TI 2021-085/III
Tinbergen Institute Discussion Paper

Heterogeneity and Dynamics in Spatial Network Models

Revision: September 2022

Enzo D'Innocenzo[1]
André Lucas[1,2]
Anne Opschoor[1,2]
Xingmin Zhang[3]

1 Vrije Universiteit Amsterdam

2 Tinbergen Institute

3 School of Finance, Southwestern University of Finance and Economics

Heterogeneity and Dynamics in Spatial Network Models*

Enzo D'Innocenzo[a], André Lucas[a,b] Anne Opschoor[a,b], Xingmin Zhang[c,⋆]

September 20, 2022

[a] *Vrije Universiteit Amsterdam*

[b] *Tinbergen Institute*

[c] *School of Finance, Southwestern University of Finance and Economics*

Abstract

We propose an empirical spatial modeling framework that allows for both heterogeneity and time-variation in economic network connections and spillovers. We establish the model's stationarity and ergodicity properties and also show that the model-implied filter is invertible. While highly flexible, the model is straightforward to estimate. We apply the model to several datasets related to Eurozone sovereign credit risk over the period December 2009 to July 2020. Accounting for both heterogeneity and time-variation turns out to be empirically highly important: the new model uncovers intuitive patterns that would go unnoticed in either homogeneous and/or static spatial network models.

Keywords: dynamic networks, spatial auto-regressions, heterogeneous spatial contagion, network heterogeneity, sovereign risk dynamics.

*Corresponding author: Xingmin Zhang, 555 Liutai Avenue, Wenjiang District, zip code 611130, Chengdu, Sichuan, P. R. China, email zhangxingmin@swufe.edu.cn. We thank Bernd Schwaab and seminar participants at SoFiE 2022 Cambridge, and Vrije Universiteit Amsterdam for suggestions that helped to improve the paper. This research was initiated when Zhang was visiting Vrije Universiteit Amsterdam on a Chinese Research Council scholarship. Zhang acknowledges financial support from National Natural Science Foundation of China (No.72101209). Opschoor acknowledges financial support from the Dutch National Science Foundation (NWO, grant VI.Vidi.201.079).

1 Introduction

In this paper we study the spatial network dynamics of Eurozone sovereigns. Spatial models have recently attracted quite some attention to describe networks of markets and market participants. For instance, spatial models have been used to study international stock market linkages (Fernandez, 2011; Fernández-Avilés et al., 2012; Asgharian et al., 2013), asset pricing implications of network connections (Kou et al., 2018), cross-sectional market volatility patterns as related to firm connections (Herskovic et al., 2020), financial markets and banking sector stability (Tonzer, 2015; Milcheva and Zhu, 2016; Pino and Sharma, 2019), contagion across energy markets (Demirer et al., 2020), credit rating contagion (Asgharian et al., 2013), and pricing and portfolio implications in real estate markets (Zhu and Lizieri, 2021; Zhu and Milcheva, 2020).

In all these models, the linkages between network players are typically captured by a (possibly time-varying) matrix of observable variables that measures the connection strengths. For instance, in our application we use quarterly cross-border banking claims in the Eurozone as in Tonzer (2015) and Blasques et al. (2016) for this purpose. On top of this observable matrix of network connections, the relative connection strengths in the network are generally captured by a single, static parameter, also known as the spatial correlation. This is quite restrictive. In dynamic and possibly stressed markets different network players can have a time-varying sensitivity to other players in the network. Moreover, this player-specific sensitivity may also increase or decrease disproportionately in a player-specific way during stressed periods. This heterogeneity may not be captured by only the observable variables that are typically used to describe the network connections. This poses two main challenges.

First, as indicated by Aquaro et al. (2021), there is a need to allow for more heterogeneity in the spatial model parameters that describe the sensitivities of network players to shocks elsewhere in the network. Aquaro et al. (2021) show that standard scalar spatial models are typically inadequate and may obscure and bias important patterns in the data. More heterogeneity in the spillover parameters is called for to capture the network structure correctly. This finding is also in line with Herskovic et al. (2020), who allow for different network sensitivities of firms in a network setting. Both Aquaro et al. (2021) and Herskovic et al. (2020) use heterogeneous extensions of the baseline spatial model for financial networks. These extensions, however, are all static.

This brings us to the second challenge: time-variation in network connections. Blasques et al. (2016), Catania and Billé (2017), and Billé et al. (2019) show that the overall strength of network connections can change substantially over time in stock markets, debt markets, and real estate markets. These papers introduce dynamic extensions of the baseline spatial model that allow for additional dynamics in network propagation strength on top of what is captured by simple observables. All three papers do so with one or two dynamic parameters. This allows one to

obtain additional measures of systemic stress and network fragility that exhibit more dynamics during turbulent times compared to commonly used static models. Still, these models allow for only a limited amount of (dynamic) heterogeneity across the network players; compare Aquaro et al. (2021). This may severely bias the results against time-variation in network propagation strength and hide important features of the network and its dynamics. For example, Blasques et al. (2016) find evidence of time-varying spillover strength for European sovereign debt, but the time-variation is surprisingly modest given the variety of countries included in their analysis. From our empirical application in this paper, we show this is likely due to over-restrictive pooling assumptions in the original model specification.

This paper develops a new spatial dependence model for dynamic networks with heterogeneous, dynamic spatial spillover parameters, thus filling an important gap in the empirical literature. On top of time-variation in network connections via observed variables, we introduce separate time-varying spatial dependence parameters. This enables each network player to have its own time-varying sensitivity to what happens elsewhere in the network. The dynamics in our model are based on the score-driven dynamics of Creal et al. (2013) and Harvey (2013). We show that the model has a stationary and ergodic solution, and that the model-implied filter is invertible, such that asymptotically we can correctly recover the paths of the time-varying spatial correlations from the data.

Compared to earlier score-driven correlation-type models we introduce three innovations to the score-driven specification. First, we allow for a vector of time-varying correlations. Second, we do not force the spatial correlations to remain in the [0,1] interval. This is empirically important, as several spatial correlations lie empirically outside this range. Third, we introduce time-varying step lengths for the scaled score steps to ensure the spatial stability of the model at all times. This approach also facilitates some of the theoretical derivations for the statistical properties of the model. Though the proposed mechanism is not called upon in our empirical application, it may provide useful in other settings to ensure the model remains spatially stable at all times by construction.

Empirically and in simulations, we find that it is important to allow for both heterogeneity and time-variation. If the dynamics are omitted, differences in network fragility over time are not captured well. Conversely, if heterogeneity is omitted, a scalar dynamic spatial parameter only provides a very blurred picture of the positions of the different network players over time. For instance, if the network importance of some countries increases precisely at the moment where that of others wanes, the overall scalar summary measure might incorrectly signal nothing is happening in the network at all. Again, this may lead to flawed inference and to an incorrect assessment of the true underlying economic mechanisms. Our new model avoids these pitfalls and allows

one to capture the heterogeneity, dynamics, and asymmetric risk propagation between network players more accurately. Despite all this added flexibility, the new model is still straightforward to estimate.

We apply our model to three different datasets related to European sovereign credit risk. Our baseline analysis uses weekly changes in 5-year government bond spreads over the OIS EONIA rate for 7 European countries over the period 2009-2020. We find that allowing for heterogeneous dependence parameters improves the statistical fit compared to the scalar spatial model of Blasques et al. (2016), as well as compared to static heterogeneous models or static scalar models, like the original model of Anselin (2009). Important features characterizing the dynamics of the European sovereign debt crisis only become clear in the new heterogeneous dynamic model. This includes the anchoring role of Germany as well as the risk sensitivities of countries like Spain, Portugal, Ireland, and particularly Italy. Allowing for heterogeneous, time-varying spatial dependence parameters also leads to an increase in both short- and long-run spillover risk measures, sometimes by a factor of more than three or four. Again, we stress that all such features would go unnoticed in homogeneous or static versions of the model such that it is empirically important to allow for both heterogeneity and time-variation in financial networks.

Our study contributes to several lines of literature. On the one hand, it opens an new avenue to investigate network properties in a much more flexible way than with the static models used in for instance Asgharian et al. (2013), Kou et al. (2018), Herskovic et al. (2020), and other references mentioned earlier. Given the empirical relevance of both heterogeneity and time-variation, it may also call for further theoretical advances of baseline frameworks such as in Denbee et al. (2021) to explain such time-variation. Our paper also links to the extant literature on modeling financial networks to study risk propagation and financial stability, such as for instance Acharya et al. (2014), Williamson (2003), Elliott et al. (2014), and Fernandes and Artes (2016). Recent literature on systemic risk highlights the importance of spatial dependence and network spillovers (e.g., Asgharian et al., 2013; Babii et al., 2019). Finally, methodologically our work relates to the literature on score-driven time-varying parameter models as in Creal et al. (2013). A similar score-driven approach is found in Blasques et al. (2016), Catania and Billé (2017), and Billé et al. (2019). Our work differs from theirs in that we allow for more dynamic heterogeneity across network players, which proves empirically highly important. As such, our new spatial specification could also be used to measure contagion effects and systemic risk; see Franklin and Douglas (2000); Billio et al. (2012); Acemoglu et al. (2015). In addition, our score mechanism is adapted to ensure spatial stability by construction. We are also aware of a parallel line of literature that measures network effects via vector autoregressive models and a variance decomposition; see Diebold and Yılmaz (2014). In contrast to this approach, the spatial approach allows one to include prior

information on connection strenghts and can, therefore, be more parsimonious.

The remainder of this paper is organized as follows. Section 2 discusses the new model and establishes its statistical properties. The data are discussed in Section 3, followed by the empirical results in Section 4. Section 5 presents robustness checks. Section 6 concludes.

2 The model

2.1 Modeling time-variation in heterogeneous spatial dynamics

Rather then building the model step-by-step, we provide the most general model specification first and then highlight how earlier models fit in as special. We gather the relevant measurements of a network into the vector $\boldsymbol{y}_t$. For instance, in our setting $\boldsymbol{y}_t$ contains spread changes of sovereigns in the Eurozone, but the measurements might similarly relate to for instance trade data, social network data, or firm relationships. We use the following vector-valued time-varying parameter spatial autoregressive (SAR) model for $\boldsymbol{y}_t$,

$$\boldsymbol{y}_t = \boldsymbol{R}_t \boldsymbol{W}_t \boldsymbol{y}_t + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} p_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t; \boldsymbol{\Sigma}, \nu), \tag{1}$$

where $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{N,t})^\top \in \mathbb{R}^{N \times 1}$ denotes the vector of cross sectional measurements for time $t = 1, \ldots, T$, where the cross-section dimension i relates to the network players, $\boldsymbol{\varepsilon}_t \in \mathbb{R}^{N \times 1}$ is a serially independently and identically distributed error term or vector of 'structural' shocks with a density $p_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t; \boldsymbol{\Sigma}, \nu)$ such as the normal distribution or alternatively a fat-tailed Student's t distribution. The density is characterized by a diagonal covariance matrix $\boldsymbol{\Sigma}$ and a static shape parameter ν. For now, we keep $\boldsymbol{\Sigma}$ fixed, but in Section 4 we also consider specifications with time-varying volatility. The matrix $\boldsymbol{X}_t \in \mathbb{R}^{N \times K}$ denotes a matrix holding exogenous regressors with corresponding static parameter vector $\boldsymbol{\beta} \in \mathbb{R}^{K \times 1}$, $\boldsymbol{W}_t \in \mathbb{R}^{N \times N}$ is an observable matrix containing the spatial weights, and $\boldsymbol{R}_t \in \mathbb{R}^{N \times N}$ is a diagonal matrix containing the unobserved time-varying heterogeneous spatial spillover parameters, also called spatial autoregressive coefficients. The spatial spillover parameters in $\boldsymbol{R}_t$ play a major role in the subsequent analysis. The model in equation (1) is already dynamic via the observed $\boldsymbol{W}_t$ matrix that measures the time-varying connections in the network in the standard way. On top of this, (1) also allows for further unobserved heterogeneity via the matrix $\boldsymbol{R}_t$.

Model (1) nests several models from the literature. For example, if $\boldsymbol{R}_t \equiv \rho \cdot \mathbf{I}_N$ for a scalar ρ with $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ the identity matrix, then (1) collapses to the standard spatial regression model; see for instance Anselin (2009), Asgharian et al. (2013), Kou et al. (2018), Denbee et al. (2021). For a static $\boldsymbol{R}_t = \boldsymbol{R}$, we obtain the model with static heterogeneous spillover strengths as in Aquaro

et al. (2021) and Herskovic et al. (2020). The static model $\boldsymbol{R}_t = \rho \cdot \mathbf{I}_N$ was generalized to a setting with a time-varying spatial autoregressive parameter ρ_t by Blasques et al. (2016) and Catania and Billé (2017). We also obtain this scalar dynamic model as a special case by setting $\boldsymbol{R}_t = \rho_t \cdot \mathbf{I}_N$. However, our new model provides a much richer description of the spatial dynamics by allowing $\boldsymbol{R}_t$ to be a non-scalar matrix. In particular, some network units may be important contributors to systemic risk at some times, whereas other units may take over this role at other times. This is captured by the different diagonal elements of $\boldsymbol{R}_t$, the ith element measuring the spatial sensitivity of unit i to the other units. The model can easily be extended further to include, for instance, a spatial lag structure for the error term, a non-diagonal covariance structure for $\boldsymbol{\varepsilon}_t$, or as mentioned earlier a time-varying covariance matrix $\boldsymbol{\Sigma}_t$ to allow for periods where the flow of information is high compared to other periods.

To model the time variation in $\boldsymbol{R}_t$, we endow $\boldsymbol{R}_t$ with score-driven dynamics as proposed in Creal et al. (2011, 2013) and Harvey (2013). The model can then be easily estimated via standard maximum likelihood methods. Score-driven models also possess information theoretic optimality properties and yield updates of the parameters that improve the expected Kullback-Leibler divergence; see Blasques et al. (2015) and Creal et al. (2020).

We assume that the errors $\boldsymbol{\varepsilon}_t$ follow a fat-tailed Student's t distribution with zero mean, covariance matrix $\boldsymbol{\Sigma}$, and ν degrees of freedom,

$$p_{\boldsymbol{\varepsilon}}\left(\boldsymbol{\varepsilon}_t\,;\boldsymbol{\Sigma},\nu\right) = \frac{\Gamma\left(\frac{1}{2}(\nu+N)\right)}{\Gamma\left(\frac{1}{2}\nu\right)\ |(\nu-2)\pi\boldsymbol{\Sigma}|^{1/2}}\ \left(1+\boldsymbol{\varepsilon}_t^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t/(\nu-2)\right)^{-0.5(\nu+N)}. \tag{2}$$

The normal distribution is recovered as a special case for $\nu \to \infty$. The problem with the normal distribution, however, is that it does not allow for incidentally large observations and fat tails in the data, which can be particularly problematic for financial economic data as used for instance in our application in Section 4.

We gather all time-varying spatial parameters into the vector $\boldsymbol{f}_t$, i.e.,

$$\boldsymbol{f}_t = \mathrm{diag}(\boldsymbol{R}_t) = \left(\boldsymbol{R}_{11,t},\ldots,\boldsymbol{R}_{NN,t}\right)^{\top}. \tag{3}$$

Note that unlike for instance Creal et al. (2013) or Blasques et al. (2016), we do not force the spatial *correlation* parameters $\boldsymbol{R}_{i,i,t}$ to be in the unit interval using for instance a logistic or tanh transformation. Though intuitive, such reparameterizations may come at a considerable cost. In fact, we show in the empirical application that such transforms can actually be a bad idea if there is heterogeneity in spillover sensitivity: it unnecessarily restricts the model's ability to capture the dynamics in the data and leads to biased estimates, as some $\boldsymbol{R}_{i,i,t}$ may go outside the 0–1 range without jeopardizing the spatial stability of the model as a whole. This is evident if one realizes

that for spatial stability we only require all eigenvalues of $\boldsymbol{R}_t\boldsymbol{W}_t$ to lie inside the unit circle. This still allows for $\boldsymbol{R}_{i,i,t} > 1$ for specific network players i, as long as the maximum eigenvalue of $\boldsymbol{R}_t\boldsymbol{W}_t$ remains below one. The latter depends on all elements of $\boldsymbol{R}_t$ and $\boldsymbol{W}_t$, and not on one specific $\boldsymbol{R}_{i,i,t}$ only. The fact that we allow $\boldsymbol{R}_{i,i,t}$ to move outside the unit interval therefore provides substantial flexibility to the model. This comes on top of the flexibility provided by allowing the diagonal elements of $\boldsymbol{R}_t$ to be different in the first place compared to the scalar model $\boldsymbol{R}_t = \rho_t \cdot \mathbf{I}_N$ of Blasques et al. (2016).

The score-driven dynamics are given by

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \boldsymbol{B}\,\boldsymbol{f}_t + \boldsymbol{A}_t\,\boldsymbol{s}_t, \tag{4}$$

$$\begin{aligned}
\boldsymbol{s}_t &= \left|\mathbf{I}_N - \boldsymbol{R}_t\boldsymbol{W}_t\right| \cdot \frac{\partial \log p_{\boldsymbol{y}}\left(\boldsymbol{y}_t \mid \boldsymbol{f}_t, \boldsymbol{\Sigma}, \nu\right)}{\partial \boldsymbol{f}_t} \\
&= \left|\boldsymbol{Z}_t\right| \cdot \left(w_t \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{i,t}\boldsymbol{y}_t^{\star\top}\right) - \operatorname{diag}\left(\boldsymbol{W}_t\left(\mathbf{I}_N - \boldsymbol{R}_t\boldsymbol{W}_t\right)^{-1}\right)\right), \\
&= \left|\mathbf{I}_N - \boldsymbol{R}_t\boldsymbol{W}_t\right| \cdot w_t \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{i,t}\boldsymbol{y}_t^{\star\top}\right) - \operatorname{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^{\star}\right),
\end{aligned} \tag{5}$$

$$\begin{aligned}
&\boldsymbol{y}_t^{\star} = \boldsymbol{W}_t\boldsymbol{y}_t, \qquad \boldsymbol{e}_t = \boldsymbol{y}_t - \boldsymbol{R}_t\boldsymbol{W}_t\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta}, \qquad w_t = \left(1 + \frac{N+2}{\nu-2}\right) \Big/ \left(1 + \frac{\boldsymbol{e}_t^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t}{\nu-2}\right), \\
&\boldsymbol{Z}_t = \mathbf{I}_N - \boldsymbol{R}_t\boldsymbol{W}_t, \qquad \boldsymbol{Z}_t^{\star} = \left|\boldsymbol{Z}_t\right| \cdot \boldsymbol{Z}_t^{-1},
\end{aligned}$$

such that $\boldsymbol{Z}_t^{\star}$ is the adjoint of $\boldsymbol{Z}_t$. Derivations and proofs can be found in the Appendix A.

The score is scaled in the sense of Creal et al. (2013) by $|\mathbf{I}_N - \boldsymbol{R}_t\boldsymbol{W}_t|$. This helps the spatial stability of the model by automatically reducing the step size when the maximum eigenvalue of $\boldsymbol{R}_t\boldsymbol{W}_t$ approaches one. Also note that we introduced a time-varying parameter $\boldsymbol{A}_t$ as a variation to the original score model of Creal et al. (2013). The parameter $\boldsymbol{A}_t = v_t \cdot \boldsymbol{A}$ equals a fixed $\boldsymbol{A}$ times a weight v_t. The default value of the weight is $v_t = 1$, in which case the dynamics reduce to the standard score-driven dynamics. If however $\boldsymbol{R}(\boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{A}\boldsymbol{s}_t) \cdot \boldsymbol{W}_{t+1}$ has a maximum eigenvalue outside the unit circle, the weight (and thus the step size) is reduced towards 0 until the moment that the maximum eigenvalue of $\boldsymbol{R}(\boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{f}_t + v_t\boldsymbol{A}\boldsymbol{s}_t) \cdot \boldsymbol{W}_{t+1}$ is ϵ inside the unit circle, where ϵ is set by the user. Starting from an initial $\boldsymbol{R}(\boldsymbol{f}_1)$ such that $\boldsymbol{R}(\boldsymbol{f}_1)\,\boldsymbol{W}_1$ has all eigenvalues inside the unit circle, these modified score dynamics ensure that the spatial stability of the model is (by design) never violated. This mechanism also in the theoretical derivations later on, as we can now directly exclude situations of spatial instability. In a way, it generalizes the univariate stability device $\rho_t = \epsilon \cdot \tanh(f_t)$ of Blasques et al. (2016) to the current heterogeneous spatial spillover context, without being overly restrictive.

The score in equation (5) consists of two terms that have an intuitive interpretation. The first term is the regressor $\boldsymbol{y}_t^\star$ weighted standardized error term $\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{i,t}$. It captures the score-driven dynamics for a standard regression model with a time-varying regression parameter and follows from the specification in (1), where the elements of $\boldsymbol{R}_t$ act as regression coefficients. This part of the score adjusts the parameters upwards or downwards depending on whether the most recent observations lie below or above the estimated regression line at the previous point in time. The second term in (5) corrects for the simultaneity bias in the regression specification, which is due to the endogeneity of $\boldsymbol{y}_t$ as a regressor. The endogeneity correction naturally becomes smaller if either $\boldsymbol{R}_t$ or $\boldsymbol{W}_t$ lie closer to zero, i.e., if the endogeneity problem is less. For $\nu < \infty$, the weight w_t provides a robustness feature to the time-varying parameter dynamics: if an observation $\boldsymbol{y}_t$ is an incidental outlier or influential observation with a large $\boldsymbol{e}_t$, the weight w_t tends to zero. Such observations thus receive less impact on the spatial correlation dynamics; see also similar features for other location and scale models in for instance Creal et al. (2013) and Harvey and Luati (2014). For the normal distribution, we have $w_t = 1$ and this robustness feature disappears.

2.2 Stationarity, ergodicity, and filter invertibility

In this section we study the asymptotic properties of the model as a data generating process (dgp) and as a filter. Compared to earlier work such as Blasques et al. (2016) or Blasques et al. (2022), we have two complications. These are the eigenvalue condition on $\boldsymbol{R}_t\boldsymbol{W}_t$ and the time-variation in $\boldsymbol{A}_t$. Regarding the latter, we note that the design of the weighted steps using $\boldsymbol{A}_t$ only enhances the stability of the model further. In particular, the stability of the unweighted version of the model (using $\boldsymbol{A}$) implies stability of the original model (using $\boldsymbol{A}_t$) as $\|\boldsymbol{A}_t\| \leq \|\boldsymbol{A}\|$ due to the weighting mechanism.

We make the following assumptions.

Assumption 1. *The distribution of the i.i.d. errors* $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ *is a standardized multivariate Student t distribution with* $\nu > 2$ *degrees of freedom and diagonal variance-covariance matrix* $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_{1,1}, \ldots, \boldsymbol{\Sigma}_{N,N})$*, with* $\inf_i \boldsymbol{\Sigma}_{i,i} \geq k$ *and* $\sup_i \boldsymbol{\Sigma}_{i,i} \leq K$*, with* $0 < k < K < \infty$ *and* $i = 1, \ldots, N$*.*

Assumption 2. *The exogenous regressors* $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ *are stationary and ergodic with* $\mathrm{E}[\|\boldsymbol{X}_t\|^{2+\delta}] \leq K < \infty$*.*

Assumption 3. *The time-varying spatial weight matrices* $\boldsymbol{W}_t$ *are observable, stationary and ergodic, and satisfy*

$$\limsup_{t\to\infty} \max\left\{\|\boldsymbol{W}_t\|_1, \|\boldsymbol{W}_t\|_\infty\right\} \leq K < \infty.$$

Assumption 4. *Define* $\mathcal{F}_t = \{\boldsymbol{f} \mid \varrho(\boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t) \leq 1-\epsilon\}$, *for some* $\epsilon > 0$, *where* $\varrho(\cdot)$ *denotes the spectral radius of a matrix and let*

$$\mathrm{E}\left[\log \sup_{\boldsymbol{f}\in\mathcal{F}_t} \|\boldsymbol{B} + \boldsymbol{A}\dot{\boldsymbol{s}}_t^{\boldsymbol{\varepsilon}}(\boldsymbol{f})\|_\infty\right] < 0,$$

where

$$\begin{aligned}
\dot{\boldsymbol{s}}_t^{\boldsymbol{\varepsilon}}(\boldsymbol{f}) &= -S_N^\top \left(\left(w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top \otimes \mathbf{I}_N\right) - (\mathbf{I}_N \otimes \boldsymbol{W}_t)\right)\dot{\boldsymbol{Z}}_t^\star(\boldsymbol{f})\left(\boldsymbol{W}_t^\top \otimes \mathbf{I}_N\right) S_N, \\
w_t^{\boldsymbol{\varepsilon}} &= \left(1+\frac{N+2}{\nu-2}\right)\Big/\left(1+\frac{\boldsymbol{\varepsilon}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t}{\nu-2}\right), \qquad \dot{\boldsymbol{Z}}^\star(\boldsymbol{f}_t) = \frac{\partial \mathrm{vec}\left(\boldsymbol{Z}^\star(\boldsymbol{f})\right)}{\partial \mathrm{vec}\left(\boldsymbol{Z}(\boldsymbol{f})\right)^\top}, \\
\dot{\boldsymbol{Z}}^\star(\boldsymbol{f})_{ij} &= -(1)^{i+j}\mathrm{trace}\left(\mathbf{I}_{-i,N}\ \left(\mathbf{I}_{-j,N}^\top\ \boldsymbol{Z}(\boldsymbol{f})\ \mathbf{I}_{-i,N}\right)^\star\ \mathbf{I}_{-j,N}^\top\right), \quad i,j = 1,\ldots,N,
\end{aligned}$$

with S_N *a selection matrix that selects the main diagonal from the vectorized matrix* $\mathrm{diag}(\boldsymbol{W}_t) = S_N^\top \mathrm{vec}(\boldsymbol{W}_t)$, *and* $\mathbf{I}_{-i,N}$ *is an* $N \times (N-1)$ *matrix obtained by deleting the* i^{th}*-column from* $\mathbf{I}_N$.

Assumptions 1–3 are rather standard. They ensure the regressors, weight matrices, and regression errors are well-behaved, with non-degenerate errors that have a finite covariance matrix. Assumption 4 is key to our result and ensures that the recursion for $\boldsymbol{f}_t$ under the dgp is contracting. It is important to note that the condition in Assumption 4 results in a non-degenerate region, as the supremum of $\dot{\boldsymbol{s}}_t^{\boldsymbol{\varepsilon}}$ over $\boldsymbol{f}$ is never degenerate: the derivative of the adjoint never explodes for any value of $\boldsymbol{f}$ as the adjoint is a simple sum of products of elements of $\boldsymbol{Z}_t(\boldsymbol{f}) = \mathbf{I}_N - \boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t$, and all elements of $\boldsymbol{Z}_t(\boldsymbol{f})$ are well-behaved given the supremum is taken over the set where all eigenvalues of $\boldsymbol{Z}_t(\boldsymbol{f})\boldsymbol{W}_t$ remain inside the unit circle. This is also clear from the explicit expression for $\dot{\boldsymbol{Z}}^\star(\boldsymbol{f}_t)_{ij}$.

Using these assumptions, we can prove strict stationarity and ergodicity of the stochastic matrix process $\{\boldsymbol{R}(\boldsymbol{f}_t)\}_{t\in\mathbb{Z}}$ that solves (4)–(5). We note that this result also implies that $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ generated by (1) is stationary and ergodic. The proof can be found in Appendix B. It is important to note here again that the elements of $\boldsymbol{R}(\boldsymbol{f}_t)$ may easily move outside the unit circle without jeopardizing the stability of the model, as long as $\varrho(\boldsymbol{R}(\boldsymbol{f}_t)\boldsymbol{W}_t)$ remains strictly smaller than one, which is ensured by design by the weighted score step mechanism involving $\boldsymbol{A}_t$. The empirical relevance of this flexibility is shown in Section 4.

Theorem 2.1. *Consider the model defined by* (1) *and* (4)–(5) *for a given i.i.d. sequence* $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ *and let Assumptions 1–4 hold true. If* (4)–(5) *are initialized at* $\boldsymbol{f}_1$ *with* $\varrho(\boldsymbol{R}(\boldsymbol{f}_1)\boldsymbol{W}_1) \leq 1-\epsilon$, *then the (initialized) solution* $\{\hat{\boldsymbol{f}}_t\}_{t\in\mathbb{N}}$ *to* (1) *and* (4)–(5) *converges exponentially fast almost surely (e.a.s.) to a unique (unititialized) strictly stationary and ergodic process* $\{\boldsymbol{f}_t\}_{t\in\mathbb{Z}}$. *As a result,* $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ *is also stationary and ergodic with a finite* $2+\delta$ *moment for some* $\delta > 0$.

Theorem 2.1 establishes the statistical properties of $\boldsymbol{y}_t$ and $\boldsymbol{f}_t$ if the model is correctly specified. When estimating the model, it is also important to establish the properties of the model when used as a filter rather than a dgp. In particular, we want to prove that the recursions (4)–(5) converge to a stationary and ergodic solution for stationary and ergodic inputs $\boldsymbol{y}_t$. This property is known as filter invertibility; see for instance Wintenberger (2013) and Blasques et al. (2022). It allows us to (asymptotically) recover the spatial correlation parameters consistently from the data even if the filter is initialized incorrectly. As the initial true value of $\boldsymbol{f}_1$ is never known with certainty, filter invertibility is an important property. To establish filter invertibility, we make the following additional assumption.

Assumption 5. *The parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^{dim(\boldsymbol{\theta})}$ is compact and satisfies $\nu \geq 2+\delta$ and $|\boldsymbol{A}| \neq 0$ for some $\delta > 0$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. It also holds that*

$$\mathrm{E}\left[\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\boldsymbol{\mathcal{F}}_t} \left\|\boldsymbol{B} + \boldsymbol{A}\ \dot{\boldsymbol{s}}_t(\boldsymbol{f})\right\|_\infty\right] < 0$$

where

$$\begin{aligned}\dot{\boldsymbol{s}}_t(\boldsymbol{f}) = &-\dot{w}_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})|\ \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)^\top \\ &- w_t(\boldsymbol{f})\ \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \mathrm{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)^\top \\ &- w_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})|\ \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t^\star\boldsymbol{y}_t^{\star\top}\odot\mathbf{I}_N\right) + S_N^\top\left(\mathbf{I}_N\otimes\boldsymbol{W}_t\right)\dot{\boldsymbol{Z}}^\star(\boldsymbol{f}_t)\left(\boldsymbol{W}_t^\top\otimes\mathbf{I}_N\right)S_N,\end{aligned}$$

$$\dot{w}_t(\boldsymbol{f}) = -\left(1+\frac{N+2}{\nu-2}\right)\Big/\left(1+\frac{\boldsymbol{e}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t}{\nu-2}\right)^2,$$

and $\dot{\boldsymbol{Z}}_t^\star = \partial\mathrm{vec}(\boldsymbol{Z}^\star)/\partial\mathrm{vec}(\boldsymbol{Z})^\top$ is defined in Assumption 4.

We again note that there are no explosive elements in the definition of $\dot{\boldsymbol{s}}_t(\boldsymbol{f})$ for $\boldsymbol{f} \in \boldsymbol{\mathcal{F}}_t$. This provides a non-trivial region over which the filter is contracting. We can now prove the following result.

Theorem 2.2. *Let $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ be generated by (1) and (4)–(5) for some $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. Consider the filtering equations defined by (4)–(5) at some arbitrary $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and let Assumptions 1–5 hold true. If the filter is initialized at $\boldsymbol{f}_1$ with $\varrho(\boldsymbol{R}(\boldsymbol{f}_1)\boldsymbol{W}_1) \leq 1-\epsilon$, then the (initialized) solution $\{\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ to these filtering equations converges exponentially fast almost surely (e.a.s.) to a unique (uninitialized) strictly stationary and ergodic process $\{\boldsymbol{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, uniformly over the compact parameter space $\boldsymbol{\Theta}$,*

i.e.,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \hat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta}) \right\|_\infty \xrightarrow{e.a.s.} 0.$$

The effect of a fixed initialization of the filter at $\boldsymbol{f}_1$ thus wears of exponentially fast. This allows us to recover the true $\boldsymbol{f}_t$ from the data $\boldsymbol{y}_t$ and the true parameter value $\boldsymbol{\theta}_0$. The next section explains how to estimate $\boldsymbol{\theta}_0$ by maximum likelihood.

2.3 Estimation

Score-driven models can easily be estimated by maximum likelihood methods via a standard prediction error decomposition. Gathering all static parameters in the vector $\boldsymbol{\theta}$, we obtain the objective function

$$\mathcal{L}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log p_{\boldsymbol{y}}(\boldsymbol{y}_t \mid \boldsymbol{f}_t, \boldsymbol{\theta}) \equiv \sum_{t=1}^{T} \log p_{\boldsymbol{y}}(\boldsymbol{y}_t \mid \boldsymbol{R}_t, \boldsymbol{\Sigma}, \nu). \tag{6}$$

To compute the likelihood, one proceeds as follows. Given a value of $\boldsymbol{\theta}$ and an initial value $\boldsymbol{f}_1$, one obtains the time-varying parameter values $\boldsymbol{f}_t$ for all times $t = 1, \ldots, T$ using the recursion (4). With these values of $\boldsymbol{f}_t$, we obtain the values of $\boldsymbol{R}_t = \boldsymbol{R}(\boldsymbol{f}_t)$. These can then directly be plugged into (6) to obtain the value of the log-likelihood function. The initial $\boldsymbol{f}_1$ is obtained by estimating a static version of the model on the initial 2 years of observations.

We obtain the maximum likelihood estimator as $\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} \mathcal{L}_T(\boldsymbol{\theta})$. Given stationarity and ergodicity and filter invertibility as established in the previous section, consistency and asymptotic normality of the maximum likelihood estimator can be obtained along the lines of for instance Blasques et al. (2022). We omit such a proof, as it contains no novel arguments when applied to the model studies in this paper.

We estimate covariance matrix of $\hat{\boldsymbol{\theta}}_T$ in the usual way as

$$\hat{\boldsymbol{V}}_T = \hat{\boldsymbol{H}}_T^{-1} \cdot \hat{\boldsymbol{J}}_T \cdot \hat{\boldsymbol{H}}_T^{-1}, \quad \hat{\boldsymbol{J}}_T = \sum_{t=1}^{T} \frac{\mathrm{d} \log p_{\boldsymbol{y}}(\boldsymbol{y}_t \mid \boldsymbol{f}_t, \hat{\boldsymbol{\theta}}_T)}{\mathrm{d}\boldsymbol{\theta}} \frac{\mathrm{d} \log p_{\boldsymbol{y}}(\boldsymbol{y}_t \mid \boldsymbol{f}_t, \hat{\boldsymbol{\theta}}_T)}{\mathrm{d}\boldsymbol{\theta}^\top}, \quad \hat{\boldsymbol{H}}_T = -\frac{\partial^2 \mathcal{L}_T(\hat{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}, \tag{7}$$

where the computation of the outer-product-of-gradients $\hat{\boldsymbol{J}}_T$ uses the total rather than the partial derivatives. If the model is correctly specified, $\hat{\boldsymbol{H}}_T - \hat{\boldsymbol{J}}_T$ converges to zero and the covariance matrix can be estimated by $\hat{\boldsymbol{V}}_T = \hat{\boldsymbol{H}}_T^{-1}$.

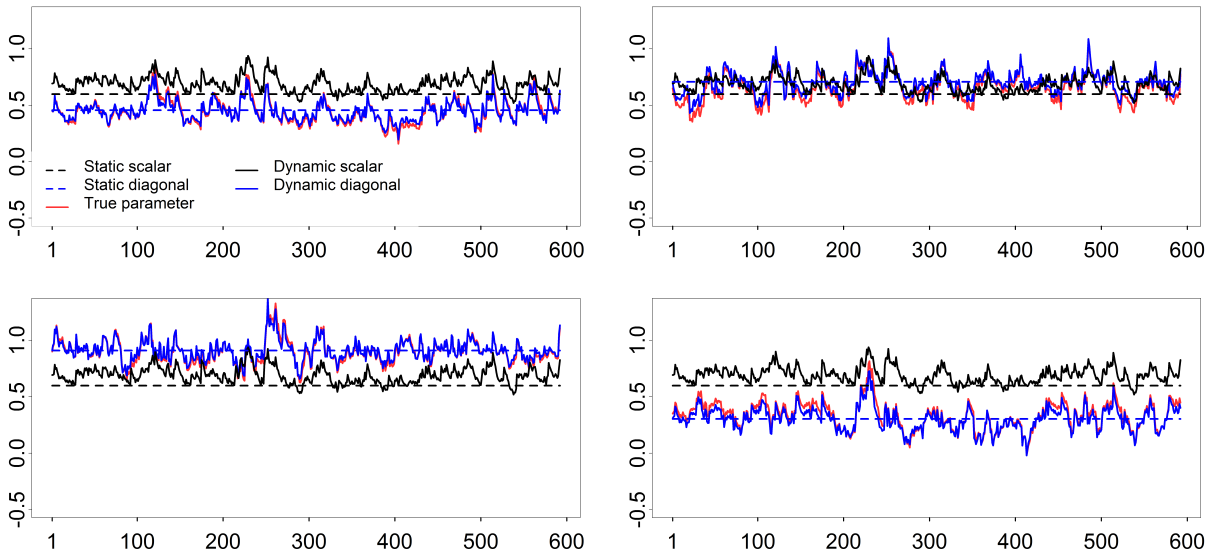


Figure 1: Simulated true and fitted spatial dependence parameters

Each panel shows the simulated and estimated paths of one out of $N = 4$ spatial autoregressive parameters. The true data generating process has heterogeneous time varying spatial parameters with unconditional means equal to 0.45 (top-left), 0.65 (top-right), 0.90 (bottom-left), and 0.35 (bottom right), respectively. The four models estimated are the static scalar spatial autoregression ($\boldsymbol{R}_t = \rho \cdot \mathbf{I}_N$), the static diagonal model ($\boldsymbol{R}_t = \boldsymbol{R}$), the dynamic scalar model of Blasques et al. (2016) ($\boldsymbol{R}_t = \rho_t \cdot \mathbf{I}_N$), and the new diagonal spatial model (diagonal $\boldsymbol{R}_t$).

2.4 Simulation evidence

To investigate the properties of the new model and the effects of heterogeneity in the spatial autoregressive parameters, we report the results of a small simulation experiment. The experiment is set up as follows. We consider a data generating process (dgp) with dynamic, heterogeneous spatial parameters $\boldsymbol{R}_t$. We use a setting with fat-tailed structural shocks ($\nu = 5$), no regressors ($\boldsymbol{\beta} = 0$), and $N = 4$. We simulate around 600 observations, similar in magnitude to the number of time series observations in the application. For $\boldsymbol{W}_t$, we use the row-normalized empirical spatial weight matrices $\boldsymbol{W}_t$ for Germany, France, Italy, and the Netherlands from Section 4. For concreteness, we set $\boldsymbol{A} = 0.07 \cdot \mathbf{I}_N$, $\boldsymbol{B} = 0.9 \cdot \mathbf{I}_N$, and $(\mathbf{I}_N - \boldsymbol{B})^{-1}\boldsymbol{\omega} = (0.45, 0.65, 0.90, 0.35)^\top$. Qualitatively similar results are obtained for other parameter settings.

We estimate four different models: a static scalar model (ρ), a dynamic scalar model (ρ_t), a static diagonal model ($\boldsymbol{R}$), and a dynamic diagonal model ($\boldsymbol{R}_t$). This allows us to clearly see the separate contributions of dynamics and heterogeneity. Figure 1 reports the results for a typical simulation. The simulated time-variation and heterogeneity are actually quite modest compared to the empirical patterns found in the data in Section 4, where heterogeneity and dynamics are even more prevalent.

Figure 1 shows that the correctly specified model (solid blue curve) captures the movements in

the dgp (solid red) highly accurately for sample sizes similar to those in the empirical application.[1] The static version of the model $\boldsymbol{R}_t = \boldsymbol{R}$ (dashed blue) with heterogeneity but without dynamics still captures the unconditional level of the true time-varying spatial correlations, but misses all of the dynamics which can be up to a magnitude of 0.4, which is sizeable for a spatial correlation parameter. The standard static scalar spatial autoregression of Anselin (2009) with $\boldsymbol{R}_t = \rho \cdot \mathbf{I}_N$ fares even worse. It has to balance the different levels of the spatial autocorrelations both in the cross section and over time and therefore lands at some sort of average spatial autocorrelation of ρ around 0.55 to 0.6 (dashed black).

We see a similar bias for the dynamic scalar spatial regression model (solid black) of Blasques et al. (2016). The true spatial correlations (in red) are below the black solid curve for the top-left and bottom-right panels, whereas it is above and on top of the black curve in the other two panels. Again, given the heterogeneity in the true spatial correlations, the scalar model can only take some kind of average of these, both in terms of the overall level and in terms of the specific dynamics over time. For the latter, we note for instance the peaks of similar magnitude in the black curve in the bottom-left graph around observations 230 and 260, respectively. The true (in red) dynamics of $\boldsymbol{R}_{i,i,t}$ for this cross-sectional unit, however, only have a peak around observation 260, and not around time 230. The latter peak in the black curve appears to be attributable to the bottom-right cross-sectional unit, which clearly peaks around observation 260, but not around the earlier observation 230. As the heterogeneous dynamics of all series have to be captured by one single ρ_t for the scalar dynamic model, a mixed-up message emerges about what is actually happening in the data. Our new model with heterogeneous, diagonal $\boldsymbol{R}_t$ avoids all these issues and captures the cross-sectional and time-series heterogeneity in the dynamics much better.

3 Data

We illustrate the model by applying it to analyze the network dynamics of sovereign systemic risk in the Eurozone. We consider seven countries: Germany, France, Ireland, Italy, Netherlands, Portugal, and Spain. Our sample period spans from December 10, 2009 until July 2, 2020. As our dependent variable y_t we consider the changes of 5-year government yield spreads, defined as the difference between the euro-denominated 5-year government bond yield and the 5-year EONIA OIS rate. This allows us to concentrate fully on the credit risk component of sovereign bonds. We also provide a robustness analysis using a 1-year rather than 5-year maturity, as well as using CDS spreads rather than bond yield spreads; see Section 5. Results are qualitatively similar. Data are

[1] Additional unreported simulations show that the new model can also adequately track the true model parameters even in cases where the statistical model is misspecified. This is in line with the theoretical results in Blasques et al. (2015) and Creal et al. (2020).

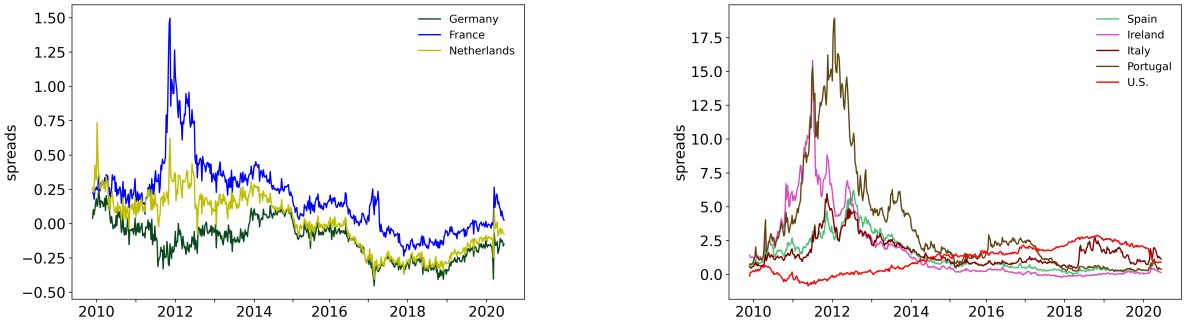


Figure 2: Government bond yield spreads of seven Eurozone sovereigns and the U.S.

Weekly euro-denominated 5-year government yield spreads of seven European countries and the United States from December 2009 - July 2020.

taken from Datastream and Bloomberg, and the sample contains 543 weekly observations for each country.

We base our main analysis on changes rather than log changes in 5-year government yield spreads. Log changes can lead to non-intuitive results if spreads hover near zero, which is the case in our sample where central banks have kept key rates artificially low for a long period. In such cases, taking the logarithmic change of near-zero spreads can result in large numbers, which would for instance lead to Germany becoming the most volatile country in the sample. Taking normal changes instead, this problem is avoided and results in Germany being one of the least volatile economies in the Eurozone, a result that is much more intuitive.

We base our main analysis on 5-year maturity bond yields rather than CDS spreads as in Blasques et al. (2016). Particularly during the later years in the sample, sovereign CDSs are less liquidly traded compared to their underlying bonds. As a result, all CDS spreads hover more closely to zero with little variation in the estimates of the spatial correlation parameters. Section 5 provides two robustness checks of the main analysis: one using 5-year maturity CDS spreads, and one using 1-year maturity bond yields.

Figure 2 presents the 5-year government bond yield spreads. All countries in our sample tend to co-move, especially during the 2010–2012 European sovereign debt crisis, with the exception of Germany and to some extent the Netherlands. Whereas most yields go up during the crisis, those of Germany go down, Germany acting as an anchor and safe-haven for the euro during that time. Note that the right-hand panel has a different vertical scale, indicating that the countries in that panel have much higher government bond yield spread levels compared to the countries in the left-hand panel. This makes sense, as these were precisely the countries that were at the center of the crisis.

Table 1 reports summary statistics. We see high volatilities for countries such as Portugal,

Table 1: Summary statistics 5-year government bonds

This table lists descriptive statistics of changes of 5-year government bond yield spreads of seven European countries and the United States. The yield spread is defined as the difference between the 5-year government bond yield and the 5-year EONIA OIS rate. In addition, we list summary statistics of stock returns and changes in the VSTOXX index and term spread. The term spread is computed as the absolute change between the 10-year and 1-year government bond yield. All variables are weekly from December 10, 2009 until July 2, 2020.

	mean	s.d.	med	min	max	skew	kurt
5 year gov bond spread changes							
Germany	0.000	0.052	0.001	-0.20	0.33	0.30	7.38
Spain	0.000	0.180	-0.003	-1.07	0.82	-0.22	8.64
France	0.000	0.069	-0.001	-0.64	0.35	-0.95	19.94
Ireland	-0.002	0.361	0.001	-3.07	2.76	-0.90	29.65
Italy	0.001	0.196	-0.001	-0.93	0.99	0.52	8.29
Netherland	-0.001	0.062	0.000	-0.27	0.41	0.72	10.90
Portugal	0.000	0.505	-0.004	-3.31	3.95	0.30	21.44
U.S.	0.002	0.088	0.004	-0.40	0.27	-0.43	4.50
Stock index returns							
Germany	0.001	0.029	0.004	-0.22	0.10	-1.31	11.62
Spain	-0.001	0.032	0.002	-0.23	0.10	-0.92	8.54
France	0.000	0.028	0.003	-0.22	0.10	-1.28	11.24
Ireland	0.001	0.026	0.003	-0.20	0.08	-1.47	11.18
Italy	0.000	0.033	0.002	-0.26	0.10	-1.34	11.20
Netherlands	0.001	0.026	0.003	-0.20	0.09	-1.50	12.59
Portugal	-0.001	0.029	0.001	-0.20	0.07	-1.11	7.92
Gov. yield spread changes (10-year - 1-year)							
Germany	-0.004	0.081	-0.011	-0.278	0.523	0.994	7.44
Spain	-0.005	0.224	0.000	-3.732	1.175	-8.107	144.5
France	-0.004	0.083	-0.007	-0.335	0.578	0.838	8.11
Ireland	-0.005	0.279	-0.012	-2.283	2.628	0.790	30.65
Italy	-0.002	0.154	-0.008	-0.855	1.125	1.022	14.74
Netherlands	-0.006	0.108	-0.009	-0.890	0.828	-0.112	22.57
Portugal	-0.003	0.431	-0.007	-2.475	2.536	-0.035	13.00
Control variable							
Δ VSTOXX	-0.001	3.906	-0.150	-16.98	31.31	1.68	16.81

Ireland, Italy and Spain compared to Germany. The kurtosis is high for all series, indicating that the data are (unconditionally) peaked and that a combination of volatility clustering and conditional non-normality may be called for. Our model used in Section 4 will account for both.

We use three control variables from the literature; see Blasques et al. (2016). To control for economic conditions, we include the (log) returns in each country's main stock index as well as the slope of the term structure (government bond yield 10-year minus 1-year) for each country. We also control for market stress by including the changes in the VSTOXX index. The VSTOXX index measures volatility in options markets and is therefore a forward looking measure of market stress and investor sentiment. Being a European index, it directly relates to the local Eurozone market developments and its risk perception and appetite.

For the spatial weight matrices $\boldsymbol{W}_t$, we choose domestic banks cross-border exposures as in Tonzer (2015) using data extracted from the BIS website.[2] A small number of missing values are back-filled by the last available observations. The consolidated statistics in the BIS database cover the positions of international bank cross border holdings and thus provide a measure of interconnectedness of the financial system in the Eurozone. Financial system interconnectedness can be regarded as one of the prominent determinants for sovereign credit risk spillovers given the bail-out incentive of governments for their local banking sector, and its potential effect on government creditworthiness. The quarterly data on cross-exposures are converted to weekly values by taking the latest value of the BIS data available in that week. We lag the entries by two quarters to prevent endogeneity concerns and the use of contemporaneous or future information in $\boldsymbol{W}_t$. The spatial weight matrices are row-normalized.[3]

4 Empirical application

In this section we discuss the empirical results. Section 4.1 discusses the parameter estimates and the filtered correlations for the different models. Section 4.2 presents their implications for the risk spillovers between countries.

4.1 Spatial dependence dynamics in the Eurozone

Table 2 presents the parameter estimates for four alternative model specifications. Given that the data exhibits volatility clustering, we augment the original model from Section 2 with a score-driven dynamic covariance matrix $\boldsymbol{\Sigma}_t$ as explained in Appendix A. The first two columns in Table 2 relate to the static and dynamic model with homogeneous spatial dependence parameter ρ_t. The last two columns correspond to the models with heterogeneous spatial autoregressive parameters $\boldsymbol{R}_t$. For parsimony, the dynamic version of the model with $\boldsymbol{R}_t$ uses a pooled common persistence parameter $\boldsymbol{B} = \beta_\rho \cdot \mathbf{I}_N$ and heterogeneous adjustment speeds, $\boldsymbol{A} = \text{diag}(\alpha_1, \ldots, \alpha_N)$, with α_i the adjustment speed of $\boldsymbol{R}_{i,i,t}$. For the variances in $\boldsymbol{\Sigma}_t$, we use a common persistence parameter β_{vol} and adjustment speed α_{vol}. This specification is supported by the data. Allowing for different volatility adjustment parameters α_{vol} or persistence parameters β_ρ or β_{vol} does not impact the main results in any substantial way.

[2] https://stats.bis.org/, retreived July 17, 2020. Like Blasques et al. (2016), we use the immediate counterparty risk measure of banking groups' country risk exposures.

[3] In the empirical application we have also considered alternative ways of scaling $\boldsymbol{W}_t$. Examples include the sample maximum (over all t) of the maximum eigenvalue of $\boldsymbol{W}_t$ to account for any possible time-variation in $\boldsymbol{W}_t$ spilling over into $\boldsymbol{R}_t$. The effects are limited. Though the *level* of the $\boldsymbol{R}_{i,i,t}$ changes (as expected), their heterogeneity and dynamic behavior remain largely unaffected. In particular, the risk measures as presented in Section 4.2 remain highly similar.

Spatial heterogeneity turns out to be empirically important: allowing for heterogeneous spatial spillover parameters $\boldsymbol{R}_t$ compared to a single ρ_t considerably improves the fit of the model. The log-likelihood increases by roughly 195 points in the static case, and by around 214 points for the dynamic case. Allowing for dynamic (ρ_t or $\boldsymbol{R}_t$) rather than static (ρ or $\boldsymbol{R}$) spatial dependence also increases the fit of the model. The log-likelihood increases by 42 points in the scalar case, and by 61 points for the heterogeneous model. The value-added of the dynamic specification thus comes out more clearly in the heterogeneous case. This is in line with the simulation results from Section 2.4: not only the levels, but also the dynamic patters are corrupted if we incorrectly pool spillover parameters across countries.

The importance of the combination of both heterogeneity and dynamics clearly shows in the plots of the filtered estimates of ρ_t and $\boldsymbol{R}_{i,i,t}$ in Figure 3. These filtered estimates are easily obtained by evaluating $\boldsymbol{f}_t$ for $t = 1, \ldots, T$ at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, and then inserting $\boldsymbol{f}_t(\hat{\boldsymbol{\theta}})$ into equation (3). The left-hand panel in Figure 3 shows the results for the scalar model. The static estimate of ρ of slightly below 0.5 corresponds to the long-run average of the dynamic scalar ρ_t. The variation of ρ_t over time is quite modest between roughly $\rho_t = 0.2$ and $\rho_t = 0.7$.

The picture changes dramatically if we consider the heterogeneous spatial spillover parameters $\boldsymbol{R}_{i,i,t}$ in the right-hand panel of Figure 3. Note that the left and right panel in the figure have the same vertical scale. Whereas for the static scalar model the single value of ρ is about 0.5, for the static heterogeneous model the $\boldsymbol{R}_{i,i}$ range between a low value around 0 for Germany, to a high value of 1.1 for Italy.

If we also allow for dynamics in the heterogeneous spillover parameters $\boldsymbol{R}_{i,i,t}$, the differences become even clearer. Whereas the scalar dynamic model shows a modest variation from about 0.2 to 0.7, the dynamic heterogeneous spillover parameters exhibit a much larger variation. For instance, for Italy $\boldsymbol{R}_{i,i,t}$ ranges from a low 0.25 at the start of the sample, to a high 1.5 in the midst of the European sovereign debt crisis.

We also see that allowing for heterogeneity substantially impacts the dynamics of $\boldsymbol{R}_{i,i,t}$ compared to the scalar ρ_t case and makes them much more varied. Some countries exhibit a large time-variation, with quite different timing of the peaks and troughs. Other countries appear much more stable over time, despite the crisis years. For instance, the $\boldsymbol{R}_{i,i,t}$ for Germany tightly hovers between -0.1 and 0.25. The fact that its $\boldsymbol{R}_{i,i,t}$ becomes slightly negative at times corroborates Germany's role as an anchor and stabilizer for the Eurozone, a feature that would be completely hidden in a scalar homogeneous model. For the dynamic scalar model, ρ_t remains rather stable until mid-2011: it averages the highly opposite upward movements of Italy and to a lesser extent Ireland and the Netherlands, versus the downward movements of Spain, Portugal, France, and Germany. Only around the time that Italy also starts its descent, the scalar ρ_t starts to react and

Table 2: Parameter estimates: 5-year bond yields

This table reports the estimated parameters of four spatial dependence models, applied to weekly first differences in the spread between 5-year government bond yields and the 5-year EONIA. Results are for seven Eurozone countries. Robust (Huber sandwich) standard errors are reported in parentheses. The models are based on Student's t distributed disturbances with time-varying heteroscedasticity as in model (1)–(5). For the diagonal models, we have $\boldsymbol{B} = \beta \cdot \mathbf{I}_N$ and $\boldsymbol{A} = \text{diag}(\alpha_1, \ldots, \alpha_N, \alpha_{vol}, \ldots, \alpha_{vol})$ We report the maximum log-likelihood value (LogLike) and AIC(Akaike information criterion). The sample runs from December 2009 - July 2020. The table is continued on the next page.

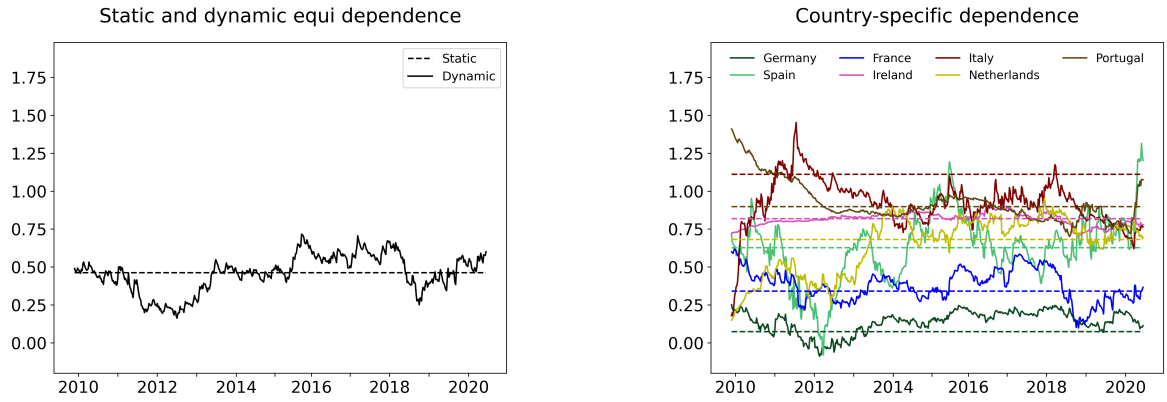
	Static Scalar	Dynamic Scalar	Static Diagonal	Dynamic Diagonal
Panel A: Spatial dependence parameters				
ω/ω_{GE}	0.463	0.009	0.074	0.001
	(0.022)	(0.005)	(0.031)	(0.002)
ω_{SP}			0.628	0.011
			(0.060)	(0.006)
ω_{FR}			0.341	0.006
			(0.026)	(0.003)
ω_{IR}			0.819	0.014
			(0.043)	(0.006)
ω_{IT}			1.111	0.028
			(0.108)	(0.011)
ω_{NE}			0.682	0.013
			(0.038)	(0.005)
ω_{PO}			0.898	0.015
			(0.069)	(0.007)
α/α_{GE}		0.005		0.010
		(0.001)		(0.004)
α_{SP}				0.049
				(0.017)
α_{FR}				0.011
				(0.004)
α_{IR}				0.009
				(0.012)
α_{IT}				0.099
				(0.044)
α_{NE}				0.020
				(0.010)
α_{PO}				0.020
				(0.018)
β_ρ		0.982		0.984
		(0.010)		(0.007)
Panel B: Volatility and control parameters				
(see next page)				

go down mildly.

One of the striking patterns in Figure 3b is the repeated exceedance of $\boldsymbol{R}_{i,i,t}$ above the threshold 1 for countries like Italy, Portugal, and Spain. We stress that the model remains spatially stable

(Table 2 continued)

	Static Scalar	Dynamic Scalar	Static Diagonal	Dynamic Diagonal
Panel B: Volatility and control parameters				
α_{vol}	0.276	0.271	0.289	0.278
	(0.026)	(0.026)	(0.027)	(0.029)
β_{vol}	0.998	0.998	0.998	0.998
	(0.002)	(0.002)	(0.002)	(0.003)
ν	3.483	3.540	4.015	4.099
	(0.326)	(0.334)	(0.412)	(0.429)
Δ VSTOXX	0.004	0.010	-0.001	0.004
	(0.032)	(0.029)	(0.030)	(0.026)
Δ term spread	0.058	0.057	0.037	0.041
	(0.049)	(0.045)	(0.046)	(0.041)
local stock	-0.022	-0.021	-0.019	-0.019
	(0.011)	(0.010)	(0.010)	(0.009)
logLik	4454	4496	4649	4710
AIC	-8865	-8946	-9244	-9350



(a) (b)

Figure 3: Filtered spatial autoregressive parameters

This figure plots the filtered spatial spillover parameters for four different models, applied to euro-denominated 5-year government bond yield spreads of seven European countries. The left figure shows the results for the static and time-varying scalar spillover model, whereas the right figure shows the results for the static and dynamic heterogeneous spillover models. The sample runs from December 2009 - July 2020.

in all these cases given that the maximum eigenvalue of $\boldsymbol{R}_t\boldsymbol{W}_t$ always remains far below 1.

Figure 3b as well as Table 2 also underline the importance of allowing for differences in adjustment speeds α_i on top of the differences in the long-run levels of $\boldsymbol{R}_{i,i,t}$. Countries like Germany, France, and Ireland have a low α_i, indicating that time-variation in $\boldsymbol{R}_{i,i,t}$ for those countries is modest in size and rather gradual. By contrast, other countries such as Spain and particularly

Italy require a much larger value of α_i as their spatial spillover changed much more rapidly during the periods of stress in the sample. Not allowing for this type of heterogeneity would again substantially distort the empirical patterns in the data and result in too much stability of $\boldsymbol{R}_t$ as the model would have to average the different adjustment speeds.[4]

Some final highlights can be found in Panel B of Table 2. We see that none of the control variables are statistically significant, a feature that we further corroborate in our robustness analyses in Section 5. We also see that it is important to allow for fat tails and volatility clustering. The persistence of volatility is high with a β_{vol} close to one, whereas α_{vol} is also strongly significant.[5] The degrees of freedom parameter ν lies between 3.5 and 4, which is quite low given that we also control for volatility clustering: it indicates that changes in bond yield spreads have fat tails and that 4th order (conditional) moments may not exist. Again, this does not result in problems for our score-driven model, unlike for a standard GARCH model where typically 4th order moments are needed for consistency; see Blasques et al. (2022) for details on the asymptotics of score-driven models.

Summarizing, the statistical improvements by allowing for heterogeneous and time-varying spatial spillovers in European sovereign credit risk changes are statistically important. The new model also results in more details on the empirical patterns and better insight in the spatial spillover dynamics compared to models that either lack time-varying parameters or models that impose homogeneity. One should, however, be cautious in over-interpreting the dynamics of ρ_t or $\boldsymbol{R}_{i,i,t}$ on their own: in terms of systemic risk implications, the $\boldsymbol{R}_{i,i,t}$ mix with the spatial weight matrix $\boldsymbol{W}_t$ and the time-varying volatilities $\boldsymbol{\Sigma}_t$ to form spatial risk spillovers. In the next subsection, we therefore put all these components together and consider them jointly to study the models' systemic risk measurement implications.

4.2 Heterogeneity in Eurozone systemic risk spillovers

To study the risk implications of our model, we consider the effect of a one-standard-deviation shock $\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_t^{1/2}\, e_j$ to the system, where e_i $(i = 1, \dots N)$ denotes the i-th column of the identity matrix $\mathbf{I}_N$. The effect is given by

$$\begin{aligned}
&\boldsymbol{y}_t = \boldsymbol{R}_t \boldsymbol{W}_t \boldsymbol{y}_t + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\Sigma}_t^{1/2}\, e_j \quad \Leftrightarrow \quad (\mathbf{I}_N - \boldsymbol{R}_t \boldsymbol{W}_t)\, \boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\Sigma}_t^{1/2}\, e_j \quad \Leftrightarrow \\
&\boldsymbol{y}_t = (\mathbf{I}_N - \boldsymbol{R}_t \boldsymbol{W}_t)^{-1} \left(\boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\Sigma}_t^{1/2}\, e_j \right)
\end{aligned}$$

[4]Note that the persistence parameters β are the same for all $\boldsymbol{R}_{i,i,t}$. Unlike the case for α_i, allowing for different β_i does not increase the likelihood by much. Also, all β_i remain close to one in that case, such that little is lost by pooling these parameters as opposed to the result for the α_is.

[5]Note that in a score-driven model we need not have $\alpha_{vol} + \beta_{vol} < 1$ for stationarity as in a GARCH context, but rather $\beta_{vol} < 1$ only; see for instance Blasques et al. (2022).

$$= \left(\mathbf{I}_N + \boldsymbol{R}_t\boldsymbol{W}_t + (\boldsymbol{R}_t\boldsymbol{W}_t)^2 + (\boldsymbol{R}_t\boldsymbol{W}_t)^3 + \ldots\right)\left(\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\Sigma}_t^{1/2}\, e_j\right), \tag{8}$$

as long as the largest absolute eigenvalue of $\boldsymbol{R}_t\boldsymbol{W}_t$ is smaller than one. It is clear that not only $\boldsymbol{R}_t$, but also $\boldsymbol{W}_t$ and $\boldsymbol{\Sigma}_t$ play a key role in this transmission mechanism. To summarize the impact, we construct two different spatial impulse response based risk measures: a short-run (SR) and a long-run (LR) risk spillover measure, defined as

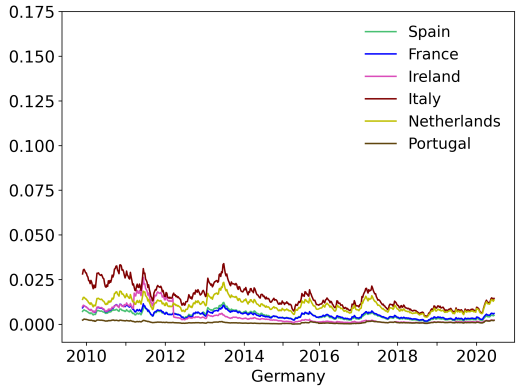
$$SR_{i,j,t} = e_i^\top\ \boldsymbol{R}_t\ \boldsymbol{W}_t\ \boldsymbol{\Sigma}_t^{1/2}\ e_j = \boldsymbol{\Sigma}_{j,j,t}^{1/2}\ \boldsymbol{R}_{i,i,t}\ \boldsymbol{W}_{i,j,t} \tag{9}$$

$$LR_{i,j,t} = e_i^\top\ (\mathbf{I}_N - \boldsymbol{R}_t\ \boldsymbol{W}_t)^{-1}\ \boldsymbol{\Sigma}_t^{1/2}\ e_j = \boldsymbol{\Sigma}_{j,j,t}^{1/2}\ (\mathbf{I}_N - \boldsymbol{R}_t\ \boldsymbol{W}_t)^{-1}_{i,j}, \tag{10}$$

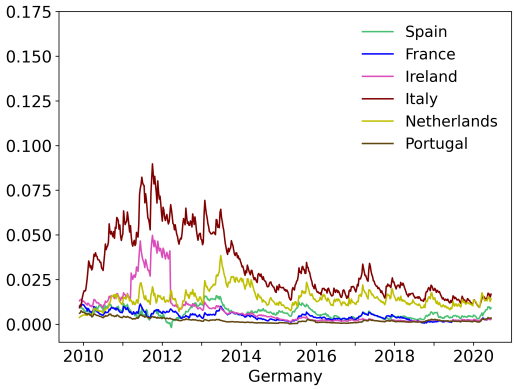
respectively. The first measure computes the first-order spatial spillover effect of a one-standard-deviation structural shock to country j on country i. The second measure calculates the compound or reduced form effect of such a structural shock. As there are 7×7 possible combinations of countries and two risk measures, we only report a subset of the results to highlight the rich empirical patterns that can be obtained by allowing for both heterogeneous and dynamic spatial spillover parameters $\boldsymbol{R}_{i,i,t}$. We focus on structural shocks to three large Eurozone countries: Germany, Spain, and Italy. To save space, we only consider the dynamic versions of all models. Results are presented in Figures 4 and 5 for $SR_{i,j,t}$ and $LR_{i,j,t}$, respectively.

The left and right panels in Figures 4 and 5 relate to the scalar (ρ_t) and heterogeneous ($\boldsymbol{R}_t$) model, respectively. All figures reflect a clear decrease in volatility ($\boldsymbol{\Sigma}_{j,j,t}$) towards the end of the sample for all countries considered, which results in smaller (one-standard-deviation) shocks. Looking at Figure 4a, we note that the impact of a one-standard-deviation shock to Germany in the scalar model erroneously appear modest and comparable across all other countries. Allowing for heterogeneous spillovers in Figure 4b changes this picture dramatically. It now becomes clear that particularly Italy has a sizeable reaction to German shocks during the crisis. All other sensitivities are an order of magnitude smaller compared to Italy, with the exception of Ireland during 2011. This is intuitive, Italy at the time being one of the largest European economies in potential distress (Romano, 2021), and Germany being the stable anchor economy. Also note that the effect on Italy is up to three times as large as for the scalar model.

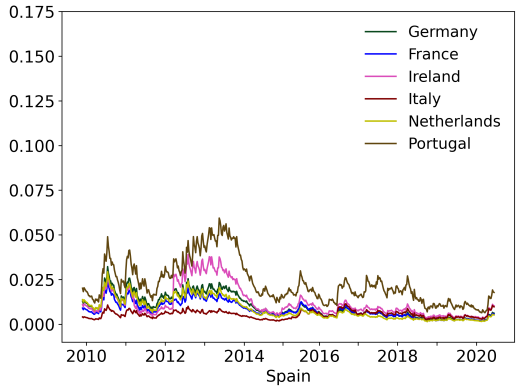
We see a similar effect in Figures 4c and 4d. For the pooled, scalar ρ_t model the effects on all other countries erroneously appear modest and comparable. If we allow for heterogeneity via $\boldsymbol{R}_t$, however, Spanish shocks mainly reflect on Portugal and are up to triple the size of the scalar model's effects. This again makes intuitive sense given the close connection of these economies. Second in line are Ireland and Italy, which were also at the center of the European sovereign debt crisis. The other countries only follow at a larger distance. Also note that Germany reacts very



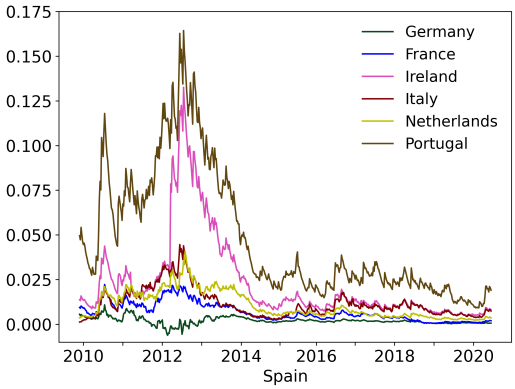
(a) German shock; ρ_t model



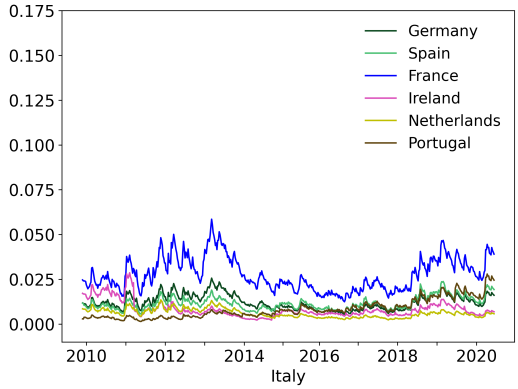
(b) German shock; $\boldsymbol{R}_t$ model



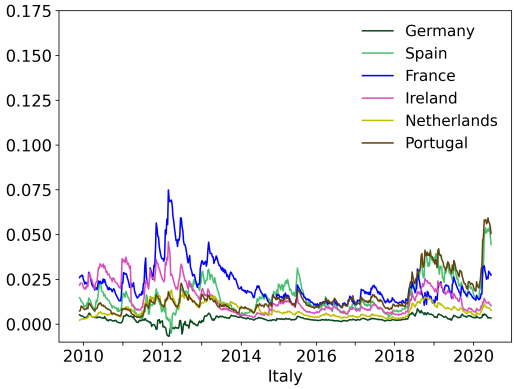
(c) Spanish shock; ρ_t model



(d) Spanish shock; $\boldsymbol{R}_t$ model



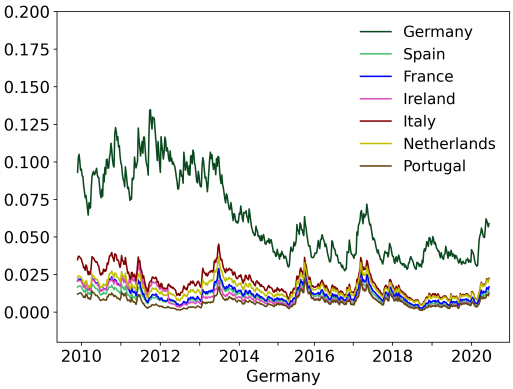
(e) Italian shock; ρ_t model



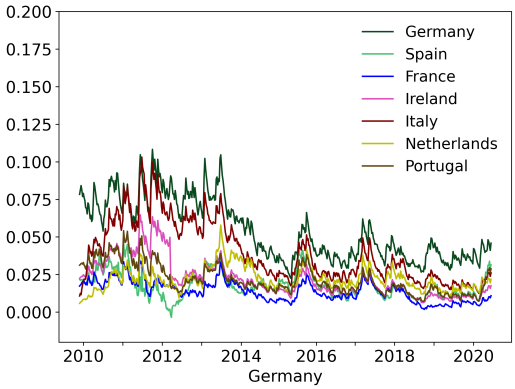
(f) Italian shock; $\boldsymbol{R}_t$ model

Figure 4: First-order spillover effects

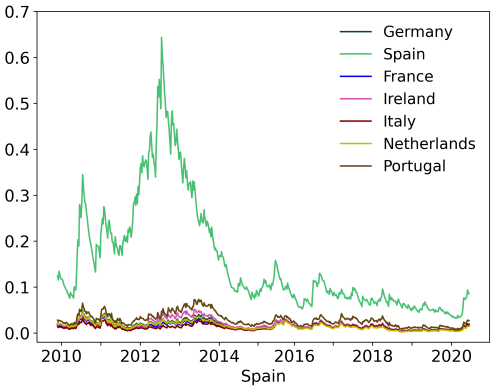
This figure gives the first-order effect $SR_{i,j,t}$ of a one-standard-deviation shock to the government bond yield spread change of country j on country i for j equal to Germany (top panels), Spain (middle panels), or Italy (lower panels). The left and right panels relate to the dynamic scalar and diagonal spatial model, respectively. Spreads are denoted in basis points (bp). The sample ranges from December 2009 - July 2020.



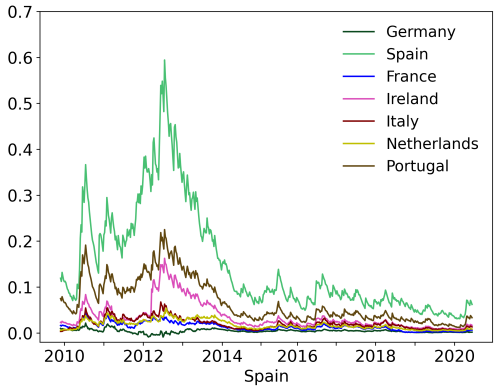
(a) German shock; ρ_t model



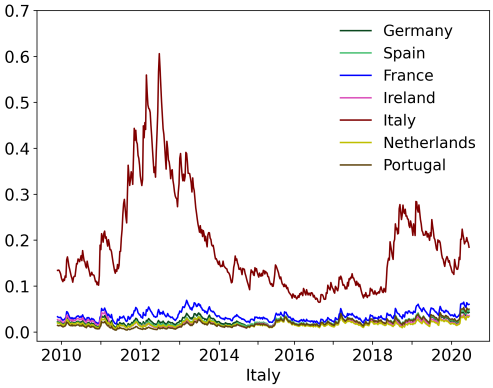
(b) German shock; $\boldsymbol{R}_t$ model



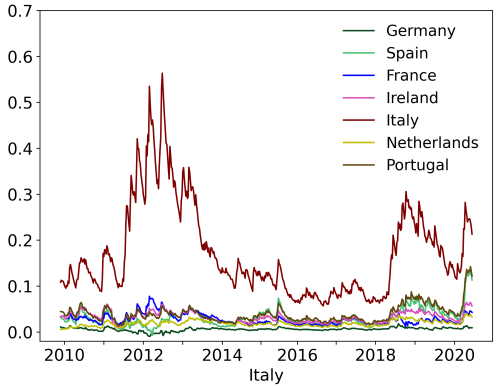
(c) Spanish shock; ρ_t model



(d) Spanish shock; $\boldsymbol{R}_t$ model



(e) Italian shock; ρ_t model



(f) Italian shock; $\boldsymbol{R}_t$ model

Figure 5: Long-term spillover effects

This figure gives the compound reduced-form effect $LR_{i,j,t}$ of a one-standard-deviation shock of the government bond yield spread of country j on country i for j equal to Germany (top panels), Spain (middle panels), or Italy (lower panels). The left and right panels relate to the dynamic and diagonal spatial model, respectively. The sample ranges from December 2009 - July 2020.

differently to Italian shocks in the scalar versus the diagonal model. In the scalar model in Figure 4e the impact on Germany directly follows that of France in magnitude. In the heterogeneous model in Figure 4f, by contrast, Germany reflects a much more stabilizing role in the Eurozone during the crisis and even reacts negatively to Italian structural shocks, something that is not possible in the scalar homogeneous model.

We also see that during the onset of the pandemic Germany is affected much more in the heterogeneous model than in the scalar model. This again makes sense as the pandemic was detached from economic fundamentals, and particularly Italy was severely affected as one of the first European countries at the start of the pandemic. The dynamics of the heterogeneous model thus appear to follow the dynamics of the different crises much more closely and provide a more intuitive economic description of the data. It also allows the spillovers to peak at different moments for the different countries in the sample, which is clearly relevant as some countries peak at the moment where other countries exhibit a trough in $\boldsymbol{R}_{i,i,t}$.

Compounding the first-order spillover effects from equation (9) into their reduced from effects, we obtain the patterns in Figure 5. We can now also see the reduced form effect of a shock to a country on itself.[6] For the homogeneous scalar models, a one-standard-deviation shock only appears to have a sizeable reduced form effect on the country itself. This holds for all countries considered, though the effect in basis points (vertical axes) differs, the effect being understandably higher for Spain and Italy than for Germany. However, allowing for heterogeneous spillovers again changes the story considerably. Figure 5a shows that a one-standard-deviation shock to Germany not only has an effect on Germany, but that the effect is similarly sized for Italy up to almost the end of the sample. The countries thus are much more connected than in it would appear in the homogeneous scalar case.

A similar effect is seen for Spain in Figure 5d: whereas the scalar model only really shows a reduced form effect of Spain on Spain and no substantial country connections, the heterogeneous model reveals much more interconnections, with sizeable effects of Spain on both Portugal and Ireland during the sovereign debt crisis. For instance, the reduced-form effect of Spain on Portugal is only 0.05 bp in July 2012 in the scalar model, whereas it is fivefold (0.25 bp) if we allow for heterogeneity and dynamics in spillover. Figure C.1 of Appendix C shows the effects of Germany on Italy and Spain on Portugal accompanied by 95% confidence bounds as in Blasques et al. (2016) to account for parameter uncertainty. The figure again clearly shows that the differences between the scalar and heterogeneous model is substantial and highly significant.

For Italy, the homogeneous and heterogeneous reduced form model outcomes are much more comparable: the structural shocks to Italy appear to spill over less into the other Eurozone countries

[6] The $SR_{i,i,t}$ is always zero in our setting where the diagonal of $\boldsymbol{W}_t$ equals zero.

during the sovereign debt crisis. However, towards the end of the sample during the onset of the pandemic, we see Italian shocks having sizeable effect on particularly Spain and Portugal. All of the above differences between countries and the timing in terms of the connections in the network are obscured if one assumes homogeneous and/or static behavior of the spillover parameters. Only when allowing for both effects as is done with the model proposed in this paper, one uncovers the much richer empirical historical developments in the data.

5 Robustness checks

In this section, we present two robustness checks to the main analysis. First, we show that the main results persist if we correct European yield changes for U.S. yield changes as a possibly omitted common factor. Second, we show that results do not hinge on the use of the 5-year government bond yields. Results remain similar if we use 1-year bond yields as well 5-year CDS spreads.

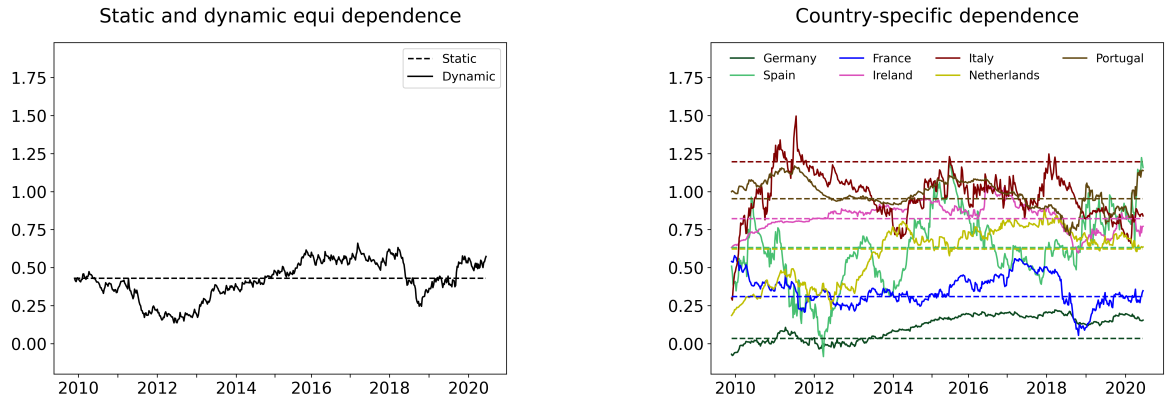
5.1 Accounting for a common global (U.S.) factor

In spatial regression models, there is a concern (Hale and Lopez, 2019) that the network effects may in part be confounded by a missing common factor. In our current context of spillover effects in Eurozone countries, network effects could mix in with an omitted global common government credit risk factor. As a robustness check, we therefore introduce the U.S. government bond yield spread as an observed common factor. In a first step, we regress all Eurozone yield changes on the U.S. yield changes to take out this common factor. Subsequently, we redo our analysis using the residuals of these regressions as our dependent variables.

Figure 6 reports the new filtered heterogeneous spatial dependence parameters. Though some of the patterns change slightly when taking out the common U.S. component, the overall picture in Figure 6 remains similar compared to that in Figure 3. In particular, the spatial spillover strengths remain remarkably heterogeneous across countries: countries like Italy, Spain, and Portugal still exhibit high spatial dependence strength during the crisis, whereas a country like Germany has low and sometimes even negative spatial dependence.

5.2 Alternative dependent variables

As a second robustness check, we confirm that the importance of time-variation and heterogeneity in dynamic networks for the Eurozone is not confined to the use of 5-year government bond yields. In particular, we perform the analysis also for two alternative datasets: euro-denominated 5-year CDS (credit default swap) spread changes and 1-year government yields over the 1-year EONIA



(a) (b)

Figure 6: Filtered spatial autoregressive parameters after U.S. correction

This figure shows fitted spillover parameters for four different models, applied to changes of government bond yield spreads of seven European countries after taking out the effect of U.S. yield spread changes over EONIA. The left and right panels show results for the (static and dynamic) scalar homogeneous and diagonal heterogeneous model, respectively. The sample spans the period December 2009 - July 2020.

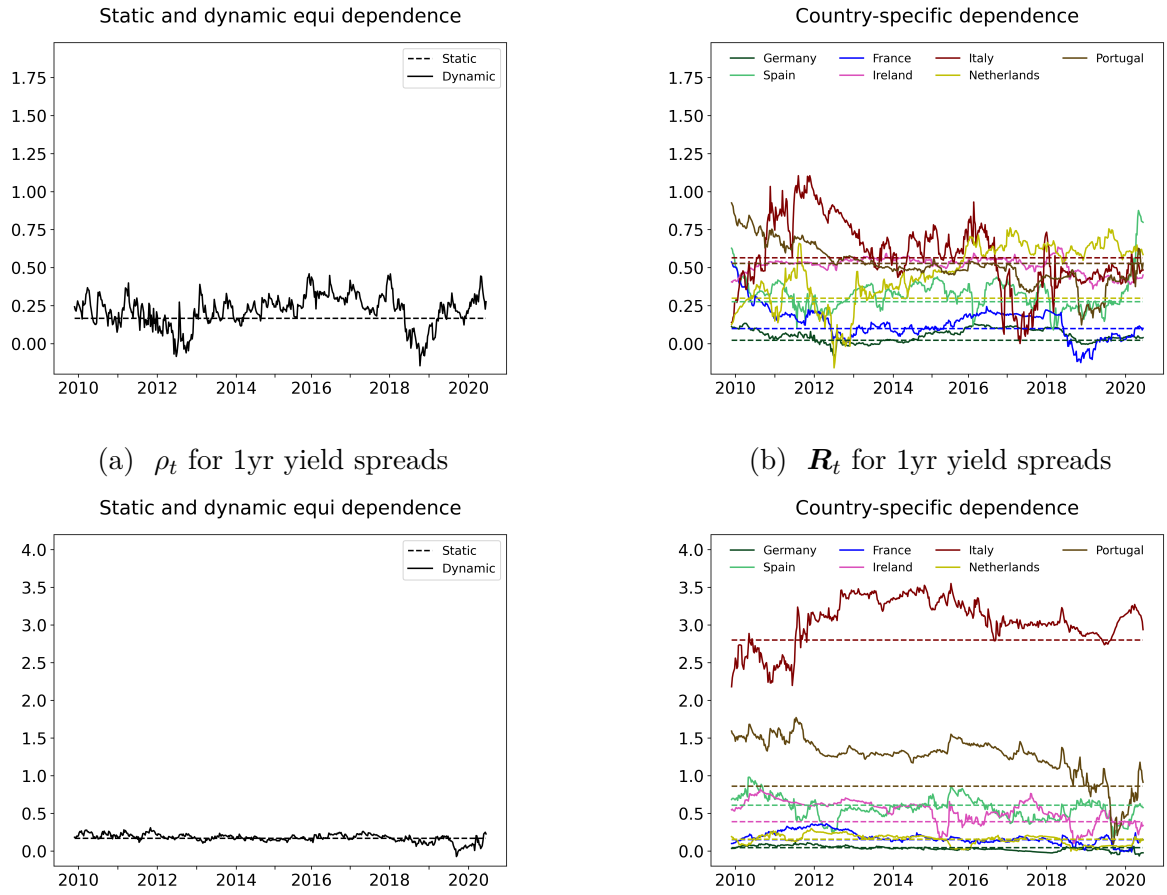
Table 3: Log-likelihood AIC values for models based on 1-year government bond yield spreads and 5-year CDS spread data.

Set-up is similar as in Table 2.

	Static Scalar	Dynamic Scalar	Static Diagonal	Dynamic Diagonal
Panel A: 1y bond yields over EONIA				
logLik	4733	4764	4778	4813
AIC	-9424	-9482	-9503	-9555
Panel B: 5y euro-denominated CDS				
logLik	-10562	-10540	-10284	-10228
AIC	21167	21126	20622	20526

OIS rate. Summary statistics are provided in the Appendix C. Like the 5-year bond yield data, both the 1-year bond yields and the 5-year CDS rates exhibit clear signs of fat tails and outliers, such that our use of the fat-tailed Student t distribution and the time-variation in $\mathbf{\Sigma}_t$ seems warranted.

Table 3 and Figure 7 summarize the main results. Full estimation results including all parameters and standard errors are provided in the Appendix C. For the 1-year spreads, we confirm the value-added of both the heterogeneity (45 likelihood points increase) and the time-variation (another 35 likelihood points increase). A similar result holds for the 5 year CDS spreads, with an increase of 278 likelihood points for adding heterogeneity, and a further 56 points for the time variation. When accounting for the number of parameters using the AIC, these results are confirmed: the AIC is lowest for the model with both heterogeneity and dynamics.



(a) ρ_t for 1yr yield spreads

(b) $\boldsymbol{R}_t$ for 1yr yield spreads

(c) ρ_t for 5yr CDS spreads

(d) $\boldsymbol{R}_t$ for 5yr CDS spreads

Figure 7: Filtered spatial autoregressive parameters for 1-year bond yield EONIA spread changes and 5 year CDS spread changes

This figure shows filtered spillover parameters for the four different models, applied to changes in 1-year government bond yield spreads of seven European countries, or to their 5-year CDS spread changes. The left and right panels are for the homogeneous and heterogeneous model, respectively. The sample ranges from December 2009 - July 2020.

Figure 7 also confirms our previous findings. As in our baseline analysis, we see that the magnitude as well as the dynamics of the scalar ρ_t are heavily limited due to the cross-sectional heterogeneity and time-series variation in the $\boldsymbol{R}_{i,i,t}$ for the different countries. As a result, if all of this information is pooled and only a scalar ρ_t is estimated, hardly any interesting signal is left. By contrast, for instance for the 1-year government bond yield spreads, we again see the dominance in spatial sensitivity of countries like Italy and Portugal, as well as the opposite movements of Italy versus most other countries during the first years of the sample.

CDS markets do not move in full lock-step with bond markets over the sample period. Despite this fact, the importance of both network dynamics and network heterogeneity are also evident

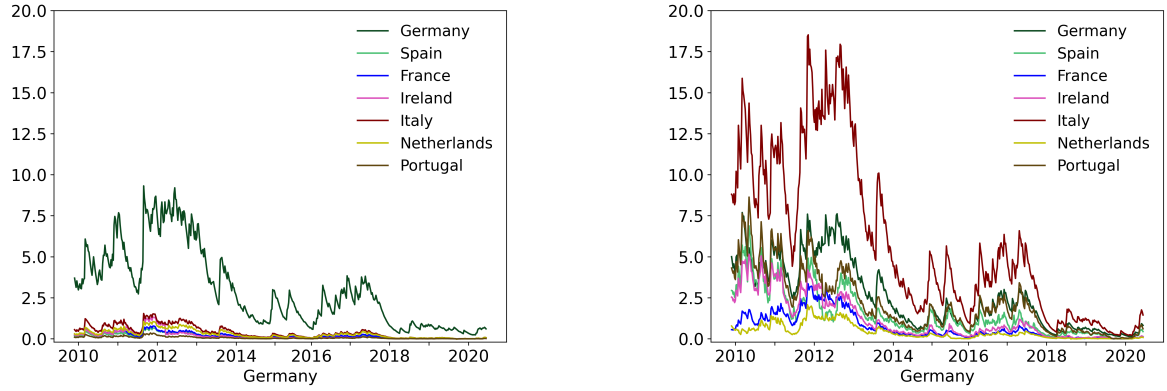
from 5-year CDS spread results. The cross-sectional variation in $\boldsymbol{R}_{i,i,t}$ for the CDS spreads is even larger than for bond spreads, ranging from slightly below zero for Germany to a high value of 3.5 for Italy. Again, Italy and Portugal have the highest $\boldsymbol{R}_{i,i,t}$, followed by Spain and Ireland. As before, the size of $\boldsymbol{R}_{i,i,t}$ is substantially above 1 for both Italy and Portugal without jeopardizing the stability of the model, i.e., $\boldsymbol{R}_t\boldsymbol{W}_t$ still has all eigenvalues well inside the unit circle at all times. Also the CDS data show the importance of heterogeneity in time-variation: we identify an increase in $\boldsymbol{R}_{i,i,t}$ for Italy during the first years of the sample, whereas countries like Spain and Portugal show a downward trend during that same period. Similarly, Ireland and Portugal exhibit troughs that are not matched by any of the other series. As a result, hardly any pattern is left in the scalar ρ_t (lower-left panel). Pooling of $\boldsymbol{R}_{i,i,t}$ is therefore not a good idea: it obscures the empirical patterns and impedes a good empirical interpretation of the dynamics of network connections.

Finally, Figure 8 gives an example of the robustness of the risk measure results. We concentrate on the CDS spread data models and the total, reduced form risk spillover $LR_{i,j,t}$ measure for a one-standard-deviation shock to Germany. We see an even more pronounced effect of heterogeneity and dynamics than in Figure 5. In the scalar model, the effect of a German shock erroneously mainly affects Germany and none of the other economies. However, if we allow for heterogeneity, the picture changes dramatically. Particularly during the sovereign debt crisis, German shocks also heavily affect Spain, Portugal, Ireland, and Italy. This is fully in line with the story of the European sovereign debt crises, and not recovered by the simpler, pooled standard model. The effect on Italy even turns out to be higher than on Germany itself. All of this is missed by the scalar model.

6 Conclusion

In this paper, we proposed a new model for dynamic risk networks with heterogeneous and dynamic spatial spillover coefficients. We proved that such models are much more effective in describing dynamic risk networks than standard, homogeneous static spatial models. Discarding heterogeneity or time-variation of the network connections can lead to flawed conclusions about the major players in the network and their position over time. By contrast, the new model was successful in uncovering such patterns. Though flexible, the model remains tractable and straightforward to estimate using maximum likelihood methods, and we proved that the model has a stationary and ergodic solution and that its filter is invertible.

We applied the new model to three different datasets related to Europe perceived sovereign credit risk over the period 2009-2020. Using pooled spatial dynamics as in standard network models failed dramatically in uncovering the importance of the different players over time. In particular,



(a) (b)

Figure 8: Reduced-form risk spillovers ($LR_{i,j,t}$) for a one-standard-deviation German shock based on 5-year CDS spread data (left homogeneous, right heterogeneous model)

This figure shows the compound reduced-form effect $LR_{i,j,t}$ of a one-standard-deviation shock of the government bond yield spread of Germany to other Eurozone countries based on models estimated with 5 year CDS spread changes as the dependent variable. The left-hand panels relate to the dynamic scalar spatial model. The right-hand panels correspond tot he dynamic diagonal spatial model. The sample ranges from December 2009 - July 2020.

the anchoring role of Germany for the Eurozone only became clear after sufficient heterogeneity was allowed for. Similarly, risk connections between players like Spain and Portugal, and the special behavior of Italy during the crisis only emerged upon allowing for heterogeneity as well as time-variation, irrespective of which of the three datasets was used. Empirical models for network dynamics should thus carefully account for such differences in spillover strength and their time-variation if they are to help the researcher in uncovering the true story in the data and a correct assessment of risk.

References

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Systemic risk and stability in financial networks. *American Economic Review 105*(2), 564–608.

Acharya, V., I. Drechsler, and P. Schnabl (2014). A pyrrhic victory? bank bailouts and sovereign credit risk. *The Journal of Finance 69*(6), 2689–2739.

Anselin, L. (2009). Spatial regression. *The SAGE handbook of spatial analysis 1*, 255–276.

Aquaro, M., N. Bailey, and M. H. Pesaran (2021). Estimation and inference for spatial models

with heterogeneous coefficients: an application to us house prices. *Journal of Applied Econometrics 36*(1), 18–44.

Asgharian, H., W. Hess, and L. Liu (2013). A spatial analysis of international stock market linkages. *Journal of Banking & Finance 37*(12), 4738–4754.

Babii, A., X. Chen, and E. Ghysels (2019). Commercial and residential mortgage defaults: Spatial dependence with frailty. *Journal of Econometrics 212*(1), 47–77.

Billé, A. G., F. Blasques, and L. Catania (2019). Dynamic spatial autoregressive models with time-varying spatial weighting matrices. *Available at SSRN 3241470*.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics 104*(3), 535–559.

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

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

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

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics 227*(2), 325–346.

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

Bougerol, P. and N. Picard (1992). Strict Stationarity of Generalized Autoregressive Processes. *The Annals of Probability 20*(4), 1714 – 1730.

Catania, L. and A. G. Billé (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics 32*(6), 1178–1196.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*(4), 552–563.

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

Creal, D., S. J. Koopman, A. Lucas, and M. Zamojski (2020). Generalized autoregressive method of moments.

Demirer, R., R. Ferrer, and S. J. H. Shahzad (2020). Oil price shocks, global financial markets and their connectedness. *Energy Economics 88*, 104771.

Denbee, E., C. Julliard, Y. Li, and K. Yuan (2021). Network risk and key players: A structural analysis of interbank liquidity. *Journal of Financial Economics 141*, 831–859.

Diebold, F. X. and K. Yılmaz (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics 182*(1), 119–134.

Elliott, M., B. Golub, and M. O. Jackson (2014). Financial networks and contagion. *American Economic Review 104*(10), 3115–53.

Fernandes, G. B. and R. Artes (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research 249*(2), 517–524.

Fernandez, V. (2011). Spatial linkages in international financial markets. *Quantitative Finance 11*(2), 237–245.

Fernández-Avilés, G., J.-M. Montero, and A. G. Orlov (2012). Spatial modeling of stock market comovements. *Finance Research Letters 9*(4), 202–212.

Franklin, A. and G. Douglas (2000). financial contagion. *Journal of Political Economy 108*(1), 1–33.

Hale, G. and J. A. Lopez (2019). Monitoring banking system connectedness with big data. *Journal of Econometrics 212*(1), 203–220.

Harvey, A. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Volume 52. Cambridge University Press.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2020). Firm volatility in granular networks. *Journal of Political Economy 128*(11), 4097–4162.

Kou, S., X. Peng, and H. Zhong (2018). Asset pricing with spatial interaction. *Management Science 64*(5), 2083–2101.

Milcheva, S. and B. Zhu (2016). Bank integration and co-movements across housing markets. *Journal of Banking & Finance 72*, S148–S171.

Pino, G. and S. C. Sharma (2019). On the contagion effect in the us banking sector. *Journal of Money, Credit and Banking 51*(1), 261–280.

Romano, S. (2021). The 2011 crisis in italy: A story of deep-rooted (and still unresolved) economic and political weaknesses. In *Financial Crisis Management and Democracy*, pp. 173–184. Springer, Cham.

Tonzer, L. (2015). Cross-border interbank networks, banking risk and contagion. *Journal of Financial Stability 18*, 19–32.

Williamson, S. D. (2003). Payments systems and monetary policy. *Journal of Monetary Economics 50*(2), 475–495.

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

Zhu, B. and C. Lizieri (2021). Connected markets through global real estate investments. *Real Estate Economics 49*, 618–654.

Zhu, B. and S. Milcheva (2020). The pricing of spatial linkages in companies' underlying assets. *The Journal of Real Estate Finance and Economics 61*(3), 443–475.

Appendix to: Heterogeneity and Dynamics in Spatial Network Models

Enzo D'Innocenzo[a], André Lucas[a,b] Anne Opschoor[a,b], Xingmin Zhang[c]

[a] *Vrije Universiteit Amsterdam*

[b] *Tinbergen Institute*

[c] *School of Finance, Southwestern University of Finance and Economics*

A Score-driven time-varying parameters

A.1 Time varying spatial correlations

The score updates in our model drive the diagonal time-varying matrix $\boldsymbol{R}_t = \boldsymbol{R}(\boldsymbol{f}_t)$, where we gathered the time-varying parameters in a vector $\boldsymbol{f}_t$ as

$$\boldsymbol{f}_t^\top = (\boldsymbol{R}_{1,1,t}, \dots, \boldsymbol{R}_{N,N,t},)\,.$$

The log predictive density is given by

$$\log p_{\boldsymbol{y}}\left(\boldsymbol{y}_t; \boldsymbol{R}(\boldsymbol{f}_t), \boldsymbol{\Sigma}, \nu\right) = \log\Gamma\left(\frac{\nu+N}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2}\log\left(\nu-2\right) - \frac{N}{2}\log\left(\pi\right)$$

$$-\frac{1}{2}\log\mid\boldsymbol{\Sigma}\mid+\log\mid\mathbf{I}_N-\boldsymbol{R}_t\boldsymbol{W}_t\mid-\frac{\nu+N}{2}\log\left(1+\frac{\boldsymbol{e}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t}{(\nu-2)}\right), \tag{A.1}$$

with $\boldsymbol{e}_t = \boldsymbol{y}_t - \boldsymbol{R}_t\,\boldsymbol{W}_t\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta}$. Let $\boldsymbol{E}_i$ be an $N\times N$ matrix of zeros, with a a single 1 on the ith diagonal element. Scaling the score by $|\mathbf{I}_N - \boldsymbol{R}_t\boldsymbol{W}_t|\cdot\mathbf{I}_N$, we obtain the unscaled score as

$$\begin{aligned}
\boldsymbol{\nabla}_{i,t} &= \frac{\partial\log p_{\boldsymbol{y}}\left(\boldsymbol{y}_t;\boldsymbol{R}_t,\boldsymbol{\Sigma},\nu\right)}{\partial\boldsymbol{R}_{i,i,t}} \\
&= w_t\;\boldsymbol{e}_t^\top\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{R}_t}{\partial f_{i,t}}\boldsymbol{W}_t\boldsymbol{y}_t - \operatorname{vec}\left(\left(\left(\mathbf{I}_N-\boldsymbol{R}_t\boldsymbol{W}_t\right)^{-1}\right)^\top\right)^\top\operatorname{vec}\left(\frac{\partial\boldsymbol{R}_t}{\partial f_{i,t}}\boldsymbol{W}_t\right) \\
&= w_t\;\boldsymbol{e}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{E}_i\boldsymbol{y}_t^\star - \operatorname{vec}\left(\left(\left(\mathbf{I}_N-\boldsymbol{R}_t\boldsymbol{W}_t\right)^{-1}\right)^\top\right)^\top\operatorname{vec}\left(\boldsymbol{E}_i\boldsymbol{W}_t\right) \\
&= w_t\;\boldsymbol{e}_{i,t}^\top\boldsymbol{\Sigma}_{i,i,t}^{-1}\boldsymbol{y}_{i,t}^\star - \operatorname{trace}\left(\left(\mathbf{I}_N-\boldsymbol{R}_t\boldsymbol{W}_t\right)^{-1}\boldsymbol{E}_i\boldsymbol{W}_t\right)
\end{aligned}$$

$$= w_t \; \boldsymbol{e}_{i,t}^{\top} \boldsymbol{\Sigma}_{i,i,t}^{-1} \boldsymbol{y}_{i,t}^{\star} - \left(\boldsymbol{W}_t \left(\mathbf{I}_N - \boldsymbol{R}_t \boldsymbol{W}_t \right)^{-1} \right)_{i,i},$$

such that for the scaled score we obtain

$$\begin{aligned} \boldsymbol{s}_{i,t} &= |\mathbf{I}_N - \boldsymbol{R}_t \boldsymbol{W}_t| \cdot w_t \; \boldsymbol{e}_{i,t}^{\top} \boldsymbol{\Sigma}_{i,i,t}^{-1} \boldsymbol{y}_{i,t}^{\star} - \left(\boldsymbol{W}_t \boldsymbol{Z}_t^{\star} \right)_{i,i} \\ &= |\mathbf{I}_N - \boldsymbol{R}_t \boldsymbol{W}_t| \cdot \operatorname{diag} \left(w_t \; \boldsymbol{e}_t^{\top} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_t^{\star \top} \right) - \operatorname{diag} \left(\boldsymbol{W}_t \boldsymbol{Z}_t^{\star} \right)_{i,i}. \end{aligned}$$

For $\nu \to +\infty$, we see $w_t \to 1$ and the score function vector collapses to that of multivariate normal distribution.

A.2 Extension: time varying log error variances

We can generalize the set-up to also allow for a time varying diagonal covariance matrix $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}(\boldsymbol{f}_t)$ if we set

$$\boldsymbol{f}_t^{\top} = \left(\boldsymbol{R}_{1,1,t}, \ldots, \boldsymbol{R}_{N,N,t}, \log \boldsymbol{\Sigma}_{1,1,t}, \ldots, \log \boldsymbol{\Sigma}_{N,N,t} \right).$$

The log predictive density is given by

$$\log p_{\boldsymbol{y}} \left(\boldsymbol{y}_t; \boldsymbol{R}(\boldsymbol{f}_t), \boldsymbol{\Sigma}(\boldsymbol{f}_t), \nu \right) = \log \Gamma \left(\frac{\nu + N}{2} \right) - \log \Gamma \left(\frac{\nu}{2} \right) - \frac{N}{2} \log \left(\nu - 2 \right) - \frac{N}{2} \log \left(\pi \right)$$

$$- \frac{1}{2} \log \mid \boldsymbol{\Sigma}(\boldsymbol{f}_t) \mid + \log \mid \mathbf{I}_N - \boldsymbol{R}_t \boldsymbol{W}_t \mid - \frac{\nu + N}{2} \log \left(1 + \frac{\boldsymbol{e}_t^{\top} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{e}_t}{(\nu - 2)} \right), \tag{A.2}$$

For the scores with respect to $f_{i+N,t} = \log \boldsymbol{\Sigma}_{i,i,t}$, using the diagonality of $\boldsymbol{\Sigma}_t$, we have

$$\begin{aligned} \boldsymbol{s}_{i+N,t} &= \boldsymbol{\nabla}_{i+N,t} \\ &= \frac{\partial \log p_{\boldsymbol{y}} \left(\boldsymbol{y}_t; \boldsymbol{R}(\boldsymbol{f}_t), \boldsymbol{\Sigma}(\boldsymbol{f}_t), \nu \right)}{\partial f_{i+N,t}} \\ &= \frac{\partial \log p_{\boldsymbol{y}} \left(\boldsymbol{y}_t; \boldsymbol{R}_t, \boldsymbol{\Sigma}_t, \nu \right)}{\partial \boldsymbol{\Sigma}_{i,i,t}} \; \frac{\partial \exp \left(f_{i+N,t} \right)}{\partial f_{i+N,t}} \\ &= \left(-\frac{1}{2 \boldsymbol{\Sigma}_{i,i,t}} + \tfrac{1}{2} w_t \; \frac{\boldsymbol{e}_{i,t}^2}{\boldsymbol{\Sigma}_{i,i,t}^2} \right) \cdot \exp \left(f_{i+N,t} \right) \\ &= \left(-\frac{1}{2 \boldsymbol{\Sigma}_{i,i,t}} + \tfrac{1}{2} w_t \; \frac{\boldsymbol{e}_{i,t}^2}{\boldsymbol{\Sigma}_{i,i,t}^2} \right) \cdot \boldsymbol{\Sigma}_{i,i,t} \\ &= \tfrac{1}{2} w_t \; \frac{\boldsymbol{e}_{i,t}^2}{\boldsymbol{\Sigma}_{i,i,t}} - \frac{1}{2} \end{aligned}$$

B Statistical properties of the model

Proof of Theorem 2.1

We first prove stationarity and ergodicity of the matrix process $\{\boldsymbol{R}(\boldsymbol{f}_t)\}_{t\in\mathbb{Z}}$ by checking the conditions C1 and C2 of Theorem 3.1 of Bougerol (1993). We note that if the model is the dgp, the process can be embedded in a stochastic recurrence equation of the type

$$\phi_t(\boldsymbol{f}_t) = \boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{A}\boldsymbol{s}_t^{\boldsymbol{\varepsilon}}, \tag{B.1}$$

$$\begin{aligned}
\boldsymbol{s}_t^{\boldsymbol{\varepsilon}} &= |\mathbf{I}_N - \boldsymbol{R}(\boldsymbol{f}_t)\ \boldsymbol{W}_t| \cdot w_t^{\boldsymbol{\varepsilon}} \cdot \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\ \boldsymbol{y}_t^{\star\top}\right) - |\mathbf{I}_N - \boldsymbol{R}(\boldsymbol{f}_t)\ \boldsymbol{W}_t| \cdot \operatorname{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t(\boldsymbol{f}_t)^{-1}\right) \\
&= |\boldsymbol{Z}_t(\boldsymbol{f}_t)| \cdot w_t^{\boldsymbol{\varepsilon}} \cdot \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\ \boldsymbol{y}_t^{\star\top}\right) - |\boldsymbol{Z}_t(\boldsymbol{f}_t)| \cdot \operatorname{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t(\boldsymbol{f}_t)^{-1}\right) \\
&= |\boldsymbol{Z}_t(\boldsymbol{f}_t)| \cdot w_t^{\boldsymbol{\varepsilon}} \cdot \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\ \ (\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t)^\top\ \boldsymbol{Z}_t(\boldsymbol{f}_t)^{-1\ \top}\right) - \operatorname{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f}_t)\right) \\
&= w_t^{\boldsymbol{\varepsilon}} \cdot \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\ \ (\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t)^\top\ \boldsymbol{Z}_t^\star(\boldsymbol{f}_t)^\top\right) - \operatorname{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f}_t)\right),
\end{aligned}$$

$$w_t^{\boldsymbol{\varepsilon}} = \left(1 + \frac{N+2}{\nu-2}\right) \Big/ \left(1 + \frac{\boldsymbol{\varepsilon}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t}{\nu-2}\right),$$

where we put the steplength weigth v_t of $\boldsymbol{A}_t = v_t \cdot \boldsymbol{A}$ directly to its maximum value $v_t = 1$, which provides the most conservative bound later on. From equation (B.1) it is clear that the sequence $\{\phi_t(\,\cdot\,)\}_{t\in\mathbb{Z}}$ is stationary and ergodic, since under Assumptions 1–3 $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is i.i.d., and $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$ and $\{\boldsymbol{W}_t\}_{t\in\mathbb{N}}$ are stationary and ergodic. Then, in order check condition C1 of Theorem 3.1 of Bougerol (1993), we consider

$$\begin{aligned}
\mathrm{E}\left[\log^+ \|\phi_t(\boldsymbol{f}) - \boldsymbol{f}\|_\infty\right] \leq\ & 2\log 2 + \log^+ \|\boldsymbol{\omega}\|_\infty + \log^+ \|(\boldsymbol{B} - \mathbf{I}_N)\boldsymbol{f}\|_\infty + \log^+ \|\boldsymbol{A}\|_\infty \\
&+ \mathrm{E}\left[\log^+ \left\|w_t^{\boldsymbol{\varepsilon}} \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\ \ (\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t)^\top\ \boldsymbol{Z}_t^\star(\boldsymbol{f})^\top\right)\right\|_\infty\right] \\
&+ \mathrm{E}\left[\log^+ \|\operatorname{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)\|_\infty\right],
\end{aligned} \tag{B.2}$$

where $\boldsymbol{f}$ is fixed at a value such that $\varrho(\boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t) \leq 1 - \epsilon$ lie at least ϵ inside the unit circle. The first four terms in (B.2) on the right-hand side are immediately finite.
The fifth term in (B.2) is finite because

$$\begin{aligned}
&\mathrm{E}\left[\log^+ \left\|w_t^{\boldsymbol{\varepsilon}} \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\ \ (\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t)^\top\ \boldsymbol{Z}_t^\star(\boldsymbol{f})^\top\right)\right\|_\infty\right] \\
&\qquad \leq \mathrm{E}\left[\log^+ \left\|w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\ \ (\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t)^\top\ \boldsymbol{Z}_t^\star(\boldsymbol{f})^\top\right\|_\infty\right]
\end{aligned}$$

$$\leq \log^+ \left\|\boldsymbol{\Sigma}^{-1}\right\|_\infty + \mathrm{E}\left[\log^+ \left\| w_t^{\boldsymbol{\varepsilon}} \boldsymbol{\varepsilon}_t \ (\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t)^\top \right\|_\infty\right] + \mathrm{E}\left[\log^+ \left\| \boldsymbol{Z}_t^\star(\boldsymbol{f})^\top \right\|_\infty\right]$$

The first term is finite due to the non-singularity of $\boldsymbol{\Sigma}$ from Assumption 1. The last term is finite due to the eigenvalues of $\boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t$ always lying ϵ inside the unit circle, and thus the adjoint of $\mathbf{I}_N - \boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t$ being finite by a standard spectral decomposition. The middle term is finite by a Hölder inequality, the uniform boundedness of $w_t^{\boldsymbol{\varepsilon}} \cdot \boldsymbol{\varepsilon}_t$, and the existence of a second moment of $\boldsymbol{X}_t$ from Assumption 2.

The sixth term in (B.2) is finite because

$$\begin{aligned}\mathrm{E}\left[\log^+ \left\|\mathrm{diag}\left(\boldsymbol{W}_t \boldsymbol{Z}_t^\star(\boldsymbol{f})\right)\right\|_\infty\right] &\leq \mathrm{E}\left[\log^+ \left\|\boldsymbol{W}_t \boldsymbol{Z}_t^\star(\boldsymbol{f})\right\|_\infty\right] \\ &\leq \mathrm{E}\left[\log^+ \left\|\boldsymbol{W}_t\right\|_\infty\right] + \mathrm{E}\left[\log^+ \left\|\boldsymbol{Z}_t^\star(\boldsymbol{f})\right\|_\infty\right],\end{aligned}$$

where the first term is finite due to Assumption 3, and the second term is finite using the same argument as before.

Define the commutation matrix $\mathcal{K}_N$ with $\mathrm{vec}(\boldsymbol{W}_t) = \mathcal{K}_N \ \mathrm{vec}(\boldsymbol{W}_t^\top)$ and the selection matrix S_N with $\mathrm{diag}(\boldsymbol{W}_t) = S_N^\top \mathrm{vec}(\boldsymbol{W}_t)$ and $\mathrm{vec}(\boldsymbol{W}_t) = S_N \ \mathrm{diag}(\boldsymbol{W}_t)$. Also define the restricted set $\boldsymbol{\mathcal{F}}_t = \{\boldsymbol{f} \mid |\varrho(\boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t)| \leq 1 - \epsilon\}$, where $\varrho(\,\cdot\,)$ denotes the spectral radius of a matrix. To prove that condition C2 of Bougerol (1993) is satisfied, we only need to show that our contraction condition in Assumption 4 implies condition C2 of Bougerol. To see this, we take the derivative of $\phi_t(\,\cdot\,)$ and obtain

$$\begin{aligned}\sup_{\boldsymbol{f}\in\boldsymbol{\mathcal{F}}_t} \left\|\dot{\phi}_t(\boldsymbol{f})\right\|_\infty &\leq \sup_{\boldsymbol{f}\in\boldsymbol{\mathcal{F}}_t} \left\|\boldsymbol{B} + \boldsymbol{A}\cdot \dot{\boldsymbol{s}}_t^{\boldsymbol{\varepsilon}}(\boldsymbol{f})\right\|_\infty, \\ \dot{\boldsymbol{s}}_t^{\boldsymbol{\varepsilon}}(\boldsymbol{f})^\top &= w_t^{\boldsymbol{\varepsilon}} \cdot \frac{\partial \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top \boldsymbol{Z}_t^\star(\boldsymbol{f})^\top\right)}{\partial \boldsymbol{f}^\top} - \frac{\partial \mathrm{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\ &= w_t^{\boldsymbol{\varepsilon}} \cdot S_N^\top \frac{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top \boldsymbol{Z}_t^\star(\boldsymbol{f})^\top\right)}{\partial \boldsymbol{f}^\top} - S_N^\top \frac{\partial \mathrm{vec}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\ &= S_N^\top \left(\left(\mathbf{I}_N \otimes w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top\right)\mathcal{K}_N - (\mathbf{I}_N \otimes \boldsymbol{W}_t)\right)\frac{\partial \mathrm{vec}\left(\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\ &= S_N^\top \left(\left(\mathbf{I}_N \otimes w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top\right)\mathcal{K}_N - (\mathbf{I}_N \otimes \boldsymbol{W}_t)\right)\frac{\partial \mathrm{vec}\left(\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\ &= S_N^\top \left(\mathcal{K}_N\mathcal{K}_N\left(\mathbf{I}_N \otimes w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top\right)\mathcal{K}_N - (\mathbf{I}_N \otimes \boldsymbol{W}_t)\right)\frac{\partial \mathrm{vec}\left(\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\ &= S_N^\top \left(\mathcal{K}_N\left(w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top \otimes \mathbf{I}_N\right) - (\mathbf{I}_N \otimes \boldsymbol{W}_t)\right)\frac{\partial \mathrm{vec}\left(\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\ &= S_N^\top \left(\left(w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top \otimes \mathbf{I}_N\right) - (\mathbf{I}_N \otimes \boldsymbol{W}_t)\right)\frac{\partial \mathrm{vec}\left(\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\ &= -S_N^\top \left(\left(w_t^{\boldsymbol{\varepsilon}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t)^\top \otimes \mathbf{I}_N\right) - (\mathbf{I}_N \otimes \boldsymbol{W}_t)\right)\dot{\boldsymbol{Z}}_t^\star(\boldsymbol{f})\left(\boldsymbol{W}_t^\top \otimes \mathbf{I}_N\right)S_N.\end{aligned}$$

Therefore, we can conclude that, under maintained Assumptions, the diagonal matrix process $\{\boldsymbol{R}(\hat{\boldsymbol{f}}_t)\}_{t\in\mathbb{N}}$ initialized at $\boldsymbol{f}_1$ converges e.a.s. to a unique stationary ergodic process $\{\boldsymbol{R}(\boldsymbol{f}_t)\}_{t\in\mathbb{Z}}$. This clearly implies the stationarity of the sequence $\{\boldsymbol{R}(\boldsymbol{f}_t)\boldsymbol{W}_t\}_{t\in\mathbb{Z}}$, since $\boldsymbol{W}_t$ is stationary and ergodic and has a log moment by Assumption 3.

To prove that the vector process $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic, we note that $\boldsymbol{Z}_t(\boldsymbol{f}_t)$ is never singular because $\sup_t \varrho(\boldsymbol{R}(\boldsymbol{f}_t)\boldsymbol{W}_t)) \leq 1-\epsilon$. This allows us to write

$$\boldsymbol{y}_t = \boldsymbol{Z}_t^{-1}\left(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t\right) = \left(\mathbf{I}_N - \boldsymbol{R}(\boldsymbol{f}_t)\boldsymbol{W}_t\right)^{-1}\left(\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{\varepsilon}_t\right),$$

which is stationary as $\boldsymbol{\varepsilon}_t$ is assumed i.i.d. in Assumption 1, and $\boldsymbol{X}_t$ and $\boldsymbol{W}_t$ are assumed stationary and ergodic in Assumptions 2–3, and all of these have at least a log moment. We can apply Theorem 1.1 of Bougerol and Picard (1992) and conclude that $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic. The existence of the second moment of $\boldsymbol{y}_t$ follows immediately from $\nu \geq 2+\delta$ for some positive δ, and the uniform boundedness of $\boldsymbol{R}(\boldsymbol{f}_t)$. ∎

Proof of Theorem 2.2

Filter invertibility can be proved by using similar arguments as in the proof of Theorem 2.1 by checking if the conditions C1 and C2 of Theorem 3.1 of Bougerol (1993) hold for our (initialized) filtered process $\{\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$. As the filtering equations are now used as a filter rather than as a dgp, the stochastic recurrence equation for a given parameter vector $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ takes the form

$$\phi_t(\boldsymbol{f}) = \boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{f} + \boldsymbol{A}\boldsymbol{s}(\boldsymbol{f}),$$

where $\boldsymbol{s}(\boldsymbol{f})$ is defined in (5). Note that $\phi_t(\boldsymbol{f})$ takes the observed data $\{\boldsymbol{y}_t\}_{t\in\mathbb{N}}$ rather than the innovations $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ as inputs.

To verify the log-moment condition C1 in Theorem 3.1 of Bougerol (1993), note that $\boldsymbol{y}_t$ has at least two bounded moments given Theorem 2.1, and that $\boldsymbol{W}_t$ has a uniform bound given Assumption 3. Also note that $\sup_{\boldsymbol{f}\in\boldsymbol{\mathcal{F}}_t}|\boldsymbol{Z}_t(\boldsymbol{f})| < \infty$ given the constraint on the spectral radius of $\boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t$ over $\boldsymbol{\mathcal{F}}_t$ and Assumption 3. We then have

$$\begin{aligned}
&\mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\phi_t(\boldsymbol{f}(\boldsymbol{\theta})) - \boldsymbol{f}(\boldsymbol{\theta})\right\|_\infty\right] \\
&\qquad \leq 2\log 2 + \log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{\omega}\right\|_\infty + \log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|(\boldsymbol{B}-\boldsymbol{\iota})\boldsymbol{f}(\boldsymbol{\theta})\right\|_\infty + \log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{A}\right\|_\infty \\
&\qquad\quad + \mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| w_t(\boldsymbol{f}(\boldsymbol{\theta}))\ \ |\boldsymbol{Z}_t(\boldsymbol{f}(\boldsymbol{\theta}))|\ \ \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right\|_\infty\right]
\end{aligned}$$

$$+ \mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathrm{diag}\left(\boldsymbol{W}_t \boldsymbol{Z}_t^\star(\boldsymbol{f}(\boldsymbol{\theta}))\right)\|_\infty\right].$$

The first four terms are finite due to the compactness of $\boldsymbol{\Theta}$.
For the fifth term, note

$$\begin{aligned}
&\mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| w_t(\boldsymbol{f}(\boldsymbol{\theta}))\ |\boldsymbol{Z}_t(\boldsymbol{f}(\boldsymbol{\theta}))|\ \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right\|_\infty\right] \\
&\qquad \le C + \mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| |\boldsymbol{Z}_t(\boldsymbol{f}(\boldsymbol{\theta}))|\ \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right\|_\infty\right] \\
&\qquad \le C + \log^+(1-\epsilon)^N + \mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right\|_\infty\right] \\
&\qquad \le C' + \mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right\|_\infty\right] \\
&\qquad \le C'' + \mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left(\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right\|_\infty\right] \\
&\qquad \le C''' < \infty,
\end{aligned}$$

where the last inequality follows from a standard Hölder inequality given the existence of a second moment for both $\boldsymbol{y}_t$ (from Theorem 2.1) and $\boldsymbol{X}_t$ (from Assumption 2), and the uniform boundedness of $\boldsymbol{W}_t$ from Assumption 3.
For the sixth term, we note

$$\mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathrm{diag}\left(\boldsymbol{W}_t \boldsymbol{Z}_t^\star(\boldsymbol{f}(\boldsymbol{\theta}))\right)\|_\infty\right] \le C + \mathrm{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{Z}_t^\star(\boldsymbol{f}(\boldsymbol{\theta}))\|_\infty\right] < \infty,$$

where the first inequality follows from Assumption 3, and the second inequality follows from the fact that the adjoint is a sum of products of elements of $\boldsymbol{Z}_t(\boldsymbol{f}(\boldsymbol{\theta})) = \mathbf{I}_N - \boldsymbol{R}(\boldsymbol{f}(\boldsymbol{\theta}))\boldsymbol{W}_t$, whose elements are bounded given the bound $\sup_{\boldsymbol{f}\in\boldsymbol{\mathcal{F}}_t} \varrho(\boldsymbol{R}(\boldsymbol{f})\boldsymbol{W}_t) \le 1-\epsilon$.

Having proved the existence of the bounded log-moment of the stochastic recurrence equation, the proof follows immediately from Theorem 3.1 in Bougerol (1993) by noting that our contraction condition in Assumption 5 imply conditions C2 in his Theorem. To see the latter, note that the contraction condition is obtained by taking the derivative of $\phi_t(\boldsymbol{f})$ which takes the observed data $\{\boldsymbol{y}_t\}_{t\in\mathbb{N}}$ as inputs, that is

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\boldsymbol{\mathcal{F}}_t} \left\|\dot{\phi}_t(\boldsymbol{f})\right\|_\infty \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\boldsymbol{\mathcal{F}}_t} \|\boldsymbol{B} + \boldsymbol{A}\cdot\dot{\boldsymbol{s}}_t(\boldsymbol{f})\|_\infty,$$

$$\dot{\boldsymbol{s}}_t(\boldsymbol{f}) = \frac{\partial}{\partial \boldsymbol{f}^\top}\left(w_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})|\ \mathrm{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right) - \frac{\partial \mathrm{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top}$$

$$
\begin{aligned}
&= |\boldsymbol{Z}_t(\boldsymbol{f})| \ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\frac{\partial\left(w_t(\boldsymbol{f})\ \right)}{\partial \boldsymbol{f}^\top} + w_t(\boldsymbol{f})\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\frac{\partial\left(|\boldsymbol{Z}_t(\boldsymbol{f})|\right)}{\partial \boldsymbol{f}^\top} \\
&\quad + w_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ S_N^\top \frac{\partial\left(\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\right)}{\partial \boldsymbol{f}^\top} - S_N^\top \frac{\partial \operatorname{vec}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)}{\partial \boldsymbol{f}^\top} \\
&= -\dot{w}_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)^\top \\
&\quad - w_t(\boldsymbol{f})\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ |\boldsymbol{Z}_t(\boldsymbol{f})|\ \operatorname{vec}\left(\boldsymbol{Z}_t(\boldsymbol{f})^{\top\,-1}\right)^\top \left(\boldsymbol{W}_t^\top \otimes \mathbf{I}_N\right) S_N \\
&\quad - w_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t^\star\boldsymbol{y}_t^{\star\top} \odot \mathbf{I}_N\right) \\
&\quad + S_N^\top\left(\mathbf{I}_N \otimes \boldsymbol{W}_t\right)\dot{\boldsymbol{Z}}^\star(\boldsymbol{f}_t)\left(\boldsymbol{W}_t^\top \otimes \mathbf{I}_N\right) S_N \\
&= -\dot{w}_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)^\top \\
&\quad - w_t(\boldsymbol{f})\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \left(\left(\boldsymbol{W}_t \otimes \mathbf{I}_N\right)\operatorname{vec}\left(\boldsymbol{Z}_t^\star(\boldsymbol{f})^\top\right)\right)^\top S_N \\
&\quad - w_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t^\star\boldsymbol{y}_t^{\star\top} \odot \mathbf{I}_N\right) \\
&\quad + S_N^\top\left(\mathbf{I}_N \otimes \boldsymbol{W}_t\right)\dot{\boldsymbol{Z}}^\star(\boldsymbol{f}_t)\left(\boldsymbol{W}_t^\top \otimes \mathbf{I}_N\right) S_N \\
&= -\dot{w}_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)^\top \\
&\quad - w_t(\boldsymbol{f})\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \operatorname{vec}\left(\boldsymbol{Z}_t^\star(\boldsymbol{f})^\top\boldsymbol{W}_t^\top\right)^\top S_N \\
&\quad - w_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t^\star\boldsymbol{y}_t^{\star\top} \odot \mathbf{I}_N\right) \\
&\quad + S_N^\top\left(\mathbf{I}_N \otimes \boldsymbol{W}_t\right)\dot{\boldsymbol{Z}}^\star(\boldsymbol{f}_t)\left(\boldsymbol{W}_t^\top \otimes \mathbf{I}_N\right) S_N \\
&= -\dot{w}_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)^\top \\
&\quad - w_t(\boldsymbol{f})\ \operatorname{diag}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t\boldsymbol{y}_t^{\star\top}\right)\ \operatorname{diag}\left(\boldsymbol{W}_t\boldsymbol{Z}_t^\star(\boldsymbol{f})\right)^\top \\
&\quad - w_t(\boldsymbol{f})\ |\boldsymbol{Z}_t(\boldsymbol{f})| \ \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t^\star\boldsymbol{y}_t^{\star\top} \odot \mathbf{I}_N\right) \\
&\quad + S_N^\top\left(\mathbf{I}_N \otimes \boldsymbol{W}_t\right)\dot{\boldsymbol{Z}}^\star(\boldsymbol{f}_t)\left(\boldsymbol{W}_t^\top \otimes \mathbf{I}_N\right) S_N
\end{aligned}
$$

$$
\dot{w}_t(\boldsymbol{f}) = -\left(1 + \frac{N+2}{\nu-2}\right) \Big/ \left(1 + \frac{\boldsymbol{e}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_t}{\nu-2}\right)^2.
$$

This proves the result. ∎

C Extra empirical results

This appendix presents additional empirical results. Parameter estimates are given for 1-year government bond yield spread changes (over EONIA) and for 5-year CDS spread changes. In addition, long term risk spillovers are plotted with a 95% confidence bound for 5-year government bond yield spread changes (over EONIA).

Table C.1: Summary statistics 1-year Government Yield spread and CDS spreads
This table lists descriptive statistics of absolute changes of euro-denominated CDS spreads and 1-year government bond yield spreads of seven European countries and the United States. The 1-year government bond yield spread is defined as the difference between the 1-year government bond yield and the 1-year EONIA OIS rate. All variables are weekly from December 10, 2009 until July 2, 2020.

	mean	s.d.	med	min	max	skew	kurt
1-year gov bond spread changes (percentage points)							
Germany	-0.001	0.041	0.001	-0.19	0.21	-0.21	6.83
Spain	0.001	0.228	-0.001	-1.27	2.86	3.45	52.46
France	-0.001	0.057	0.001	-0.59	0.45	-0.91	32.73
Ireland	-0.001	0.397	-0.001	-2.73	3.42	0.01	29.41
Italy	0.000	0.228	-0.002	-1.02	1.36	0.86	12.07
Netherland	0.000	0.060	0.000	-0.54	0.64	0.16	45.60
Portugal	-0.001	0.553	-0.002	-3.12	2.78	-0.19	14.11
U.S.	0.002	0.062	0.006	-0.56	0.25	-1.71	17.04
5-year CDS spread changes (basis points)							
Germany	-0.025	2.867	0.000	-19.02	15.19	0.09	14.05
Spain	-0.036	17.570	-0.080	-109.51	87.30	-0.56	12.61
France	-0.028	6.382	0.000	-54.84	37.90	-0.84	20.69
Ireland	-0.231	24.841	-0.200	-154.33	223.13	0.83	25.33
Italy	0.051	19.320	-0.400	-99.03	98.74	-0.29	11.33
Netherland	-0.033	4.861	0.000	-42.02	28.31	-1.95	32.72
Portugal	-0.042	39.428	-0.050	-255.23	252.21	0.30	15.53
U.S.	-0.026	2.237	0.000	-12.81	16.76	1.24	17.70

Table C.2: Parameter estimates of spatial dependence models 1-year government bond yield spreads

This table reports the estimated parameters of four spatial dependence models, applied to weekly 1-year government bond yield spread changes of seven Eurozone countries. Robust (Huber sandwich) standard errors are reported in parentheses. The models are based on Student's t distributed disturbances with time-varying heteroscedasticity as in model (1)–(5). We report the maximum log-likelihood value (LogLike) and AIC(Akaike information criterion). The sample runs from December 2009 - July 2020.

	Static Scalar	Dynamic Scalar	Static Diagonal	Dynamic Diagonal
Panel A: Spatial dependence parameters				
ω/ω_{GE}	0.167	0.020	0.022	0.000
	(0.026)	(0.010)	(0.030)	(0.001)
ω_{SP}			0.276	0.007
			(0.055)	(0.006)
ω_{FR}			0.099	0.001
			(0.035)	(0.002)
ω_{IR}			0.529	0.009
			(0.107)	(0.010)
ω_{IT}			0.566	0.030
			(0.171)	(0.021)
ω_{NE}			0.299	0.010
			(0.058)	(0.009)
ω_{PO}			0.528	0.010
			(0.157)	(0.015)
α/α_{GE}		0.011		0.004
		(0.002)		(0.005)
α_{SP}				0.023
				(0.020)
α_{FR}				0.007
				(0.004)
α_{IR}				0.027
				(0.068)
α_{IT}				0.181
				(0.082)
α_{NE}				0.024
				(0.016)
α_{PO}				0.046
				(0.097)
β_{ρ}		0.919		0.984
		(0.046)		(0.017)

(Table C.2 continued)

	Static Scalar	Dynamic Scalar	Static Diagonal	Dynamic Diagonal
Panel B: Volatility and control parameters				
α_{vol}	0.334	0.328	0.328	0.319
	(0.047)	(0.039)	(0.041)	(0.032)
β_{vol}	0.996	0.997	0.996	0.997
	(0.004)	(0.004)	(0.004)	(0.003)
ν	3.374	3.354	3.517	3.509
	(0.275)	(0.273)	(0.293)	(0.290)
Δ VSTOXX	-0.015	-0.016	-0.019	-0.014
	(0.039)	(0.036)	(0.038)	(0.035)
Δ term spread	-0.050	-0.047	-0.057	-0.041
	(0.047)	(0.044)	(0.047)	(0.045)
local stock	-0.001	0.001	0.000	0.001
	(0.010)	(0.009)	(0.009)	(0.009)
logLik	4733	4764	4778	4813
AIC	-9424	-9482	-9503	-9555

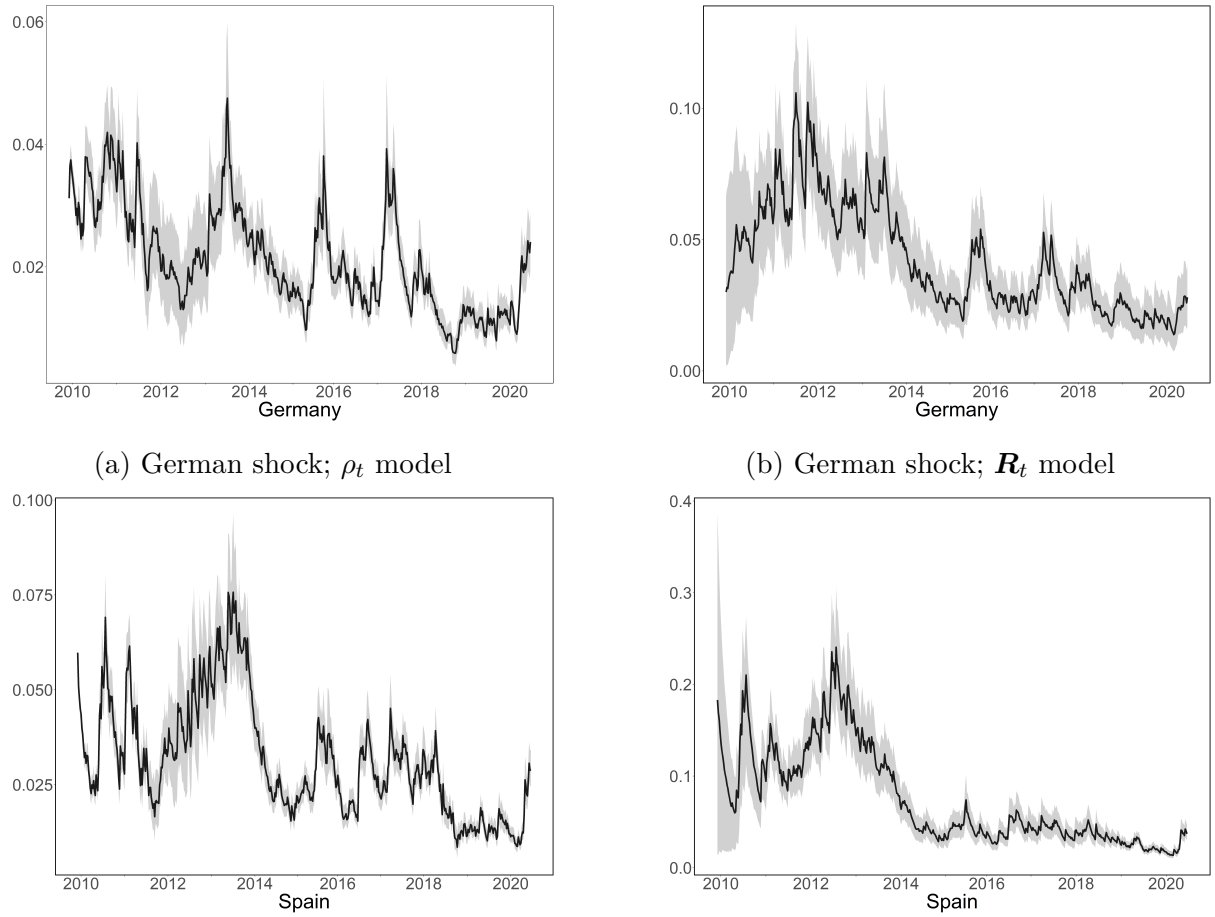
Table C.3: Parameter estimates of spatial dependence models

This table reports the estimated parameters of four spatial dependence models, applied to weekly CDS spread changes of seven Eurozone countries. Robust (Huber sandwich) standard errors are reported in parentheses. The models are based on Student's t distributed disturbances with time-varying heteroscedasticity as in model (1)–(5). To account for the different measurement units of the CDS changes (basis points) compared to the yield spread changes in the other tables, the VSTOXX change was rescaled by a factor 100. We report the maximum log-likelihood value (LogLike) and AIC(Akaike information criterion). The sample runs from December 2009 - July 2020.

	Static Scalar	Dynamic Scalar	Static Diagonal	Dynamic Diagonal
Panel A: Spatial dependence parameters				
ω/ω_{GE}	0.170	0.014	0.045	0.000
	(0.014)	(0.011)	(0.020)	(0.001)
ω_{SP}			0.609	0.009
			(0.055)	(0.012)
ω_{FR}			0.153	0.003
			(0.023)	(0.004)
ω_{IR}			0.391	0.009
			(0.037)	(0.009)
ω_{IT}			2.802	0.195
			(0.277)	(0.124)
ω_{NE}			0.156	0.003
			(0.026)	(0.004)
ω_{PO}			0.861	0.030
			(0.109)	(0.028)
α/α_{GE}		0.003		0.002
		(0.002)		(0.001)
α_{SP}				0.048
				(0.028)
α_{FR}				0.005
				(0.002)
α_{IR}				0.025
				(0.012)
α_{IT}				0.247
				(0.167)
α_{NE}				0.008
				(0.019)
α_{PO}				0.062
				(0.055)
β_{ρ}		0.923		0.982
		(0.059)		(0.017)

(Table C.3 continued)

	Static Scalar	Dynamic Scalar	Static Diagonal	Dynamic Diagonal
Panel B: Volatility and control parameters				
α_{vol}	0.344	0.338	0.350	0.336
	(0.033)	(0.034)	(0.036)	(0.037)
β_{vol}	0.999	0.999	0.999	0.999
	(0.000)	(0.000)	(0.001)	(0.001)
ν	2.537	2.512	2.780	2.842
	(0.249)	(0.246)	(0.313)	(0.332)
Δ VSTOXX	-0.780	-0.973	-0.377	-0.323
	(0.570)	(0.596)	(0.953)	(0.901)
local gov spread	-2.648	-2.857	-2.573	-2.142
	(1.134)	(1.079)	(1.843)	(1.541)
local stock	-0.111	-0.075	-0.226	-0.166
	(0.270)	(0.231)	(0.252)	(0.304)
logLik	-10562	-10540	-10284	-10228
AIC	21167	21126	20622	20526



(a) German shock; ρ_t model

(b) German shock; $\boldsymbol{R}_t$ model

(c) Spanish shock; ρ_t model

(d) Spanish shock; $\boldsymbol{R}_t$ model

Figure C.1: Confidence intervals for long-term spillover effects

This figure gives the confidence intervals for the long-term risk spillover effect $LR_{i,j,t}$ of a 1 standard deviation shock of the 5-yr government bond yield spread of country j to country i, where j equals Germany and i equals Italy (top panels) or Spain (lower panels). The left and right panels relate to the scalar and diagonal spatial model, respectively. The sample ranges from December 2009 - July 2020.