TI 2014-107/III
Tinbergen Institute Discussion Paper

Spillover Dynamics for Systemic Risk Measurement using Spatial Financial Time Series Models

Francisco Blasques
Siem Jan Koopman
Andre Lucas
Julia Schaumburg

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

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

August 2014

Abstract

We introduce a new model for time-varying spatial dependence. The model extends the well-known static spatial lag model. All parameters can be estimated conveniently by maximum likelihood. We establish the theoretical properties of the model and show that the maximum likelihood estimator for the static parameters is consistent and asymptotically normal. We also study the information theoretic optimality of the updating steps for the time-varying spatial dependence parameter. We adopt the model to empirically investigate the spatial dependence between eight European sovereign CDS spreads over the period 2009–2014, which includes the European sovereign debt crisis. We construct our spatial weight matrix using cross-border lending data and include country-specific and Europe-wide risk factors as controls. We find a high, time-varying degree of spatial spillovers in the sovereign CDS spread data. There is a downturn in spatial dependence after the first half of 2012, which is consistent with 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 score.

[1]We thank conference participants at the 2nd workshop on "Models driven by the score of predictive likelihoods" in Tenerife 2014, the 7th Annual SoFiE Conference in Toronto 2014, the International Association of Applied Econometrics conference in London 2014, the SYRTO workshop at the Bundesbank 2014 and seminar participants at VU University Amsterdam for helpful comments. Lucas and Blasques thank the Dutch Science Foundation (NWO, grant VICI453-09-005) for financial support. Koopman, Lucas, and Schaumburg thank the European Union Seventh Framework Programme (FP7-SSH/2007-2013, grant agreement 320270 - SYRTO) for financial support. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78) at Aarhus University, Denmark, and funded by the Danish National Research Foundation. Research support by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk" is gratefully acknowledged.

1 Introduction

We propose a new parsimonious model to measure the time-varying cross-sectional dependence in European sovereign credit spreads. 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.

It is important to model sovereign default risk in the Euro area using a joint model. Financial markets in the Euro area are largely integrated, and the financial sectors in the separate member countries are heavily engaged in cross-border borrowing and lending. The cross-border entanglement of financial firms and sovereigns exacerbated further during the financial crisis; see for example Acharya et al. (2013) and Dieckmann and Plank (2012). Our model accounts for the possibility that shocks affecting the credit quality of one Eurozone member are likely to be propagated to the other members via the linkages of their financial sectors. Possible feedback loops that amplify systemic risk are incorporated in this way as well. The transmission channels in our model are defined explicitly as economic distances in a spatial weights matrix constructed from cross-border debt data.

The model is adopted to empirically study a time series sample of eight Eurozone sovereign credit default swaps (CDS) over the period 2009–2014. We find strong evidence for time-varying spatial dependence. The dependence parameter displays clear but transitory downward throughs around the Long Term Refinancing Operations (LTROs) by the European Central Bank at the end 2011 and start of 2012, but effectively remains at high levels throughout up to the second half of 2012. It is only after the announcements and implementation of the Outright Monetary Transactions (OMT), and the implementation of the European Stability Mechanism (ESM) that systemic risk settles more permanently at a spatial correlation level that is roughly 30% to 40% lower than during the crisis. The empirical implications of the model are robust to a range of model extensions and alternative specifications, includuding common unobserved factors in the levels of CDS changes, alternative distributional assumptions, alternative spatial weight matrices, and time varying volatilities.

Our paper contributes to two strands of literature. First, we contribute to the applied spatial econometrics literature. Spatial models are 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. The study closest to ours is Keiler and Eder (2013). They model CDS spreads of financial institutions in a static spatial lag model, additionally accounting for firm-specific covariates and market risk factors. Their spatial weight matrix is constructed from stock return correlations.

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 new to the literature. We model the dynamics using the generalized autoregressive score 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. 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. 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.

The remainder of the paper is organized as follows. Section 2 introduces our spatial model with time-varying parameters. We examine its theoretical properties 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. We provide our study of European sovereign CDS spread dynamics in Section 5. Section 6 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

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

where y_t 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 $n \times 1$ vector e_t is the disturbance vector with multivariate density $p_e(e_t, \Sigma; \lambda)$ that has mean zero, an unknown $k \times k$ variance (or scale) matrix Σ and possibly other parameters collected in vector λ. For example, p_e may represent the Student's t distribution where λ is then the degrees of freedom parameter. This model (1) implies that each entry y_{it}, for $i = 1, \ldots, n$, of the vector y_t depends on k individual-specific regressors x_{it} as well as on possibly 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 modelling 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 a standard option to row-normalize W such that $\sum_{j=1}^{n} w_{ij} = 1$ for $i = 1, \ldots, n$, where w_{ij} is the (i, j)th element from W. The impact of the (spatially weighted) contemporaneous dependent variables Wy_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 Lee (2004).

We show that the basic form of the spatial lag model (1) can capture nonlinear feedback effects across units by rewriting the model as

$$y_t = ZX_t\beta + Ze_t, \tag{2}$$

where we assume that the inverse matrix $Z = (\mathrm{I}_n - \rho W)^{-1}$ exists with I_n as the $n \times n$ identity matrix. Using 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 shocks e_{it} or $x'_{it}\beta$ to unit i spill over to other units $j \neq i$ to an extent that depends on their relative proximity to i via the weight matrix W and the spatial dependence parameter ρ. At the same time, there are also 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 structure of (1) causes an endogeneity problem and leads to an inconsistency in the least squares estimation of the coefficients when n becomes large. An alternative solution in the cross-sectional literature is to 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 Quasi MLE (QMLE) for $n \to \infty$, and Hillier and Martellosio (2013) investigate its finite sample distribution. Large n and large T asymptotics for QMLE of the spatial model with static dependence parameter are studied in Yu et al. (2008). For further textbook treatments of spatial econometric models and their estimation, 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).

[1]Alternatively, we can use GMM as in, for example, Kelejian and Prucha (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 propose to introduce a time-varying ρ 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 $\rho_t \in (-1, 1)$. We adopt the autoregressive score framework of Creal et al. (2011, 2013) and Harvey (2013) to introduce a time-varying f_t. The score framework for time-varying parameters has been adopted successfully in a range of different model 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), the location and scale models with fat tails of Harvey and Luati (2014), and many more.[2]

The score framework is based on 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 is a scalar coefficient, A is the score coefficient, $s_t = S_t \nabla_t$ is the scaled score function and B is the autoregressive (or mean-reverting) coefficient. All three coefficients are treated as fixed and unknown parameters. 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. The score function is given by $\nabla_t = \partial \ell_t / \partial f_t$ where

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

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 $\ln |(\mathrm{I}_n - h(f_t) W)|$. This term accounts for the nonlinearity of the model in ρ_t as shown in equation (2). We define the vector of static parameters

[2] See `www.gasmodel.com` for a full compilation.

[3] In a simulation (not reported here) it turned out 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.

$\boldsymbol{\theta} = (\omega, A, B, \beta, \lambda)'$ and estimate $\boldsymbol{\theta}$ via the numerical maximization of the likelihood function given by

$$\ell_T = \sum_{t=1}^{T} \ell_t. \tag{7}$$

We consider two specifications for the error term densities p_e, namely the multivariate normal distribution and the multivariate Student's t distribution. The latter is particularly relevant for our empirical study because spread changes in credit default swaps (CDS) may be fat-tailed. Also, 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

$$\ell_t = \ln|\mathrm{I} - h(f_t)W| - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(y_t - h(f_t)Wy_t - X_t\beta)'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta), \tag{8}$$

and score

$$\nabla_t = \big(y_t'W'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta) - \mathrm{tr}(Z(f_t)W)\big) \cdot \dot{h}(f_t), \tag{9}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, $Z(f_t) = (\mathrm{I}_n - f(f_t)W)^{-1}$, 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 with λ degrees of freedom, we have the tth likelihood contribution as given by

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

with the score function given by

$$\begin{aligned}\nabla_t \;&=\; \big(\tilde{w}_t \cdot y_t'W'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta) - \mathrm{tr}(Z(f_t)W)\big) \cdot \dot{h}(f_t), \quad \text{where} \\ \tilde{w}_t \;&=\; (1 + \lambda^{-1}n)\big/\big(1 + \lambda^{-1}(y_t - h(f_t)Wy_t - X_t\beta)'\Sigma^{-1}(y_t - h(f_t)Wy_t - X_t\beta)\big).\end{aligned} \tag{10}$$

We can verify that if $\lambda \to \infty$, we have $\tilde{w}_t \to 1$ and the score expression in (10) collapses to the one in (9). 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. It 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) and

Harvey (2013). This intuition is also 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 $h(f_t)$.

The score expressions in (9) and (10) also depart from the expressions for the standard linear regression model. In particular, we have the additional correction term $-\mathrm{tr}(Z(f_t)W)$. This term accounts for the simultaneity bias in the standard least squares estimator and follows from the presence of the term $0.5 \ln |Z(f_t)|$ in the likelihood at time t. In effect, 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 effects and indirect effects; both are used to determine the appropriate transition dynamics for ρ_t.

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 choose to 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. We also discuss how the spatial score update is the optimal observation-driven parameter update in an information theoretic sense, thus providing a further theoretical backbone to the use of the score for updating f_t. All proofs of the results stated in this section can be found in the 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), (9), and (10). The filtered f_t 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 Θ.

We first need 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 $y_t(\omega) \in \mathcal{Y} \subseteq$

[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 $\boldsymbol{y} : \Omega \to \mathcal{Y}_\infty \subseteq \mathbb{R}^n_\infty$ and $\boldsymbol{X} : \Omega \to \mathcal{X}_\infty \subseteq \mathbb{R}^k_\infty$ where $\mathbb{R}^n_\infty := \times_{t=-\infty}^{t=\infty}\mathbb{R}^n$ and $\mathbb{R}^k_\infty := \times_{t=-\infty}^{t=\infty}\mathbb{R}^k$ denote Cartesian products of infinite

$\mathbb{R}^n \; \forall(\omega,t) \in \Omega \times \mathbb{Z}$ and $X_t(\omega) \in \mathcal{X} \subseteq \mathbb{R}^k \; \forall(\omega,t) \in \Omega \times \mathbb{Z}$. 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 $\boldsymbol{\theta} \in \Theta$. 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

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

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

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

as well as the supremum of its derivative,

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

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

The following theorem states sufficient conditions for the stochastic sequence $\{\tilde{f}_t(\boldsymbol{\theta}, f_1)\}_{t\in\mathbb{N}}$ initialized at $\bar{f}_1 \in \mathcal{F}$ to converge almost surely, uniformly in $\boldsymbol{\theta} \in \Theta$, 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 $\exists\, \gamma > 1$ such that

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

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 $\mathfrak{B}(\mathcal{Y}_\infty) \equiv \mathfrak{B}(\mathbb{R}^n_\infty) \cap \mathcal{Y}_\infty$ and $\mathfrak{B}(\mathcal{X}_\infty) \equiv \mathfrak{B}(\mathbb{R}^k_\infty) \cap \mathcal{X}_\infty$; see (Billingsley, 1995, p.159). Here, $\mathfrak{B}(\mathbb{R}^n_\infty)$ and $\mathfrak{B}(\mathbb{R}^k_\infty)$ denote the Borel σ-algebras generated by the finite dimensional product cylinders of $\mathbb{R}^n_\infty$ and $\mathbb{R}^k_\infty$ 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.

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 Θ with random elements taking values in the Banach space $(\mathbb{C}(\Theta, \mathcal{F}), \|\cdot\|_\Theta)$ for every $t \in \mathbb{N}$, to an SE limit $\{f_t(\cdot)\}_{t\in\mathbb{Z}}$ with elements taking values in $(\mathbb{C}(\Theta), \|\cdot\|_\Theta)$, where $\|\cdot\|_\Theta$ denotes the supremum norm on Θ. We have the following result.

THEOREM 1. *Let $\mathcal{F}$ be convex, Θ 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) $\mathbb{E}\ln^+ \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} |s(\bar{f}_1, y_t, X_t; \beta, \lambda)| < \infty$*;*

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

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′) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \bar{\phi}_1'(\boldsymbol{\theta})^{N_f} < 1$ *and* $\tilde{f}_t(\boldsymbol{\theta}, f_1) \perp \bar{\phi}_t'(\boldsymbol{\theta})\ \forall\ (t, f_1) \in \mathbb{N} \times \mathcal{F}$*, where $\perp$ denotes independence;*

then both $\{f_t(\boldsymbol{\theta}, f_1)\}_{t\in\mathbb{N}}$ and the limit SE process $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ have N_f bounded moments, i.e., $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |f_t(\boldsymbol{\theta}, f_1)|^{N_f} < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |f_t(\boldsymbol{\theta})|^{N_f} < \infty$.

The first claim of Theorem 1 makes use of the conditions in Bougerol (1993a). 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* $\sup_{(\beta,\lambda,f,y,X)\in\mathcal{B}\times\mathcal{F}\times\mathcal{Y}\times\mathcal{X}} |s(f, y, X; \beta, \lambda)| < \infty$*, then* $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |f_t(\boldsymbol{\theta}, f_1)|^{N_f} < \infty$ *holds for very $N_f \geq 1$.*

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

3.2 Asymptotic properties of the maximum likelihood estimator

The observation-driven structure of the time-varying spatial model allows for a simple implementation of a maximum likelihood (ML) estimation procedure. Following equation (7), we define the ML estimator (MLE) of the spatial score parameters more precisely as an element of the $\arg\max$ set of the sample log likelihood function $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$,

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

where

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

with $Z(f_t)$ defined below (9).

We can now use Theorem 1 to establish existence, consistency and asymptotic normality of the MLE of the static parameters in the time-varying spatial model. For existence, we make the following assumptions.

ASSUMPTION 1. $(\Theta, \mathfrak{B}(\Theta))$ *is a measurable space and* Θ *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* $\mathfrak{F}/\mathfrak{B}(\Theta)$*-measurable map* $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) : \Omega \to \Theta$ *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* $\Theta \subset \mathbb{R}^{3+d_\lambda}$ *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′) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \bar{\phi}'_{t,N_f}(\boldsymbol{\theta}) = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |B + A\,\partial s(f, y_t, X_t; \beta, \lambda)/\partial f| < 1$ *and* $f_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}, f_1) \perp \bar{\phi}'_{t+1,N_f}(\boldsymbol{\theta})\ \forall\,(t, f_1) \in \mathbb{N}\times\mathcal{F}$.

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

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* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\mathbf{x}_t(\boldsymbol{\theta})|^n < \infty$ *implies* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |H(\mathbf{x}_t(\boldsymbol{\theta}); \boldsymbol{\theta})|^m < \infty$.[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* $\log p_e \in \mathbb{M}_\Theta\big(N, N_{\log p_e}\big)$*, with* $N = \min\big\{N_y, N_x\big\}$*, 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 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 $N = \min\big\{N_y, N_x\big\}$ follows directly by the fact that the argument $(y_t - h(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)Wy_t - X_t\beta)$ 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 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.*

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

Furthermore, let $\boldsymbol{\theta}_0 \in \Theta$ be the unique maximizer of $\ell_\infty(\boldsymbol{\theta})$ on the parameter space Θ. Then the MLE satisfies $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \stackrel{a.s.}{\to} \boldsymbol{\theta}_0$ 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) $\mathbb{E}\ln\sup_{\boldsymbol{\theta}\in\Theta}\bar{\phi}'_{1,1}(\boldsymbol{\theta}) < 0$ *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. For notational simplicity, we define the function $q_t := q(\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1), y_t, X_t;\beta,\lambda) := \log p_e(y_t - h(\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)Wy_t - X_t\beta;\lambda)$, 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)* $s^{(\boldsymbol{K})} \in \mathbb{M}_\Theta(\boldsymbol{N}, N_s^{(\boldsymbol{K})})$, $q^{(\boldsymbol{K}')} \in \mathbb{M}_\Theta(N, N_q^{(\boldsymbol{K}')})$, $\boldsymbol{N} := (N_f, N_y, N_x)$, *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*

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

Again, 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, $\mathrm{int}(\Theta)$ denotes the interior of Θ.

THEOREM 4. (Asymptotic Normality) *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 1–4 hold. Furthermore, let $\boldsymbol{\theta}_0 \in \mathrm{int}(\Theta)$ be the unique maximizer of $\ell_\infty(\boldsymbol{\theta})$ on Θ. 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 time-varying spatial model. 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 set the correlation between $\underline{\rho} = -\bar{\rho}$ and $\bar{\rho}$, 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 Θ with $|B| < 1 \; \forall \; \boldsymbol{\theta} \in \Theta$, 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 is particularly applicable 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.

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

We note that the score-driven framework is not only intuitively appealing as a way to update time-varying parameters. More importantly, score based updates are also optimal in an information theoretic sense under very mild regularity conditions; see Blasques et al. (2014a).

Let $p_t := p(\cdot\,|f_t, X_t)$ denote the true unknown conditional density of y_t, which depends on the true unobserved time-varying parameter $\{f_t\}_{t\in\mathbb{Z}}$ and the regressors X_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, and the postulated innovation density p_e. To simplify the notation, note that we have dropped the dependence of $\tilde{f}_t$ on $\boldsymbol{\theta}$ and $\bar{f}_1$ 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 $\tilde{p}_{t+1} := \tilde{p}(\cdot|\tilde{f}_{t+1}, X_t)$ be 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

$$\mathcal{D}_{\mathrm{KL}}\big(p_t\,,\,\tilde{p}_{t+1}\big) = \int_Y p(y|X_t)\ln\frac{p(y|\tilde{f}_t, X_t)}{\tilde{p}(y|\tilde{f}_{t+1}, X_t;\boldsymbol{\theta})}\,\mathrm{d}y, \tag{14}$$

where $Y \subseteq \mathbb{R}$ is the set over which the divergence is evaluated; ; see Blasques et al. (2014a) for further details. In particular, we would like an update $\tilde{f}_{t+1}$ for which $\mathcal{D}_{\mathrm{KL}}(p(\cdot\,|f_t, X_t)\,,\,\tilde{p}(\cdot\,|\tilde{f}_{t+1}, X_t))$ is smaller than $\mathcal{D}_{\mathrm{KL}}(p(\cdot\,|f_t, X_t)\,,\,\tilde{p}(\cdot\,|\tilde{f}_t, X_t))$, such that the update from $\tilde{f}_t$ to $\tilde{f}_{t+1}$ reduces the distance to the true unknown conditional density.

Blasques et al. (2014a) show that only score updates have the property that they locally always reduce the KL-distance and thus provide a local improvement. Though their original proofs do not account for the presence of exogenous regressors, is it easy to see from their paper that all their results continue to hold if X_t is incorporated in the conditional densities as described above. In particular, the spatial model structure and Student's t specification are sufficiently smooth for local optimality results to apply, as well as for non-local Kullback-Leibler improvement regions from Blasques et al. (2014a) to hold.

4 Monte Carlo study

To show the adequacy of the time-varying spatial model in filtering out dynamic patterns of the spatial dependence parameters, we conduct a simulation study. In this study, we also investigate whether the MLE is well-behaved and approximately normally distributed in larger samples. We set the sample size to realistic values given the empirical application in Section 5. To limit the complexity of the experiment, we consider a spatial lag model without regressors. The data generating process is

$$y_t = Z(f_t)e_t, \quad e_t \overset{i.i.d.}{\sim} \text{Student's } t(0, \mathrm{I}_n; 5), \tag{15}$$

where $Z(f_t) = (\mathrm{I}_n - \tanh(f_t)W)^{-1}$, $t = 1, ..., 500$. The spatial weight matrix W is specified similar to the one used in our empirical application. It contains row-normalized cross-border exposures of the financial sectors of nine European countries. We simulate 250 data sets according to (15) using five processes with different dynamic patterns for the spatial dependence parameter. These patterns are similar to the ones in Engle (2002), namely

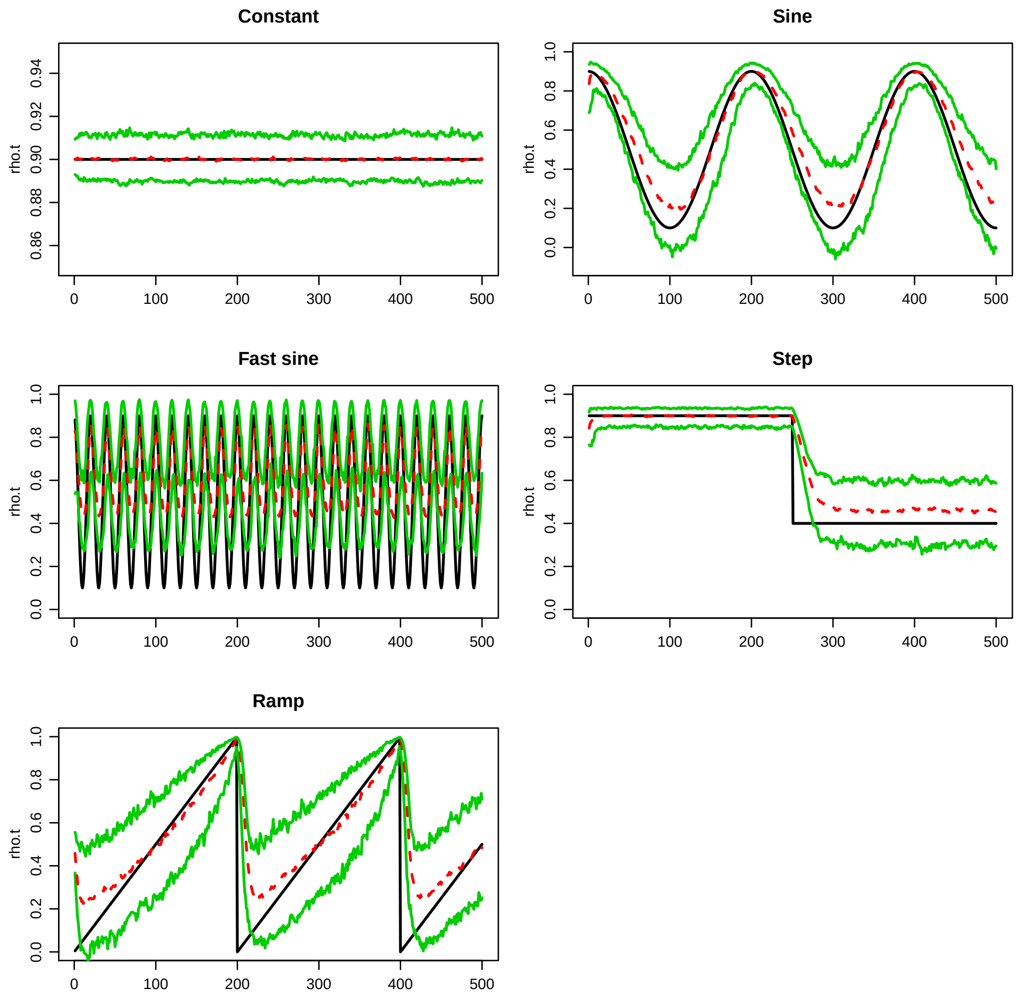
1. Constant: $\rho_t = 0.9$;
2. Sine: $\rho_t = 0.5 + 0.4\cos(2\pi t/200)$;
3. Fast sine: $\rho_t = 0.5 + 0.4\cos(2\pi t/20)$;
4. Step: $\rho_t = 0.9 - 0.5 * I(t > T/2)$;
5. Ramp: $\rho_t = \mod (t/200)$;

Figure 1 shows that the filtered spatial dependence parameters are able capture the patterns of the simulated processes quite accurately. The model has some difficulty in tracking downturns compared to upturns, but this is intuitively plausible: the signal present in strongly cross-sectionally correlated data y_t is much more apparent than that for weakly correlated data.

In our second simulation study, we again use nine cross-sectional units. We assume that the errors 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 thus correctly specified. We simulate 500 paths y_t using the parameters $\omega = 0.05$, $A = 0.05$, $B = 0.8$, $\beta = 1.5$, and $\sigma^2 = 2$. 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

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.

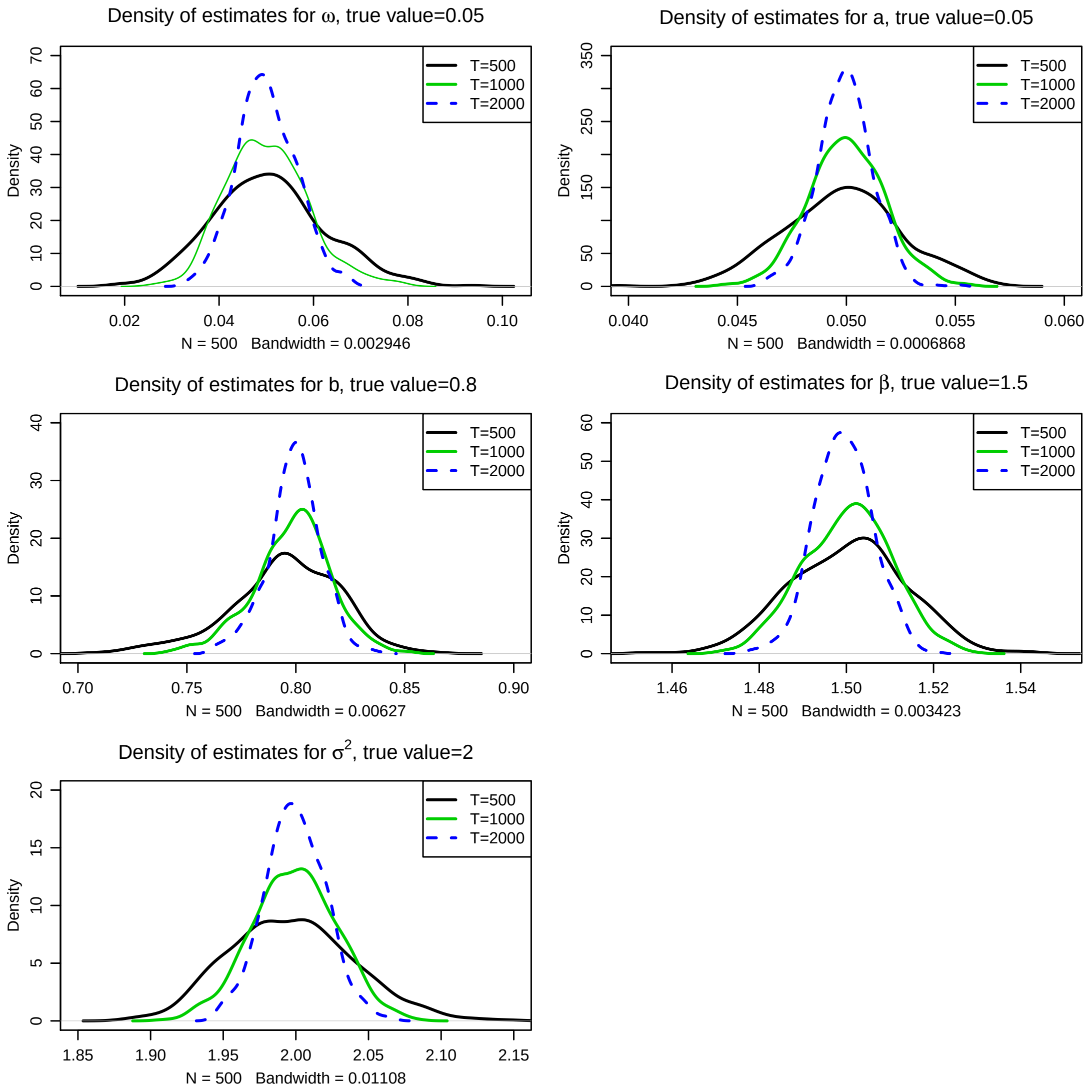


the limiting result. In particular, for empirically relevant sample sizes of around $T = 2,000$, all distributions look close to a normal centered around the true parameter values. We therefore apply the MLE and its associated standard errors in our empirical application in the next section.

5 The empirics of time-varying European CDS spread dependencies

In our empirical study we evaluate the evolution of sovereign credit risk spreads 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 sovereign credit spreads, particular 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.

Figure 2: Kernel density estimates of estimated parameters from 500 simulation replications



5.1 Data

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 spreads from February 2, 2009, until May 12, 2014 (1375 daily observations) for the nine euro area countries in our sample: Belgium, France, Germany, Ireland, Italy, the Netherlands, Portugal, and Spain. We use relative changes (log returns multiplied by 100) of Euro-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

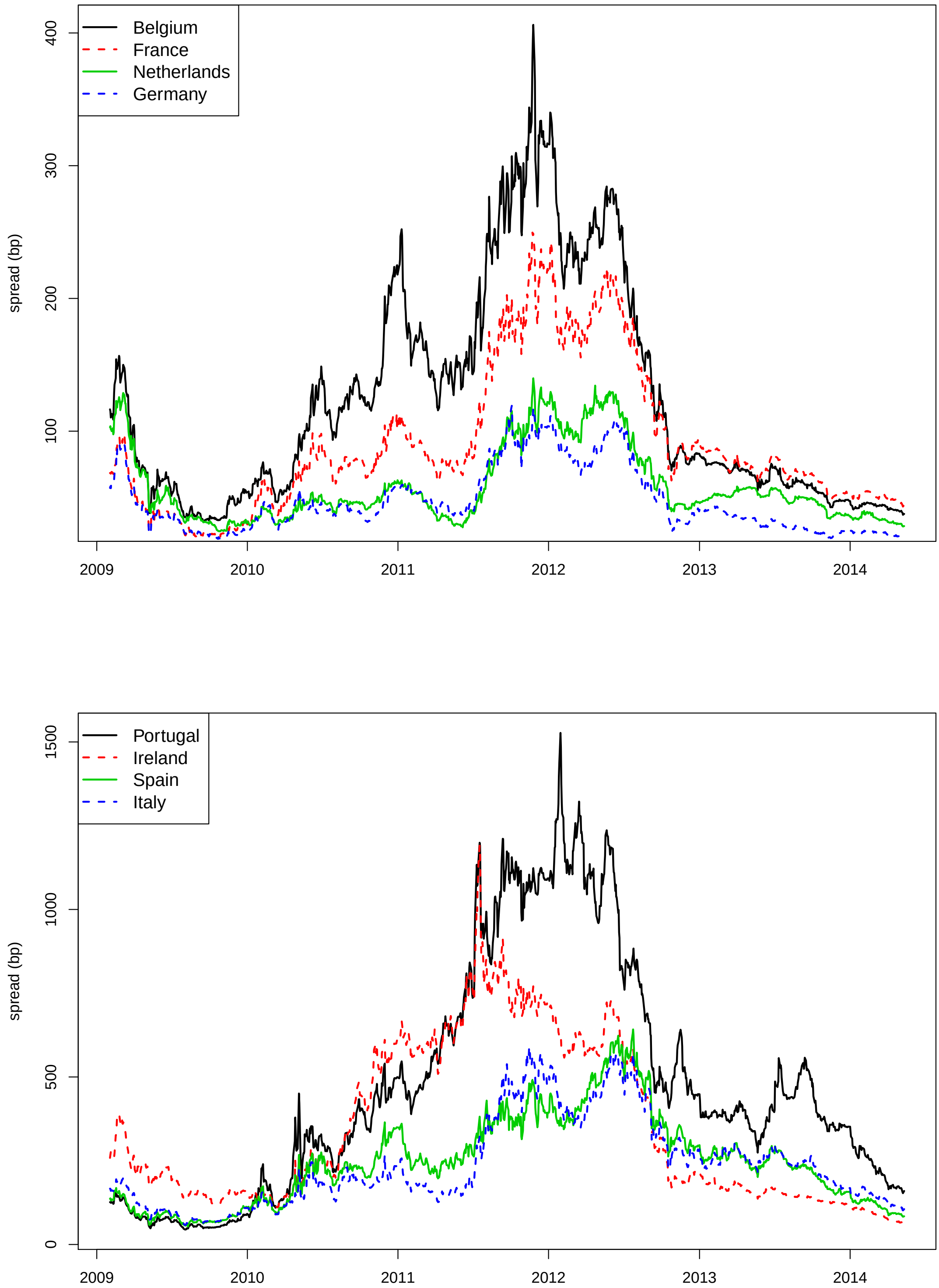


Figure 3: Credit default swap spreads of eight European sovereigns, Feb 2, 2009 – May 12, 2014. The different countries are split in two groups.

Table 1: List of country-specific stock indices included in the time-varying spatial 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

the same time, there appear to be 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 model proposed in this paper.

Other explanatory variables

Our empirical model contains two regressors that capture the state of the European financial market, 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. It captures changes in the risk appetite of financial markets. Our second variable is the difference between the three month Euribor and the overnight rate EONIA. This measure captures the stress within the financial sector and the perceived counterparty credit risk between banks.

We also incorporate two country-specific regressors, namely log returns of the respective countries' leading stock returns and absolute changes in 10-year government bond yields. Local stock market returns, see Table 1, are a measure of the well-being of the economy and the ability of governments to pay off debt in the long run. We expect a negative relation with credit spread changes. The change 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. We have computed the augmented Dickey-Fuller unit root test statistics and they indicated that all time series are stationary. Table 2 presents the summary statistics.

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	-1.9	-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	1.56	25.43
Ireland	-0.11	-32.69	-1.57	-0.03	1.32	26.81
Italy	-0.03	-43.73	-2.09	-0.1	1.76	20.27
Netherlands	-0.09	-22.2	-1.66	-0.03	1.39	14.92
Portugal	0.02	-47.38	-1.8	0	1.66	20.54
Spain	-0.04	-37.04	-2.02	0	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	0.8	9.22
Germany	0.06	-5.99	-0.57	0.07	0.75	5.9
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	10.2
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	-0.1	2.2	34.4
France	-0.14	-26.2	-2.56	-0.12	2.4	24.2
Germany	-0.13	-25.6	-2.8	-0.1	2.3	18.48
Ireland	-0.2	-102.79	-3.64	-0.25	2.8	75
Italy	-0.11	-78	-3.3	-0.09	3.1	50.9
Netherlands	-0.16	-22.4	-2.8	-0.05	2.14	15.61
Portugal	-0.07	-146.98	-5.1	-0.01	5.13	168.6
Spain	-0.11	-88.3	-3.6	0	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

Spatial weights matrix

The choice of spatial weight matrix is a key ingredient of the spatial model, as it determines the structure of the 'economic neighborhood' between the sovereign CDS spreads 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 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).[6] We average 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. This matrix is denoted by 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 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. An alternative way would be to normalize the exposure data by 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 5.3.

5.2 Results

Table 3 contains the estimation results from both the static and time-varying spatial model for normally and t-distributed error terms. For simplicity, the specifications contain a common, time-invariant variance. This assumption is relaxed in the robustness section, Section 5.3. For the static model, we find strong evidence for spatial dependence in the European sovereign CDS spreads,

[6]The data can be found at http://www.bis.org/statistics/consstats.htm, Table 9B: International bank claims, consolidated - immediate borrower basis. Last accessed on March 20, 2014.

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 errors. 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)*
$\ln(\sigma^2)$	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 sp	0.0239	0.032	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	-0.066	-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

indicated by the high estimates for ρ together with a small standard error. Given that CDS spread changes are known to exhibit fat tails, 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 after only adding 1 parameter to the model. This finding is confirmed by the AICc.

If we consider the dynamic spatial model based on the normal distribution, we see an increase of about 160 likelihood points compared to the static Gaussian model at the cost of adding 2 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. Also the increase from the static Student's t to its time-varying counterpart results in a likelihood increase, this time of about 68 points. The unconditional level of $\tanh(f_t)$ again lies close to its static counterpart.

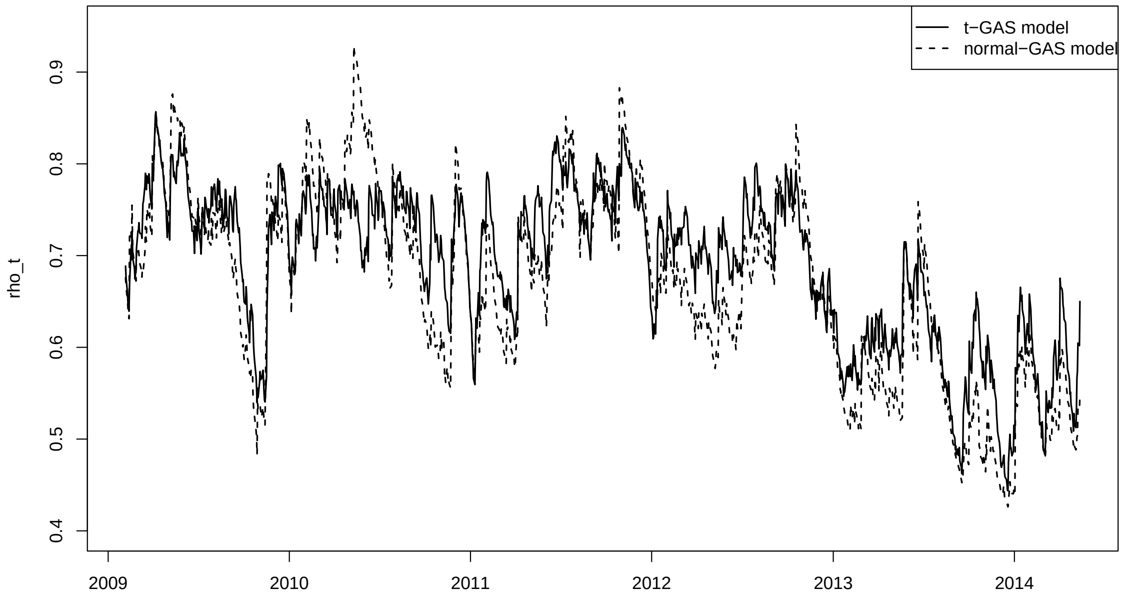
On the basis of the reported AICc values, the data clearly favors time variation in the spatial dependence parameter ρ_t using the Student's t distribution for the estimation as well as for the transition dynamics of ρ_t. The estimated degrees of freedom parameter λ for the Student's t models is around 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 5.3.

The coefficients for the included regressors have the same signs throughout the four 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 volatility on the European stock market (VStoxx) correlates with lower CDS spreads. This is consistent with the phenomenon of 'flight to quality' when the price of risk increases in financial 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. 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 spreads.

Figure 4 presents the evolution of the filtered spatial dependence parameter. We observe that the path of the spatial 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 (9) and (10), it is clear that the time-varying spatial 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 the denominator that leads to a down-weighting of large positive or negative observations. 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 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 about 0.5 to 0.6.

Figure 4: Filtered spatial dependence parameters obtained by imposing normally (dashed line) and Student's t (solid line) distributed errors.



The pattern can be related to a number of important policy events during the European sovereign debt crisis.[7] 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[8] 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, which can be seen by the value of ρ_t coming down to lower levels.

5.3 Extensions

In this section, we extend the time-varying spatial model in different directions to investigate the robustness of our results. First, we allow for sovereign-specific volatility clustering. Second, we

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

[8]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.

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 changes. We therefore extend the baseline model by adding a time-varying diagonal covariance matrix Σ_t for the errors in the spatial model,

$$y_t = h(f_t)Wy_t + X_t\beta + e_t \qquad e_t \sim p_e(0, \Sigma_t), \tag{16}$$

with

$$\Sigma_t := \Sigma(\boldsymbol{f}_t^\sigma) = \text{diag}\left(\sigma_1^2(f_t^{\sigma_1}), ..., \sigma_n^2(f_t^{\sigma_n})\right) = \text{diag}\left(\exp(f_t^{\sigma_1}), ..., \exp(f_t^{\sigma_n})\right), \tag{17}$$

where $\boldsymbol{f}_t^\sigma = (f_t^{\sigma_1}, ..., f_t^{\sigma_n})'$ is a vector of sovereign-specific variance factors. As before, we endow the factor $\boldsymbol{f}_t^\sigma$ with score updating. To enforce parsimony, we allow for sovereign-specific intercepts in the score updating equations for $\boldsymbol{f}_t^\sigma$, but impose common score and persistence parameters A^σ and B^σ; see Appendix A. Although the covariance matrix of the error terms Σ_t is diagonal, the reduced form covariance of y_t is still a full matrix $\text{Cov}(y_t) = Z(f_t)\Sigma_t Z(f_t)'$.

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,

$$y_t = h(f_t)Wy_t + X_t\beta + Z(f_t)^{-1}\boldsymbol{\lambda} f_t^\lambda + e_t, \qquad e_t \sim t_{\lambda_0}(0, \Sigma_t) \tag{18}$$

where λ_0 is the degrees of freedom parameter of the Student's t distribution, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$ 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 (18) in reduced form, we obtain

$$y_t = \boldsymbol{\lambda} f_t^\lambda + Z(f_t)X_t\beta + Z(f_t)e_t, \tag{19}$$

Table 4: Comparison of goodness of fit of all considered empirical specifications. The largest maximized loglikelihood value (logL) and the smallest Akaike Information Criterion (AICc) amongst the considered models are marked as bold.

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

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

$$y_t = X_t \beta + \boldsymbol{\lambda} f_t^\lambda + e_t. \tag{20}$$

Table 4 compares the goodness of fits of the seven empirical model specifications we consider in our analysis. Each extension improves the performance of the model. However, 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 it is done by the time-varying spatial model, is an important feature when analyzing 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, $i = 1, \ldots, n$, corresponding to the mean factor are individually significantly different from zero. Jointly, however, they improve the model fit, as is indicated by the AICc in Table 4. Furthermore, 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 regard to 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

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 error terms. 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	0.019
		(0.8937)			*(0.0142)*		*(0.007)*
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)			*(0.01)*		*(0.024)*
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	0.405	ω_7^{σ}	Portugal	0.0524	stocks	-0.102
		(0.6928)			*(0.0153)*		*(0.0183)*
λ_5	Italy	-0.1604	ω_8^{σ}	Spain	0.0591	yields	0.0173
		(0.2429)			*(0.016)*		*(0.0026)*
λ_6	Netherlands	-0.1891	A^{σ}		0.1826	λ_0	3.1357
		(0.2519)			*(0.023)*		*(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	48375.3

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 is able to substantially reduce auto-correlations and ARCH effects for most individual series.[9] Furthermore, cross-correlations are, on average, much lower for the model residuals than for the raw data. To give a full picture, we also provide the full correlation matrices in Table B.2.

To further robustify our results, we repeat the estimation, employing absolute instead of relative CDS spread changes as dependent variable. Figure B.2 shows the evolution of the corresponding filtered parameter from the full model. 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 is similar to the picture obtained by using log changes.

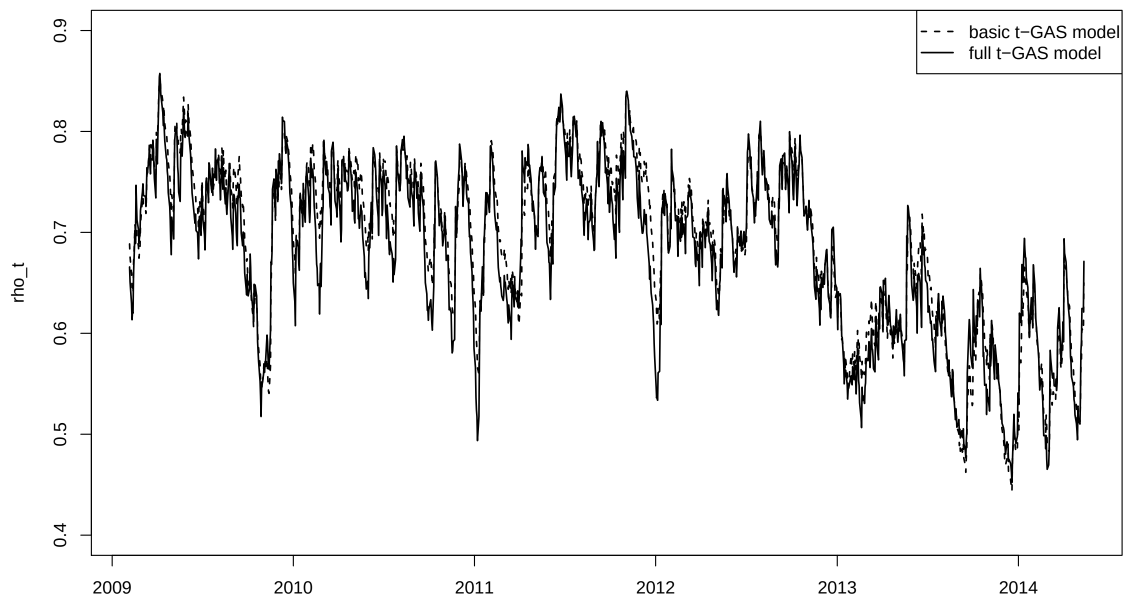
[9] Italy is the only country for which ARCH effects are not reduced by our model.

Table 6: Diagnostic tests for the residuals of the full model featuring spatial updating factor, volatilities, and additional mean factor, all driven by the score function, 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 panel contains averages of pairwise cross-correlation.

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

*Remaining effects at 5% level

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



Choice of the spatial weight matrix

So far, all results reported have been obtained using the categorical spatial weight matrix W_{cat} described in Section 5.1. In order to robustify our findings, we re-estimate the model using different

Table 7: Comparison of likelihood values for the time-varying spatial model with t-errors, using different spatial weights matrices.

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

choices of W. Candidate choices include the 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 neighborhood of the countries in our sample (W_{geo}).[10] 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 fits are quite different. The model with a categorical weights matrix provides the best fit. However, the parameter estimates are very robust towards the specification of W, and none of the qualitative implications from our model changes.

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.

6 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 they locally reduce the Kullback-Leibler distance of the statistical model

[10] We also experimented with a weights matrix in which we weighted the raw exposures of the financial markets by the countries' respective GDP. However, the fit did not improve.

to the true unknown conditional density with every score update of the spatial dependence parameter. Moreover, the maximum likelihood estimator for the model was consistent and asymptotically normal under mild regularity conditions. Also, we showed under what conditions the model is invertible, such that the filtered estimate of the time-varying spatial dependence parameters converge in the limit to a unique stationary and ergodic sequence.

In our empirical study based on our time-varying spatial model, we showed that European sovereign CDS spreads 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 to 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*.

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. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792.

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

Bougerol, P. (1993b). Kalman filtering with random coefficients and contractions. *SIAM J. Control Optim.*, 31(4):942–959.

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*.

Dieckmann, S. and Plank, T. (2012). Default risk of advanced economies: An empirical analysis of credit default swaps during the financial crisis. *Review of Finance*, 16:903–934.

Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

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.

Foland, G. B. (2009). *A Guide to Advanced Real Analysis*. Dolciani Mathematical Expositions. Cambridge University Press.

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

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*.

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

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, forthcoming*, 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.

Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Annals of Mathematical Statistics*, 33:659–680.

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.

van der Vaart, A. W. (2000). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.

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

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 errors. 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 (17) 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 (10), 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 (19), the score update for f_t changes from (10) to

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

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

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

Finally, in the benchmark model (20), the score expression equals that in (A.2) with $W = 0$ and $Z_t \equiv \mathrm{I}_n$.

Appendix B Additional tables and figures

Table B.1: Key policy events during the Eurozone crisis

Date	Event	Source
Oct. 18, 2009	Greece announces doubling of budget deficit	The Guardian[1]
Mar. 3, 2010	EU offers financial help to Greece	ECB[2]
Dec. 7, 2010	Ireland is bailed out by EU and IMF	ECB[2]
Dec. 22, 2011	ECB launches the first Longer-Term Refinancing Operation (LTRO)	ECB[2]
Mar. 1, 2012	ECB launches the second LTRO	ECB[2]
Jul. 26, 2012	M. Draghi: "[T]he ECB is ready to do whatever it takes to preserve the euro."	ECB[3]
Oct. 8, 2012	European Stability Mechanism (ESM) is inaugurated	ESM[4]
Sep. 12, 2013	European Parliament approves new unified bank supervision system	ECB[2]

[1] http://www.theguardian.com/business/2012/mar/09/greek-debt-crisis-timeline
[2] http://www.ecb.europa.eu/ecb/html/crisis.de.html
[3] http://www.ecb.europa.eu/press/key/date/2012/html/sp120726.en.html
[4] http://www.esm.europa.eu/press/releases/20121008_esm-is-inaugurated.htm
All retrieved on June 19, 2014.

Figure B.1: Filtered spatial dependence parameters obtained from the full model, together with key policy from Table B.1.

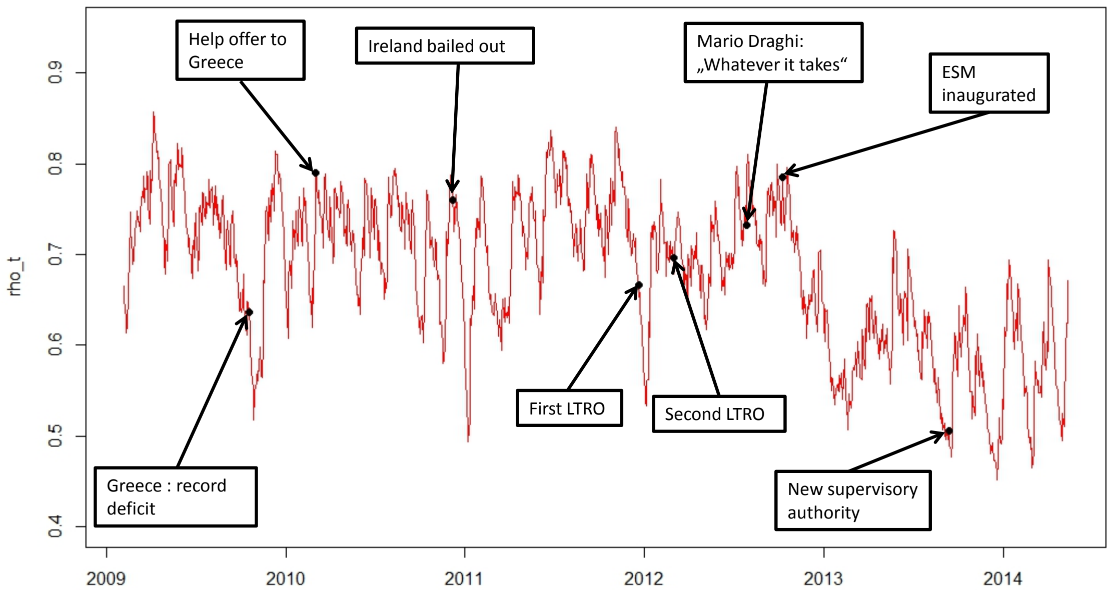


Figure B.2: Filtered spatial parameter obtained from the full time-varying spatial model with time-varying volatilities and unobserved mean factor, using absolute CDS spread changes as dependent variable.

Table B.2: Cross-correlation matrices: raw data and full model residuals

Correlation matrix of raw CDS change data

	Belgium	France	Germany	Ireland	Italy	Netherlands	Portugal	Spain
Belgium	1.000	0.724	0.697	0.630	0.738	0.737	0.643	0.707
France		1.000	0.724	0.581	0.655	0.678	0.581	0.657
Germany			1.000	0.553	0.609	0.685	0.534	0.577
Ireland				1.000	0.718	0.575	0.724	0.685
Italy					1.000	0.654	0.740	0.847
Netherlands						1.000	0.566	0.620
Portugal							1.000	0.742
Spain								1.000

Correlation matrix of residuals

	Belgium	France	Germany	Ireland	Italy	Netherlands	Portugal	Spain
Belgium	1.000	0.113	0.068	-0.114	0.186	0.079	0.03	0.145
France		1.000	0.234	-0.232	-0.076	0.029	-0.11	-0.015
Germany			1.000	-0.247	-0.231	0.099	-0.19	-0.212
Ireland				1.000	0.038	-0.115	0.205	-0.038
Italy					1.000	-0.148	0.248	0.534
Netherlands						1.000	-0.126	-0.172
Portugal							1.000	0.161
Spain								1.000

C Technical appendix: proofs

C.1 Proofs of main theorems

The lines of proof adopted here closely follow the original lines of proof in Blasques et al. (2014b), extended to the case of exogenous variables.

Proof of Theorem 1: Define the norms $\|\cdot\|^{\Theta} := \sup_{\boldsymbol{\theta}\in\Theta} |\cdot|$ and $\|\cdot\|^{\Theta}_{N_f} := \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \|\cdot\|^{N_f}$.
Following Straumann and Mikosch (2006, Proposition 3.12), we have

$$\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1) - \tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta})| \overset{e.a.s.}{\to} 0.$$

This follows directly from Bougerol (1993b, Theorem 3.1) in the context of the random sequence $\{\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\Theta})\}_{t\in\mathbb{N}}$ with elements $\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\Theta})$ taking values in the separable Banach space $\mathcal{F}_{\Theta} \subseteq (\mathbb{C}(\Theta, \mathcal{F}), \|\cdot\|_{\Theta})$, with initialization $\bar{f}_1^{\Theta}$ in $\mathbb{C}(\Theta, \mathcal{F})$, where $\bar{f}_1^{\Theta}(\boldsymbol{\theta}) = \bar{f}_1 \; \forall\, \boldsymbol{\theta} \in \Theta$, and[11]

$$\begin{aligned}
\tilde{f}_t(y^{1:t}, X^{1:t}, \cdot, \bar{f}_1^{\Theta}) &= \phi_t\big(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\Theta})\big),\\
&:= \phi\big(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\Theta})\,,\; y_t, X_t;\, \cdot\big) \;\forall\, t \in \mathbb{N},
\end{aligned}$$

where $\{\phi_t\}_{t\in\mathbb{Z}}$ is a stationary and ergodic (SE) sequence of stochastic recurrence equations $\phi_t : \Xi \times \mathbb{C}(\Theta, \mathcal{F}) \to \mathbb{C}(\Theta, \mathcal{F}) \;\forall\, t$ as in Straumann and Mikosch (2006, Proposition 3.12). Note that with a slight abuse of notation we use ϕ both to denote the functional $\phi : \mathbb{C}(\Theta, \mathcal{F}) \times \mathcal{Y} \times \mathcal{X} \to \mathbb{C}(\Theta, \mathcal{F})$ as well as the function $\phi : \mathcal{F} \times \mathcal{Y} \times \mathcal{X} \times \Theta \to \mathcal{F}$. Continuity of ϕ follows from $s \in \mathbb{C}(\mathcal{F} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \times \Lambda)$, where $\mathcal{B}$ is the domain of the regression parameters β.

The assumption that $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ are SE and the continuity of ϕ together imply that $\{\phi_t\}_{t\in\mathbb{Z}}$ is SE by Krengel (1985, Proposition 4.3). Condition C1 in Bougerol (1993b, Theorem 3.1) follows from $\mathbb{E}\ln^+ \|s(f^{\Theta}, y_t, X_t;\, \cdot\,, \cdot)\|^{\Theta} < \infty$ since, by norm sub-additivity and positive homogeneity, for any $f^{\Theta} \in \mathbb{C}(\Theta, \mathcal{F})$,

$$\begin{aligned}
\mathbb{E}\ln^+ \big\|\phi_t(f^{\Theta})\big\|^{\Theta} \le \mathbb{E}\big\|\phi_t(f^{\Theta})\big\|^{\Theta} &= \mathbb{E}\big\|\phi(f^{\Theta}, y_t, X_t;\, \cdot)\big\|^{\Theta}\\
&= \mathbb{E}\big\|\omega + A s(f^{\Theta}, y_t, X_t;\, \cdot\,, \cdot) + B f^{\Theta}\big\|^{\Theta}\\
\le \sup_{\boldsymbol{\theta}\in\Theta} |\omega| + \sup_{\boldsymbol{\theta}\in\Theta} |A|\, &\mathbb{E}\|s(f^{\Theta}, y_t, X_t;\, \cdot\,, \cdot)\|^{\Theta} + \sup_{\boldsymbol{\theta}\in\Theta} |B| \cdot \|f^{\Theta}\|^{\Theta} < \infty,
\end{aligned}$$

because $\sup_{\boldsymbol{\theta}\in\Theta} |\omega| < \infty$, $\sup_{\boldsymbol{\theta}\in\Theta} |A| < \infty$, $\sup_{\boldsymbol{\theta}\in\Theta} |B| < \infty$, and $\sup_{\boldsymbol{\theta}\in\Theta} \|f^{\Theta}\|^{\Theta} < \infty$ hold by compactness of Θ and continuity of f^{Θ}, and $\mathbb{E}\|s(f^{\Theta}, y_t, X_t;\, \cdot\,, \cdot)\|^{\Theta} < \infty$ holds by assumption. This implies that $f_{\Theta} \in \mathbb{C}(\Theta, \mathcal{F})$ satisfies

$$\begin{aligned}
\mathbb{E}\log^+ \|\phi_0(f^{\Theta}) - f^{\Theta}\|_{\Theta} \;\le\; \mathbb{E}\|\phi_0(f^{\Theta}) - f^{\Theta}\|^{\Theta} &\le\; \mathbb{E}\|\phi(f^{\Theta}, y_t, X_t; \cdot)\|^{\Theta} + \|f^{\Theta}\|^{\Theta}\\
&= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\phi(f^{\Theta}(\boldsymbol{\theta}), y_t, X_t, \boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\Theta} |f^{\Theta}(\boldsymbol{\theta})| < \infty.
\end{aligned}$$

By a similar argument $\mathbb{E}\ln^+ \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} |s(\bar{f}_1, y_t, X_t; \beta, \lambda)| < \infty$ implies $\mathbb{E}\log^+ \|\phi_0(f^{\Theta}) - f^{\Theta}\|^{N_f}_{\Theta} < \infty$.

[11] That $(\mathbb{C}(\Theta, \mathcal{F}), \|\cdot\|_{\Theta})$ is a separable Banach space under compact Θ follows from application of the Arzeláscoli theorem to obtain completeness and the Stone-Weierstrass theorem for separability.

For any pair $(f^{\Theta}, f^{'\Theta}) \in \mathbb{C}(\Theta) \times \mathbb{C}(\Theta)$, define

$$\rho_t = \rho(\phi_t) = \sup_{(f^{\Theta}, f^{'\Theta}) \in \mathcal{F}_{\Theta} \times \mathcal{F}_{\Theta}} \frac{\|\phi_t(f^{\Theta}) - \phi_t(f^{'\Theta})\|_{\Theta}}{\|f^{\Theta} - f^{'\Theta}\|_{\Theta}}.$$

Condition C2 in Bougerol (1993b, Theorem 3.1) holds if $\mathbb{E}\ln\rho_t < 0$. This is ensured by $\mathbb{E}\ln\|\bar{\phi}_t'\|^{\Theta} < 0$, with $\bar{\phi}_t(\boldsymbol{\theta})$ as defined in (11). To see this, note that

$$\begin{aligned}
\mathbb{E}\ln\rho(\phi_t) &:= \mathbb{E}\ln \sup_{\|f^{\Theta} - f^{'\Theta}\| > 0} \frac{\|\phi_t(f^{\Theta}) - \phi_t(f^{'\Theta})\|^{\Theta}}{\|f^{\Theta} - f^{'\Theta}\|^{\Theta}} \\
&= \mathbb{E}\ln \sup_{\|f^{\Theta} - f^{'\Theta}\| > 0} \frac{\sup_{\boldsymbol{\theta}\in\Theta} |\phi(f(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t, \boldsymbol{\theta}) - \phi(f'(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t, \boldsymbol{\theta})|}{\sup_{\boldsymbol{\theta}\in\Theta} |f(\boldsymbol{\theta}, \bar{f}_1) - f'(\boldsymbol{\theta}, \bar{f}_1)|} \\
&\leq \mathbb{E}\ln \sup_{\|f^{\Theta} - f^{'\Theta}\| > 0} \frac{\sup_{\boldsymbol{\theta}\in\Theta} \bar{\phi}_t'(\boldsymbol{\theta}) \;\; \sup_{\boldsymbol{\theta}\in\Theta} |f(\boldsymbol{\theta}, \bar{f}_1) - f'(\boldsymbol{\theta}, \bar{f}_1)|}{\sup_{\boldsymbol{\theta}\in\Theta} |f(\boldsymbol{\theta}, \bar{f}_1) - f'(\boldsymbol{\theta}, \bar{f}_1)|} \\
&= \mathbb{E}\ln\|\bar{\phi}_t'\|^{\Theta} < 0.
\end{aligned}$$

Also note that for the t period composition of the stochastic recurrence equation, we have $\mathbb{E}\ln\rho(\phi_t \circ \ldots \circ \phi_1) \leq \mathbb{E}\ln\prod_{j=1}^{t}\rho(\phi_j) \leq \sum_{j=1}^{t}\ln\|\bar{\phi}_j'\|^{\Theta} < 0$, where $\circ$ denotes composition. As a result, $\{\tilde{f}_t(\cdot, \bar{f}_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE solution $\{\tilde{f}_t(\cdot)\}_{t\in\mathbb{Z}}$ in $\|\cdot\|^{\Theta}$-norm. Uniqueness and e.a.s. convergence is obtained in Straumann and Mikosch (2006, Theorem 2.8).

Finally, we show that $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1)|^{N_f} < \infty$ and also $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta})|^{N_f} < \infty$. We have $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1)|^{N_f} < \infty$ if and only if $\sup_t(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1)|^{N_f})^{1/N_f} = \sup_t \|\tilde{f}_t(\cdot, \bar{f}_1)\|_{N_f}^{\Theta} < \infty$. Furthermore, for any $f^{\Theta} \in \mathbb{C}(\Theta, \mathcal{F})$, having $\|\tilde{f}_t(\cdot, \bar{f}_1) - f^{\Theta}\|_{N_f}^{\Theta} < \infty$ implies $\|\tilde{f}_t(\cdot, \bar{f}_1^{\Theta})\|_{N_f}^{\Theta} < \infty$ since continuity on the compact Θ implies $\sup_{\boldsymbol{\theta}\in\Theta} |f(\boldsymbol{\theta})| < \infty$. For $f^{\Theta} \in \mathbb{C}(\Theta, \mathcal{F})$, we define f_*^{Θ}, y_*, and X_* such that $f^{\Theta} = \phi(y, X, f_*^{\Theta}, \cdot) \in \mathbb{C}(\Theta, \mathcal{F})$. Above we showed that $\exists\, f^{\Theta} \in \mathbb{C}(\Theta, \mathcal{F})$ satisfying $\|\phi(f^{\Theta}, y_t, X_t; \cdot)\|_{N_f}^{\Theta} \leq \bar{\bar{\phi}} < \infty$ and $\|\bar{f}_1^{\Theta} - f^{\Theta}\|_{N_f}^{\Theta} = \|\bar{f}_1^{\Theta} - \phi(f_*^{\Theta}, y_*, X_*; \cdot)\|_{N_f}^{\Theta} < \infty$. From this, we obtain

$$\begin{aligned}
\sup_t \|\tilde{f}_t(\cdot, \bar{f}_1^{\Theta}) - f^{\Theta}\|_{N_f}^{\Theta} &= \sup_t \|\phi(\tilde{f}_t(\cdot, \bar{f}_1^{\Theta}), y_t, X_t; \cdot) - \phi(f_*^{\Theta}, y_*, X_*; \cdot)\|_{N_f}^{\Theta} \\
&\leq \sup_t \|\phi(\tilde{f}_t(\cdot, \bar{f}_1^{\Theta}), y_t, X_t; \cdot) - \phi(f_*^{\Theta}, y_t, X_t; \cdot)\|_{N_f}^{\Theta} + \\
&\quad \sup_t \|\phi(f_*^{\Theta}, y_t, X_t; \cdot)\|_{N_f}^{\Theta} + \sup_t \|\phi(f_*^{\Theta}, y_*, X_*; \cdot)\|_{N_f}^{\Theta} \\
&\leq \sup_t \left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) - f_*^{\Theta}|^{N_f} \times \sup_{\boldsymbol{\theta}\in\Theta} \frac{|\phi(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1^{\Theta}), y_t, X_t; \boldsymbol{\theta}) - \phi(f_*^{\Theta}(\boldsymbol{\theta}), y_t, X_t; \boldsymbol{\theta})|^{N_f}}{|\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1^{\Theta}) - f_*^{\Theta}(\boldsymbol{\theta})|^{N_f}} \right)^{1/N_f} \\
&\quad + \sup_t \|\phi(f_*^{\Theta}, y_t, X_t; \cdot)\|_{N_f}^{\Theta} + \|f^{\Theta}\|_{N_f}^{\Theta} \\
&\leq \sup_t \left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) - f_*^{\Theta}|^{N_f} \times \sup_{\boldsymbol{\theta}\in\Theta} \bar{\phi}_t'(\boldsymbol{\theta})^{N_f} \right)^{1/N_f} + \sup_t \|\phi(f_*^{\Theta}, y_t, X_t; \cdot)\|_{N_f}^{\Theta} + \|f^{\Theta}\|_{N_f}^{\Theta}.
\end{aligned}$$

Using the orthogonality condition in *(iv′)*, we can write the expectation of the product as the product of the expectations and continue

$$\begin{aligned}
&\leq \sup_t \|\tilde{f}_t(\,\cdot\,, \bar{f}_1^{\Theta}) - f_*^{\Theta}\|_{N_f}^{\Theta} \cdot \|\bar{\phi}_t'\|_{N_f}^{\Theta} + \sup_t \|\phi(f_*^{\Theta}, y_t, X_t; \cdot)\|_{N_f}^{\Theta} + \|f^{\Theta}\|_{N_f}^{\Theta} \\
&\leq \|\bar{\phi}_t'\|_{N_f}^{\Theta} \times \left(\sup_t \|\tilde{f}_t(\,\cdot\,, \bar{f}_1^{\Theta}) - f_*^{\Theta}\|_{N_f}^{\Theta} \right) + \bar{\bar{\phi}} + \bar{f},
\end{aligned}$$

with $\bar{c} = \|\bar{\phi}_t'\|_{N_f}^{\Theta} < 1$ by condition *(iv′)*, $\bar{\bar{\phi}} < \infty$, and $\bar{f} = \|f^{\Theta}\| + \bar{c} \cdot \|f^{\Theta} - f_*^{\Theta}\|_{N_f}^{\Theta} < \infty$. As a result we have the recursion $\sup_t \|\tilde{f}_t(\cdot, \bar{f}_1^{\Theta}) - f^{\Theta}\|_{N_f}^{\Theta} \leq \bar{c} \cdot \sup_t \|\tilde{f}_t(\cdot, \bar{f}_1^{\Theta}) - f^{\Theta}\|_{N_f}^{\Theta} + A$, with $A = \bar{\bar{\phi}} + \bar{f}$. Hence,

$$
\begin{aligned}
\sup_t \|\tilde{f}_t(\cdot, \bar{f}_1^{\Theta}) - f^{\Theta}\|_{N_f}^{\Theta} &\leq \sum_{j=0}^{t} (\bar{c})^j ((\bar{c}+1)\bar{f} + \bar{\bar{\phi}}) + \bar{c}^{t+1} \sup_t \|\bar{f}_1^{\Theta} - f^{\Theta}\|_{N_f}^{\Theta} \\
&\leq \frac{(\bar{c}+1)\bar{f} + \bar{\bar{\phi}}}{1-\bar{c}} + \|\bar{f}_1^{\Theta} - f^{\Theta}\|_{N_f}^{\Theta} < \infty.
\end{aligned}
$$

The same result holds using the uniform contraction in *(iv)* by taking a further supremum in y_t and X_t instead of the orthogonality condition. □

Proof of Theorem 2: Assumption 1 implies that $\ell_T(\boldsymbol{\theta}, \bar{f}_1) = (1/T)\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}, \bar{f}_1)$ is a.s. continuous (a.s.c.) in $\boldsymbol{\theta} \in \Theta$ through continuity (c.) of each

$$
\begin{aligned}
\ell_t(\boldsymbol{\theta}, \bar{f}_1) &= \ell(y_t, X_t, \tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}), \boldsymbol{\theta}) \\
&= \log p_e(Z_t(f_t)^{-1} y_t - X_t\beta; \lambda) - \log |Z_t(f_t)|
\end{aligned}
$$

ensured in turn by the differentiability of S, p_e and h and the implied a.s.c. of

$$
\begin{aligned}
\nabla(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}), y_t, X_t; \beta, \lambda) =& \frac{\partial \log p_e(Z_t(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}))^{-1} y_t - X_t\beta; \lambda)}{\partial f} \\
&- \frac{\partial \log |Z_t(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}))|}{\partial f}
\end{aligned}
$$

in $(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}); \lambda)$ and the resulting c. of $\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta})$ in $\boldsymbol{\theta}$ as a composition of t c. maps. Together with the compactness of Θ this implies by Weierstrass' theorem that the $\arg\max$ set is non-empty a.s. and hence that $\hat{\boldsymbol{\theta}}_T$ exists a.s. $\forall T \in \mathbb{N}$. Assumption 1 implies also by a similar argument that

$$
\ell_T(\boldsymbol{\theta}, \bar{f}_1) = \ell\big(\tilde{f}^{1:T}(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}); y^{1:T}, X^{1:T}, \boldsymbol{\theta}\big)
$$

is continuous in $(y^{1:T}, X^{1:T})$ $\forall\, \boldsymbol{\theta} \in \Theta$ and hence measurable w.r.t. the product Borel σ-algebra $\mathfrak{B}(\mathcal{Y}) \otimes \mathfrak{B}(\mathcal{X})$ that are, in turn, measurable maps w.r.t. $\mathcal{F}$ by Proposition 4.1.7 in Dudley (2002).[12] The measurability of $\hat{\boldsymbol{\theta}}_T$ follows from Foland (2009, p.24) and White (1994, Theorem 2.11) or Gallant and White (1988, Lemma 2.1, Theorem 2.2).[13] □

Proof of Theorem 3: We obtain $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \stackrel{a.s.}{\to} \boldsymbol{\theta}_0$ from the uniform convergence of the criterion function

$$
\sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_\infty(\boldsymbol{\theta})| \stackrel{a.s.}{\to} 0 \;\forall\, \bar{f}_1 \in \mathcal{F} \;\text{ as }\; T \to \infty, \tag{C.1}
$$

and the identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \Theta$ introduced in White (1994),

$$
\sup_{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \epsilon} \ell_\infty(\boldsymbol{\theta}) < \ell_\infty(\boldsymbol{\theta}_0) \;\forall\, \epsilon > 0; \tag{C.2}
$$

[12] Dudley's proposition states that the Borel σ-algebra $\mathfrak{B}(\mathbb{A} \times \mathbb{B})$ generated by the Tychonoff's product topology $\mathcal{T}_{\mathbb{A}\times\mathbb{B}}$ on the space $\mathbb{A} \times \mathbb{B}$ includes the product σ-algebra $\mathfrak{B}(\mathbb{A}) \otimes \mathbb{B}(\mathbb{B})$.

[13] The reference of Foland (2009) is used here to establish that a map into a product space is measurable if and only if its projections are measurable.

see for example White (1994, Theorem 3.4) or Theorem 3.3 in Gallant and White (1988) for further details.

The uniform convergence is obtained by norm sub-additivity,[14]

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_\infty(\boldsymbol{\theta})| \ \leq \ \sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_T(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})|,$$

and then showing that the initialization effect vanishes asymptotically,

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_T(\boldsymbol{\theta})| \overset{a.s.}{\to} 0 \ \text{ as } \ T \to \infty, \tag{C.3}$$

and for the second term applying the ergodic theorem for separable Banach spaces in Ranga Rao (1962), as in Straumann and Mikosch (2006, Theorem 2.7), to the sequence $\{\ell_T(\cdot)\}$ with elements taking values in $\mathbb{C}(\Theta, \mathbb{R})$ so that

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0 \quad \text{where} \quad \ell_\infty(\boldsymbol{\theta}) = \mathbb{E}\ell_t(\boldsymbol{\theta})\, \forall\, \boldsymbol{\theta} \in \Theta.$$

The criterion $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$ satisfies (C.3) if

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta}, \bar{f}_1) - \ell_t(\boldsymbol{\theta})| \overset{a.s.}{\to} 0 \quad \text{as} \quad t \to \infty.$$

The continuity of p_e ensures that $\ell_t(\cdot, \bar{f}_1) = \ell(\tilde{f}_t(y^t, X^t, \cdot, \bar{f}_1), y_t, X_t, \cdot)$ is continuous in $(\tilde{f}_t(y^t, X^t, \cdot, \bar{f}_1), y_t, X_t)$. Since all the assumptions of Theorem 1 are satisfied we know that there exists a unique SE sequence $\{\tilde{f}_t(y^t, X^t, \cdot))\}_{t\in\mathbb{Z}}$ with elements taking values in $\mathbb{C}(\Theta, \mathcal{F})$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|(\tilde{f}_t(y^{t-1}, X^{t-1}, \bar{f}_1, \boldsymbol{\theta}), y_t, X_t) - (\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}), y_t, X_t)\right| \overset{a.s.}{\to} 0,$$

and

$$\sup_t \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(y^{t-1}, X^{t-1}, \bar{f}_1, \boldsymbol{\theta})^{N_f}| < \infty \quad \text{and} \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta})|^{N_f} < \infty,$$

with $N_f \geq 1$. Hence, (C.3) follows by application of a continuous mapping theorem for $\ell : \mathbb{C}(\Theta, \mathcal{F}) \to \mathbb{C}(\Theta, \mathcal{F})$.

The ULLN $\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \mathbb{E}\ell_t(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$ as $T \to \infty$ follows, under a moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| < \infty$, by the SE nature of $\{\ell_T\}_{t\in\mathbb{Z}}$ which is implied by continuity of ℓ on the SE sequence $\{(y_t, X_t, \tilde{f}_t(y^{t-1}, X^{t-1}, \cdot))\}_{t\in\mathbb{Z}}$ and Proposition 4.3 in Krengel (1985). The moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| < \infty$ can be established as follows. First note that

$$\begin{aligned}
\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| &= \sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}| \log p_e(y_t - h(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))Wy_{t-1} - X_t\beta) \\
&\quad - \log \det Z(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))| \\
&\leq \sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}| \log p_e(y_t - h(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))Wy_{t-1} - X_t\beta)| \\
&\quad - \sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}| \log \det Z(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))| < \infty,
\end{aligned}$$

then the bounded first moment for the likelihood is implied by having

$$\mathbb{E}|y_t|^{N_y} < \infty \quad , \quad \mathbb{E}|X_t|^{N_X} < \infty \quad \text{and} \quad \sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}|\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta})|^{N_f} < \infty$$

[14] $\ell_T(\boldsymbol{\theta})$ denotes $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$ with $\tilde{f}(\boldsymbol{\theta}), \bar{f}_1)$ replaced by its limit for $T \to \infty$, i.e., by $\tilde{f}(\boldsymbol{\theta})$.

since then

$$\sup_{\boldsymbol{\theta}\in\Theta}\mathbb{E}|\log\det Z(\tilde{f}_t(y^{t-1},X^{t-1},\boldsymbol{\theta}))|<\infty,$$

because $\log|Z|\in\mathbb{M}(N_f,N_{\log|Z|})$ with $N_{\log|Z|}\geq 1$ by assumption and

$$\sup_{\boldsymbol{\theta}\in\Theta}\mathbb{E}|\log p_e(y_t-h(\tilde{f}_t(y^{t-1},X^{t-1},\boldsymbol{\theta}))Wy_{t-1}-X_t\beta)<\infty,$$

because h is uniformly bounded and hence the argument of p_e has the same moments as y_t and X_t. This ensures the desired moment since $\log p_e\in\mathbb{M}(N,N_{\log p_e})$ with $N_{\log p_e}\geq 1$ and $N=\min\{N_y,N_x\}$ by assumption.

Finally, the identifiable uniqueness (see e.g. White (1994)) of $\boldsymbol{\theta}_0\in\Theta$ in (C.2) follows from the assumed uniqueness, the compactness of Θ, and the continuity of the limit $\mathbb{E}\ell_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta}\in\Theta$ which is implied by the continuity of ℓ_T in $\boldsymbol{\theta}\in\Theta\ \forall\, T\in\mathbb{N}$ and the uniform convergence in (C.1). □

Proof of Theorem 4: As the likelihood and its derivatives depend on the derivatives of $\tilde{f}(\boldsymbol{\theta},\bar{f}_1)$ with respect to $\boldsymbol{\theta}$, we introduce the notation $\boldsymbol{f}^{(0:m)}$ as the vector containing $\tilde{f}(\boldsymbol{\theta},\bar{f}_1)$ and its derivatives up to order m, with initial condition $\bar{\boldsymbol{f}}^{(0:m)}$. We obtain the desired result from: (i) the strong consistency of $\hat{\boldsymbol{\theta}}_T\overset{a.s.}{\to}\boldsymbol{\theta}_0\in\operatorname{int}(\Theta)$; (ii) the a.s. twice continuous differentiability of $\ell_T(\boldsymbol{\theta},\bar{f}_1)$ in $\boldsymbol{\theta}\in\Theta$; (iii) the asymptotic normality of the score

$$\sqrt{T}\ell_T'\big(\boldsymbol{\theta}_0,\bar{\boldsymbol{f}}_1^{(0:1)}\big)\overset{d}{\to}\mathrm{N}\big(0,\mathcal{J}(\boldsymbol{\theta}_0)\big),\qquad \mathcal{J}(\boldsymbol{\theta}_0)=\mathbb{E}\big(\tilde{\ell}_t'(\boldsymbol{\theta}_0)\tilde{\ell}_t'(\boldsymbol{\theta}_0)^\top\big);\tag{C.4}$$

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

$$\sup_{\boldsymbol{\theta}\in\Theta}\big\|\ell_T''(\boldsymbol{\theta},\bar{\boldsymbol{f}}_1^{(0:2)})-\ell_\infty''(\boldsymbol{\theta})\big\|\overset{a.s.}{\to}0;\tag{C.5}$$

and finally, (v) the non-singularity of the limit $\ell_\infty''(\boldsymbol{\theta})=\mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta})=\mathcal{I}(\boldsymbol{\theta})$. See e.g. in White (1994, Theorem 6.2)) for further details.

The consistency condition $\hat{\boldsymbol{\theta}}_T\overset{a.s.}{\to}\boldsymbol{\theta}_0\in\operatorname{int}(\Theta)$ in (i) follows under the maintained assumptions from Theorem 3 and the additional assumption in Theorem 4 that $\boldsymbol{\theta}_0\in\operatorname{int}(\Theta)$. The smoothness condition in (ii) follows immediately from Assumption 2 and the likelihood expressions in Appendix C.2.

The asymptotic normality of the score in (C.7) follows by Theorem 18.10[iv] in van der Vaart (2000) by showing that

$$\|\ell_T'\big(\boldsymbol{\theta}_0,\bar{\boldsymbol{f}}_1^{(0:1)}\big)-\ell_T'\big(\boldsymbol{\theta}_0\big)\|\overset{e.a.s.}{\to}0\ \text{ as }\ T\to\infty,\tag{C.6}$$

plus a CLT result for $\ell_T'(\boldsymbol{\theta}_0)$. Note that from (C.6) we obtain that $\sqrt{T}\|\ell_T'\big(\boldsymbol{\theta}_0,\bar{\boldsymbol{f}}_1^{(0:1)}\big)-\ell_T'\big(\boldsymbol{\theta}_0\big)\|\overset{a.s.}{\to}0$ as $T\to\infty$. The desired CLT result follows by an application of the CLT for SE martingales in Billingsley (1961),

$$\sqrt{T}\ell_T'\big(\boldsymbol{\theta}_0\big)\overset{d}{\to}\mathrm{N}\big(0,\mathcal{J}(\boldsymbol{\theta}_0)\big)\ \text{ as }\ T\to\infty,\tag{C.7}$$

where $\mathcal{J}(\boldsymbol{\theta}_0)=\mathbb{E}(\tilde{\ell}_t'\big(\boldsymbol{\theta}_0)\tilde{\ell}_t'(\boldsymbol{\theta}_0)^\top)<\infty$, where finite (co)variances follow from the assumption $N_{\ell'}\geq 2$ in Assumption 4 and the expressions for the likelihood in Appendix C.2.

To establish the e.a.s. convergence in (C.6), we use the e.a.s. convergence

$$|\tilde{f}_t(y^{1:t-1},X^{1:t-1},\boldsymbol{\theta}_0,\bar{f}_1)-\tilde{f}_t(y^{t-1},X^{t-1},\boldsymbol{\theta}_0)|\overset{e.a.s.}{\to}0,\tag{C.8}$$

and

$$\|\tilde{\boldsymbol{f}}_t^{(1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \tilde{\boldsymbol{f}}_t^{(1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0)\| \overset{e.a.s.}{\to} 0. \tag{C.9}$$

The e.a.s. convergence in (C.8) is obtained directly by application of Theorem 1 under the maintained assumptions. The e.a.s. convergence in (C.9) is obtained by the same argument as in the proof of Theorem 1 since: (a) the expressions for the derivative process $\{\tilde{\boldsymbol{f}}_t^{(1)}\}$ in Appendix C.2 show that the contraction condition

$$\mathbb{E} \ln \sup_{\boldsymbol{\theta} \in \Theta} \bar{\phi}_{1,1}'(\boldsymbol{\theta}) < 0$$

for the recursion of the filter $\{\tilde{f}_t\}$ is the same as the contraction condition for the derivative process $\{\tilde{\boldsymbol{f}}_t^{(1)}\}$; and (b) the expressions in Appendix C.2 also reveal that the counterpart of the moment condition

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

used in Theorem 1 for the filtered process $\{\tilde{f}_t\}$, is implied by the condition that

$$\min\{N_f, N_s, N_s^{0,1,0}, N_s^{(0,0,1)}\} > 0,$$

as imposed in Assumption 4.

From the differentiability of

$$\tilde{\ell}_t'(\boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{(0:1)}) = \ell'\big(\boldsymbol{\theta}, y^{1:t}, X^{1:t}, \tilde{\boldsymbol{f}}_t^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{(0:1)})\big)$$

in $\tilde{\boldsymbol{f}}_t^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{(0:1)})$ and the convexity of $\mathcal{F}$, we use the mean-value theorem to obtain

$$\begin{aligned} \|\ell_T'\big(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \ell_T'\big(\boldsymbol{\theta}_0)\| \le \sum_{j=1}^{4+d_\lambda} \Big| \frac{\partial \ell'(y^{1:t}, X^{1:t}, \hat{\boldsymbol{f}}_t^{(0:1)})}{\partial f_j} \Big| \\ \times \big| \tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0) \big|, \end{aligned} \tag{C.10}$$

where d_λ denotes the dimension of λ, and $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}$ denotes the j-th element of $\tilde{\boldsymbol{f}}_t^{(0:1)}$, and $\hat{\boldsymbol{f}}^{(0:1)}$ is on the segment connecting $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)})$ and $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0)$. Note that $\tilde{\boldsymbol{f}}_t^{(0:1)} \in \mathbb{R}^{4+d_\lambda}$ because it contains $\tilde{f}_t \in \mathbb{R}$ (the first element) as well as $\tilde{\boldsymbol{f}}_t^{(1)} \in \mathbb{R}^{3+d_\lambda}$ (the derivatives with respect to ω, A, B, and λ). Using the expressions of the likelihood and its derivatives in Appendix C.2, the moment bounds and the moment preserving properties in Assumption 4, and the expressions in Appendix C.2 shows that

$$\big|\partial \ell'(y^{1:t}, X^{1:t}, \hat{\boldsymbol{f}}_t^{(0:1)})/\partial f_j\big| = O_p(1) \quad \forall j = 1, \ldots, 4 + d_\lambda.$$

The strong convergence in (C.10) is now ensured by

$$\|\ell_T'\big(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \ell_T'\big(\boldsymbol{\theta}_0)\| = \sum_{i=1}^{4+d_\lambda} O_p(1) o_{e.a.s}(1) = o_{e.a.s.}(1). \tag{C.11}$$

The proof of the uniform convergence in (C.5) is similar to that of Theorem 2. We note

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\ell_T''(\boldsymbol{\theta}, \bar{f}_1) - \ell_\infty''(\boldsymbol{\theta})\| \le \sup_{\boldsymbol{\theta} \in \Theta} \|\ell_T''(\boldsymbol{\theta}, \bar{f}_1) - \ell_T''(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\ell_T''(\boldsymbol{\theta}) - \ell_\infty''(\boldsymbol{\theta})\|. \tag{C.12}$$

To prove that the first term vanishes a.s., we show that

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta}, \bar{f}_1) - \tilde{\ell}_t''(\boldsymbol{\theta})\| \stackrel{a.s.}{\to} 0 \quad \text{as} \quad t \to \infty.$$

The differentiability of $\tilde{g}$, $\tilde{g}'$, $\tilde{p}$, and S from Assumption 2 ensure that

$$\tilde{\ell}_t''(\cdot, \bar{f}_1) = \ell''(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{\boldsymbol{f}}_{0:2}), \cdot)$$

is continuous in $(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \cdot, \boldsymbol{f}_{0:2}))$. Again, we note that the proof of Theorem 1 can be easily adapted to show that there exists a unique SE sequence $\{\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, X^{t-1}, \cdot)\}_{t\in\mathbb{Z}}$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\|(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_{0:2})) - (y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, X^{t-1}, \boldsymbol{\theta})\right\| \stackrel{a.s.}{\to} 0,$$

and satisfying, for for $N_f \geq 1$,

$$\sup_t \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_{0:2})\|^{N_f} < \infty,$$

and also

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, X^{t-1}, \boldsymbol{\theta})\|^{N_f} < \infty,$$

because (a) the expressions for the derivative process $\{\tilde{\boldsymbol{f}}_t^{(1)}\}$ in Appendix C.2 show that the contraction condition

$$\mathbb{E} \ln \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\phi}_{1,1}'(\boldsymbol{\theta}) < 0$$

for the recursion of the filter $\{\tilde{f}_t\}$ is the same as the contraction condition for the second derivative process $\{\tilde{\boldsymbol{f}}_t^{(2)}\}$; and (b) the expressions in Appendix C.2 show also that the counterpart of the moment condition

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

used in Theorem 1 for the filtered process $\{\tilde{f}_t\}$, is implied by the condition that

$$\min \Bigg\{ N_f^{(1)} \,,\, N_s^{(0,1,0)} \,,\, N_s^{(0,0,1)} \,,\, N_s^{(0,2,0)} \,,\, N_s^{(0,0,2)} \,,\, N_s^{(0,1,1)} \,,\, \frac{N_s^{(1,0,0)} N_f^{(1)}}{N_s^{(1,0,0)} + N_f^{(1)}} \,,$$
$$\frac{N_s^{(2,0,0)} N_f^{(1)}}{2N_s^{(2,0,0)} + N_f^{(1)}} \,,\, \frac{N_s^{(1,1,0)} N_f^{(1)}}{N_s^{(1,1,0)} + N_f^{(1)}} \,,\, \frac{N_s^{(1,0,1)} N_f^{(1)}}{N_s^{(1,0,1)} + N_f^{(1)}} \Bigg\} > 0 \,,$$

imposed in Assumption 4. By application of a continuous mapping theorem for $\ell'' : \mathbb{C}(\Theta \times \mathcal{F}^{(0:2)}) \to \mathbb{R}$ we thus conclude that the first term in (C.12) converges to 0 a.s..

The second term in (C.12) converges under a bound $\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta})\| < \infty$ by the SE nature of $\{\ell_T''\}_{t\in\mathbb{Z}}$. The latter is implied by continuity of ℓ'' on the SE sequence

$$\{(y_t, X_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \cdot))\}_{t\in\mathbb{Z}}.$$

The moment bound $\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta})\| < \infty$ follows from $N_{\ell''} \geq 1$ in Assumption 4 and the expressions in Appendix C.2. Finally, the non-singularity of the limit $\ell_\infty''(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$ in (v) below equation (C.5) is implied by the uniqueness of $\boldsymbol{\theta}_0$ as a maximum of $\ell_\infty''(\boldsymbol{\theta})$ in Θ. $\square$

C.2 Derivatives of the likelihood function

We take first derivatives of the likelihood with respect to all static parameters $\boldsymbol{\theta} = (\omega, a, b, \beta', \sigma^2)'$:

$$\frac{\partial \ell_t}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \ell_t}{\partial \omega}, \frac{\partial \ell_t}{\partial a}, \frac{\partial \ell_t}{\partial b}, \frac{\partial \ell_t}{\partial \beta}, \frac{\partial \ell_t}{\partial \sigma^2} \right)'$$

Let θ_m denote the mth element of $\boldsymbol{\theta}$. We can decompose the derivatives of the likelihood with respect to each θ_m into two parts:

$$\begin{aligned} \frac{\partial \ell_t}{\partial \theta_m} &= \frac{\partial (\tilde{p}_t + \ln \tilde{g}_t')}{\partial f_t} \cdot \frac{\partial f_t}{\partial \theta_m} + \frac{\partial \tilde{p}_t}{\partial \theta_m} \\ &= \nabla_t \cdot \frac{\partial f_t}{\partial \theta_m} + \frac{\partial \tilde{p}_t}{\partial \theta_m}, \end{aligned} \tag{C.13}$$

because $\tilde{g}_t'$ does not depend on any of the parameters directly, only through f_t. For $\theta_m \in \{\omega, a, b\}$ the second term is zero, because these parameters enter the likelihood only through f_t.

All partial derivatives contain the term $\frac{\partial f_t}{\partial \theta_m}$ given by

$$\begin{aligned} \frac{\partial f_t}{\partial \theta_m} &= \frac{\partial}{\partial \theta_m} (\omega + a s_{t-1} + b f_{t-1}) &\text{(C.14)} \\ &= \frac{\partial \omega}{\partial \theta_m} + \frac{\partial a}{\partial \theta_m} s_{t-1} + a \frac{\partial s_{t-1}}{\partial f_{t-1}} \cdot \frac{\partial f_{t-1}}{\partial \theta_m} + a \frac{\partial s_{t-1}}{\partial \theta_m} + \frac{\partial b}{\partial \theta_m} f_{t-1} + b \frac{\partial f_{t-1}}{\partial \theta_m} &\text{(C.15)} \\ &= \frac{\partial \omega}{\partial \theta_m} + \frac{\partial a}{\partial \theta_m} \nabla_{t-1} + a \nabla_{t-1}' \cdot \frac{\partial f_{t-1}}{\partial \theta_m} + a \frac{\partial \nabla_{t-1}}{\partial \theta_m} + \frac{\partial b}{\partial \theta_m} f_{t-1} + b \frac{\partial f_{t-1}}{\partial \theta_m} &\text{(C.16)} \\ &= \frac{\partial \omega}{\partial \theta_m} + \frac{\partial a}{\partial \theta_m} \nabla_{t-1} + a \frac{\partial \nabla_{t-1}}{\partial \theta_m} + \frac{\partial b}{\partial \theta_m} f_{t-1} + (a \nabla_{t-1}' + b) \frac{\partial f_{t-1}}{\partial \theta_m} &\text{(C.17)} \end{aligned}$$

We want the matrix of second derivatives of the likelihood function, i.e.

$$\frac{\partial^2 \ell_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

We take another derivative of (C.13) with respect to θ_o:

$$\frac{\partial^2 \ell_t}{\partial \theta_m \partial \theta_o} = \nabla_t' \cdot \frac{\partial f_t}{\partial \theta_o} \cdot \frac{\partial f_t}{\partial \theta_m} + \frac{\partial \nabla_t}{\partial \theta_o} \cdot \frac{\partial f_t}{\partial \theta_m} + \nabla_t \frac{\partial^2 f_{t-1}^2}{\partial \theta_m \theta_o} + \frac{\partial^2 \tilde{p}_t}{\partial \theta_m \partial \theta_o} \tag{C.18}$$

The second derivative process takes the form

$$\begin{aligned} \frac{\partial^2 f_t}{\partial \theta_m \partial \theta_o} =& \frac{\partial a}{\partial \theta_m} \cdot \frac{\partial \nabla_{t-1}}{\partial f_{t-1}} \cdot \frac{\partial f_{t-1}}{\partial \theta_o} + \frac{\partial a}{\partial \theta_m} \frac{\partial \nabla_{t-1}}{\partial \theta_o} \\ &+ \frac{\partial a}{\partial \theta_o} \frac{\partial \nabla_{t-1}}{\partial f_{t-1}} \frac{\partial f_{t-1}}{\partial \theta_m} + a \frac{\partial^2 \nabla_{t-1}}{\partial f_{t-1}^2} \frac{\partial f_{t-1}}{\partial \theta_o} \frac{\partial f_{t-1}}{\partial \theta_m} + a \frac{\partial^2 \nabla_{t-1}}{\partial f_{t-1} \partial \theta_o} \frac{\partial f_{t-1}}{\partial \theta_m} \\ &+ a \frac{\partial \nabla_{t-1}}{\partial f_{t-1}} \frac{\partial^2 f_{t-1}}{\partial \theta_m \partial \theta_o} + \frac{\partial a}{\partial \theta_o} \frac{\partial \nabla_{t-1}}{\partial \theta_m} + a \frac{\partial^2 \nabla_{t-1}}{\partial \theta_m \partial \theta_o} + a \frac{\partial^2 \nabla_{t-1}}{\partial \theta_m \partial f_{t-1}} \frac{\partial f_{t-1}}{\partial \theta_o} \\ &+ \frac{\partial b}{\partial \theta_m} \frac{\partial f_{t-1}}{\partial \theta_o} + \frac{\partial b}{\partial \theta_o} \frac{\partial f_{t-1}}{\partial \theta_m} + b \frac{\partial^2 f_{t-1}}{\partial \theta_m \partial \theta_o} \\ =& \frac{\partial a}{\partial \theta_m} \cdot \nabla_{t-1}' \cdot \frac{\partial f_{t-1}}{\partial \theta_o} + \frac{\partial a}{\partial \theta_m} \frac{\partial \nabla_{t-1}}{\partial \theta_o} \\ &+ \frac{\partial a}{\partial \theta_o} \cdot \nabla_{t-1}' \cdot \frac{\partial f_{t-1}}{\partial \theta_m} + a \nabla_{t-1}'' \cdot \frac{\partial f_{t-1}}{\partial \theta_o} \frac{\partial f_{t-1}}{\partial \theta_m} + a \frac{\partial \nabla_{t-1}'}{\partial \theta_o} \frac{\partial f_{t-1}}{\partial \theta_m} \\ &+ a \nabla_{t-1}' \frac{\partial^2 f_{t-1}}{\partial \theta_m \partial \theta_o} + \frac{\partial a}{\partial \theta_o} \frac{\partial \nabla_{t-1}}{\partial \theta_m} + a \frac{\partial^2 \nabla_{t-1}}{\partial \theta_m \partial \theta_o} + a \frac{\partial^2 \nabla_{t-1}}{\partial \theta_m \partial f_{t-1}} \frac{\partial f_{t-1}}{\partial \theta_o} \\ &+ \frac{\partial b}{\partial \theta_m} \frac{\partial f_{t-1}}{\partial \theta_o} + \frac{\partial b}{\partial \theta_o} \frac{\partial f_{t-1}}{\partial \theta_m} + b \frac{\partial^2 f_{t-1}}{\partial \theta_m \partial \theta_o}. \end{aligned} \tag{C.19}$$