozone bank risks. *Working Paper 19039, National Bureau of Economic Research*.

Ait-Sahalia, Y., Laeven, R. J. A., and Pelizzon, L. (2014). Mutual excitation in european sovereign cds. *Journal of Econometrics, forthcoming*.

Ang, A. and Longstaff, F. A. (2013). Systemic sovereign risk: Lessons from the u.s. and europe. *Journal of Monetary Economics*, 60:493–510.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer.

Arezki, R., Candelon, B., and Sy, A. N. R. (2011). Sovereign rating news and financial markets spillovers: Evidence from the european debt crisis. *IMF Working Paper WP/11/68*.

Arnold, M., Stahlberg, S., and Wied, D. (2013). Modeling different kinds of spatial dependence in stock returns. *Empirical Economics*, 44:761–774.

Asgharian, H., Hess, W., and Liu, L. (2013). A spatial analysis of international stock market linkages. *Journal of Banking and Finance*, 37:4738–4754.

Beetsma, R., Giuliodori, M., de Jong, F., and Widijanto, D. (2012). Spread the news: The impact of news on the european sovereign bond markets during the crisis. *Journal of International Money and Finance*, 34:83–101.

Billingsley, P. (1995). *Probability and Measure*. Wiley-Interscience.

Blasques, F., Koopman, S. J., and Lucas, A. (2014a). Information theoretic optimality of observation driven time series models. *Tinbergen Institute Discussion Papers 14-046/III*.

Blasques, F., Koopman, S. J., and Lucas, A. (2014b). Maximum likelihood estimation for generalized autoregressive score models. *Tinbergen Institute Discussion Papers 14-029/III*.

Bougerol, P. (1993). *Kalman Filtering with Random Coefficients and Contractions*. Prépublications de l'Institut Elie Cartan. Univ. de Nancy.

Caporin, M., Pelizzon, L., Ravazzolo, F., and Rigobon, R. (2013). Measuring souvereign contagion in europe. *NBER Working Paper No. 18741*.

Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and 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*, 28:777–795.

Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014). Observation driven mixed-measurement dynamic factor models. *Review of Economics and Statistics*, page forthcoming.

De Santis, R. A. (2012). The euro area sovereign debt crisis - safe haven, credit rating agencies and the spread of the fever from greece, ireland and portugal. *ECB Working Paper No. 1419*.

Denbee, E., Julliard, C., Li, Y., and Yuan, K. (2013). Network risk and key players: A structural analysis of interbank liquidity. *Working Paper*.

Engle, R. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20:339–350.

Fernandez, V. (2011). Spatial linkages in international financial markets. *Quantitative Finance*, 11:237–245.

Gorea, D. and Radev, D. (2013). The euro area sovereign debt crisis: Can contagion spread from the periphery to the core? *Gutenberg School of Management and Economics Discussion Paper No. 1208*.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Econometric Society Monographs, Cambridge University Press.

Harvey, A. C. and Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, page forthcoming.

Hillier, G. and Martellosio, F. (2013). Properties of the maximum likelihood estimator in spatial autoregressive models. *cemmap working paper CWP44/13*.

Kalbaska, A. and Gatkowski, M. (2012). Eurozone sovereign contagion: Evidence from the cds market (20052010). *Journal of Economic Behavior & Organization*, 83:657–673.

Kallestrup, R., Lando, D., and Murgoci, A. (2013). Financial sector linkages and the dynamics of bank and sovereign credit spreads. *Working paper, available at SSRN: http://ssrn.com/abstract=2023635*.

Keiler, S. and Eder, A. (2013). Cds spreads and systemic risk - a spatial econometric approach. *Deutsche Bundesbank Discussion Paper*, No. 1/2013.

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

Korte, J. and Steffen, S. (2013). Zero risk contagion - banks' sovereign exposure and sovereign risk spillovers. *Mimeo*.

Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72:1899–1925.

Lee, L. and Yu, J. (2010). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, 40:255–271.

LeSage, J. P. and Pace, R. K. (2008). *Introduction to Spatial Econometrics*. CRC Press.

Leschinski, C. and Bertram, P. (2013). Pure contagion dynamics in emu government bond spreads. *Mimeo*.

Lucas, A., Schwaab, B., and Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics*, 32(2):271–284.

Oh, D. H. and Patton, A. (2013). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Duke University Discussion Paper*.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70:120–126.

Pesaran, M. H. and Pick, A. (2007). Econometric issues in the analysis of contagion. *Journal of Economic Dynamics & Control*, 31:12451277.

Saldias, M. (2013). A market-based approach to sector risk determinants and transmission in the euro area. *ECB Working Paper Series*, No. 1574.

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

Tonzer, L. (2013). Cross-border interbank networks, banking risk and contagion. *mimeo*.

Wied, D. (2013). Cusum-type testing for changing parameters in a spatial autoregressive model for stock returns. *Journal of Time Series Analysis*, 34:221229.

Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1, 1) model. *Scandinavian Journal of Statistics*, 40(4):846–867.

Yu, J., de Jong, R., and Lee, L.-f. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics*, 146(118-134).

Appendix A Model extensions

We restrict the model extensions to the case of Student's t distributed disturbances. We obtain the equations for the Gaussian case as a special case by letting $\lambda_0 \to \infty$.

We assume that the vector of variance factors $\boldsymbol{f}_t^\sigma$ in (16) follows an n-dimensional score process as given by

$$\boldsymbol{f}_{t+1}^\sigma = \boldsymbol{\omega}^\sigma + A^\sigma \nabla_t^\sigma + B^\sigma \boldsymbol{f}_t^\sigma$$

with $\boldsymbol{\omega} = (\omega_1^\sigma, \ldots, \omega_n^\sigma)'$, and $A^\sigma, B^\sigma \in \mathbb{R}$. We thus allow for sovereign-specific intercepts in the variance score update, but restrict the dynamic parameters A^σ and B^σ to be common across all countries. This results in a parsimonious, yet flexible model. The score of the spatial dependence factor f_t is given in (9), with Σ replaced by Σ_t. For the variance factors, the score vector is

$$\nabla_t^\sigma = \frac{\partial \ell_t}{\partial \boldsymbol{f}_t^\sigma} = \frac{1}{2} \begin{pmatrix} \frac{(1+\lambda^{-1}n)\exp(-f_{1,t}^\sigma)\cdot\left(y_{1,t}-h(f_t)\sum_{j=1}^n w_{1j}y_{j,t}-x'_{1,t}\beta\right)^2}{1+\lambda^{-1}(y_t-h(f_t)Wy_t-X_t\beta)'\Sigma(\boldsymbol{f}_t^\sigma)^{-1}(y_t-h(f_t)Wy_t-X_t\beta)} - 1 \\ \vdots \\ \frac{(1+\lambda^{-1}n)\exp(-f_{n,t}^\sigma)\cdot\left(y_{n,t}-h(f_t)\sum_{j=1}^n w_{nj}y_{j,t}-x'_{n,t}\beta\right)^2}{1+\lambda^{-1}(y_t-h(f_t)Wy_t-X_t\beta)'\Sigma(\boldsymbol{f}_t^\sigma)^{-1}(y_t-h(f_t)Wy_t-X_t\beta)} - 1 \end{pmatrix},$$

with $X_t' = (x_{1,t}, \ldots, x_{n,t})$, and $x_{i,t} \in \mathbb{R}^{k\times 1}$.

In the presence of an additional mean factor f_t^λ as in (18), the score update for f_t changes from (9) to

$$\begin{aligned} \nabla_t &= \left[\tilde{w}_t \cdot \left(Wy_t - W\lambda f_t^\lambda\right)' \Sigma^{-1} \left(y_t - h(f_t)Wy_t - X_t\beta - Z(f_t)^{-1}\boldsymbol{\lambda}f_t^\lambda\right) - \text{tr}(Z(f_t)W)\right] \cdot \dot{h}(f_t), \\ \tilde{w}_t &= \frac{(1+\lambda^{-1}n)}{1+\lambda^{-1}(y_t-h(f_t)Wy_t-X_t\beta-Z(f_t)^{-1}\boldsymbol{\lambda}f_t^\lambda)'\Sigma^{-1}(y_t-h(f_t)Wy_t-X_t\beta-Z(f_t)^{-1}\boldsymbol{\lambda}f_t^\lambda)} \; . \end{aligned} \qquad \text{(A.1)}$$

The updating equation for f_t^λ is given by

$$f_{t+1}^\lambda = \omega^\lambda + A^\lambda \nabla_t^\lambda + B^\lambda f_t^\lambda,$$

with score

$$\nabla_t^\lambda = \tilde{w}_t \cdot (Z(f_t)^{-1}\lambda)'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta - Z(f_t)^{-1}\boldsymbol{\lambda}f_t^\lambda). \qquad \text{(A.2)}$$

Finally, in the benchmark model (19), the score expression equals that in (A.2) with $W = 0$ and $Z(f_t) \equiv I_n$.