Theory and Application of Dynamic Spatial Time Series Models

Andree, B.P.J.

2020

document version
Publisher's PDF, also known as Version of record

Chapter 4

Parametric Spatial Nonlinearities

Chapter Summary

This paper introduces a new model for spatial time series in which cross-sectional dependence varies nonlinearly over space and time. We refer to it as the Smooth Transition Spatial Autoregressive (ST-SAR) model. We study the stochastic properties for the ST-SAR as a data generating process and obtain asymptotic theoretic properties for the maximum likelihood estimator (MLE) under correct specification and potential misspecification. The asymptotic consistency of the MLE explicitly allows for failure of parameter identification, which is a well-known issue of threshold models. To tackle the implications of the identification issue on the inference of the estimation results, we propose model selection based on in-sample and validation-sample estimates of Kullback-Leibler divergence. The methods are valid when the model is correctly specified, misspecified, over-specified and when parameters are unidentified. Simulations are presented that support the use of information criteria for model selection when the true process is linear and parameters of the model are unidentified, and when the process is nonlinear and the MLE of identified parameters is in fact well-behaved. These results are shown to be robust to additive outliers and fat-tailed errors. The model is applied to study space-time dynamics in two cases that differ in spatial and temporal extent. We study clustering in urban densities and pay particular focus to the advantages of the ST-SAR over linear spatial models as a way to appropriately filter out clustering dynamics. In our second study, we apply the ST-SAR to monthly long term interest rates, and find evidence of asymmetries and cycles in spillover dynamics. In both applications, we find strong evidence for nonlinearity. The empirical evidence highlights that the ST-SAR improves significantly over the SAR and is a powerful tool to understand and predict future values in cross-sectional time series with different dependence regimes.[1]

[1]This chapter is based on *"Smooth Transition Spatial Autoregressive Models"*, available as part of the *Tinbergen Institute Discussion Papers* in the Econometrics and Operations Research research group. The full reference is Andree et al. (2017a).

4.1 Introduction

Spatial entanglement of economic agents plays an important role in the realization of many economic processes measured over space and time. Spatial autocorrelation models are capable of describing the spatial dependence between variables measured across space and are widely adopted in different research fields; see e.g. LeSage and Fischer (2008) on regional growth, Kostov (2009) on agricultural land prices, Baltagi et al. (2014) on housing prices, Debarsy et al. (2015) on foreign direct investments and Hoshino (2016) on crime.

Standard spatial models account for spatial correlation in (un)observed variables, but commonly assume that the spatial autoregressive parameter is constant across space and over time. In particular, models that allow for spatial (auto)correlation often do not sufficiently relax linearity constraints on functional representations of spatial spillovers. Specifically, spillover-processes are represented by "global" dependence parameters (Fotheringham, 2009). The literature has stressed the importance of relying on local statistics for spatial dependence instead of global measures due to parameter heterogeneity; see Anselin (1995) and Fotheringham (2009). Local statistics allow for variation in spatial correlation across grouped cross-sectional units. Typically, spatial aggregation into groups relies on the use of econometric tools to avoid ad-hoc sample divisions. Researchers have for example relied on Geographically Weighted Regression (GWR) (Fotheringham et al., 2002; Su et al., 2012), boosted trees (Crase et al., 2012), Bayesian (Glass et al., 2016), nonparametric (Frías and Ruiz-Medina, 2016), or semiparametric (Basile et al., 2014) approaches to model heterogeneity.

All of the above approaches, however, treat the spatial dependence parameter as static. That is, the (local) parameters represent effects that are fixed (locally) across the data dimensions, rather than introducing relationships that produce varying effects as a function of data itself.

Heterogeneity in interaction is instead achieved by using trend surfaces or sample divisions that allow estimation of separate parameter vectors. For example, the Spatial Autoregressive Semiparametric Geoadditive Models discussed by Basile et al. (2014) maintain linearity assumptions with respect to the spatial autoregressive component, but have smooth locally linear dependence structures with respect to exogenous variables. The GWR isolates neighborhoods in a Cartesian coordinate system using kernels, and estimates weighted parameter vectors for those different neighborhoods. These approaches typically require many observations per neighborhood to be effective and numerous studies pointed out serious drawbacks.[2] We note that grouping observations not by kernels but through fixed effect approaches also has its drawbacks because convergence rates depend on the number of observations in groups tending to infinity (Bonhomme and Manresa, 2015), while in many practical situations additional observations can only be collected over time with group sizes remaining fixed.

In this paper we propose a parsimonious model in which cross-sectional dependence varies nonlinearly over both space and time. The model builds on the well-known SAR model (Anselin, 1988) and the Smooth Transition Autoregressive (STAR) framework advocated by Teräsvirta and Anderson (1992); Granger and Teräsvirta (1993); Teräsvirta (1994); Teräsvirta et al. (2010). In the resulting ST-SAR, dynamics in the cross-sectional dimension are driven by a smooth transition function around lagged cross-sectional variables that are possibly endogenous or "self-exciting". This configuration allows for the possibility of regime-specific dynamics in spillovers with differential in intensity, and allows observations to move smoothly from one regime to another over time. Feedback loops

[2]Inadequate modeling of spatial lag and error processes (Leung et al., 2000; Fotheringham et al., 2002; Paez et al., 2002), spatial patterns revealed by GWR could be attributed to the procedure itself rather than the data generating process (Wheeler and Tiefelsdorf, 2005; Wheeler, 2007), and finally problems that relate to bandwidth selection and local violations of least-squares assumptions (Wheeler and Tiefelsdorf, 2005; Farber and Páez, 2007; Cho et al., 2010).

that amplify spillovers in the cross-section are also modeled with varying intensity in this way, both in the cross-sectional and in the temporal dimension. The entanglement structure remains exogenously determined through the specification of a spatial weights matrix following standard procedures in the spatial econometric literature.

We study the stochastic properties for the ST-SAR as a data generating process and obtain asymptotic theoretic properties for the MLE by taking the time dimension to infinity. We focus on t-distributed innovations as a generalization of the Gaussian case and as an attractive way to achieve robustness to fat tails and outliers. Our theory comes in a variety of flavors that allow for possible failure of parameter identification and potential model misspecification. In particular, we develop consistency when the model is correctly specified or possibly misspecified, and in both cases develop set-consistency when one or more parameters of the model are not identified. We establish asymptotic Gaussianity of the MLE of the correctly specified and identified parameters when the score is a martingale difference sequence and similarly when the model is mis-specified but the score it is near epoch dependent. Since the normality breaks down for unidentified parameters, estimated distributions of parameters cannot be used directly to infer the presence of nonlinear dynamics in the data. We therefore develop in-sample and out-of-sample methods that rely on unbiased estimates of log likelihood, that are still available when one or more parameters are unidentified and the model is set-consistent, to diagnose the presence and significance of nonlinearity. In particular, we investigate the usefulness of information criteria that already have been applied successfully to distinguish nonlinearity in univariate threshold time series. We highlight that information criteria consistently rank the models asymptotically according to Kullback-Leibler divergence, even if parameters are unidentified. To address possible other sources of bias, such as over-fitting, we also provide a theoretical argument for model selection based on a validation-sample estimate of Kullback-Leibler di-

vergence which is again valid when one ore several parameters of the model are not identified because it relies on assumptions that are imposed directly on the differential in forecast errors and not on model parameters. Simulations support the use of information criteria for model selection when the true process is linear and parameters of the model are unidentified, and when the process is nonlinear and the MLE of identified parameters is in fact well-behaved. These results are shown to be robust to additive outliers and fat-tailed errors.

We apply our model to study two cases with different panel dimensions. In the first application, we study clustering in residential densities in a large number of districts in the Netherlands. We test two hypotheses regarding cross-sectional dependencies that cannot be captured by linear models: (i) that spatial autocorrelation decays along the urban gradient in line with the distance decay of agglomeration effects (Fotheringham, 1981; Rosenthal and Strange, 2003); and (ii) that the relation between concentrations of urban densities and household compositions of surrounding neighborhoods inverts along the urban gradient, reflecting sorting patterns that arise under single-crossing assumptions about household preferences (Epple and Sieg, 1999). We model these nonlinearities with a threshold function specified around population densities and find strong evidence for both hypotheses.

Our second application uses a long time series of long term interest rates of a sample of European sovereigns. We assess the integration of financial systems by estimating ST-SAR dynamics in co-movements. Linear dynamics in co-movement, spillovers, and cross-sectional dependence have been explored in a number of studies on financial integration Frankel et al. (2004); Caceres et al. (2016); Kharroubi et al. (2016). We pay particular focus on the time-varying properties of sovereign-specific cross-sectional dependence parameters as a way to understand understand convergence and dispersion in interest rates. Our spatial weights matrix

is based on pair-wise correlations and allows spillovers to flow based on non-geographic linkages. We model nonlinearities with a threshold function specified around ARMA components and find strong evidence asymmetries and cycles in the spillover dynamics.

In both applications, the ST-SAR is shown to be a powerful tool for both understanding and predicting future values in cross-sectional time series in which the dependence of observations on neighbors changes once they enter a different regime. In particular, the ST-SAR improves substantially over the SAR in terms of improving log likelihood, (corrected) AIC and forecasting power. The ST-SAR also renders the residuals free of significant correlations while the SAR residuals maintain both strong spatial clustering and temporal correlations. Our most conservative tests remain significant at the highest level.

The remainder of this paper is organized as follows. Section 4.2 considers spatial autocorrelation models and proposes our nonlinear framework. It also highlights the issues related to parameter identification. Asymptotic theory for the MLE is examined in Section 4.3. Its finite-sample behavior is studied via simulations in Section 4.4. The model is applied in Section 4.5. Section 4.6 summarizes and concludes. Additional results and proofs are located in the Appendix. Additional theoretical results are provided in the Supplementary Appendix that comes with this paper.

4.2 Linear and nonlinear spatial autoregressive models

4.2.1 Linear dynamics: the SAR Model

Spatial data is often highly dependent across space. In order to model this dependence, Cliff and Ord (1969) proposed the Spatial Autoregressive (SAR) model. The SAR in the context an Autoregressive Moving Average

model with Exogenous Regressors (ARMAX) model is given by:

$$\mathbf{y}_t = \rho W \mathbf{y}_t + c + \sum_{p=1}^{P} \mathbf{y}_{t-p}\phi_p + \sum_{k=0}^{K} \mathbf{X}_{t-k}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_t + \sum_{q=1}^{Q} \boldsymbol{\varepsilon}_{t-q}\mu_q \; \forall \; t \in \mathbb{Z} \; , \quad (4.1)$$

$$\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}} \sim p_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t, \Sigma, \boldsymbol{\lambda}),$$

where $\mathbf{y}_t$ denotes a vector of N cross-sectional observations at time t, c is an intercept, ρ is the spatial dependence parameter, W is the $N \times N$ matrix of exogenous spatial weights, ϕ_p is the p-th lag autoregressive parameter, $\mathbf{X}_{t-k}$ is an $N \times D$ matrix of D exogenous regressors at lag k with $\boldsymbol{\beta}_k$ as the $D \times 1$ vector of coefficients, μ_q is a q-th lag moving average parameter and $\boldsymbol{\varepsilon}_t$ is the disturbance vector with multivariate density $p_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t, \Sigma, \boldsymbol{\lambda})$ with zero mean and unknown variance-covariance matrix Σ. Other possible parameters are contained in the vector $\boldsymbol{\lambda}$. In this model structure, each entry y_{it} for $i = 1, ..., N$, of the vector $\mathbf{y}_t$ depends on the local values in the K lags of D individual-specific regressors $\{x_{it-k,d}\}_{d=1,k=0}^{D,K}$, as well as the neighboring entries of y_{jt} and thus indirectly on $\{x_{jt-k,d}\}_{d=1,k=0}^{D,K}$ for $i \neq j$. Similarly, the (moving) error structure spills over. Spatial dependence modeling is made operational by specifying the spatial weights matrix W that defines the dependence structure between cross-sectional entries, for example as a function of geographic or economic distances. It is standard procedure to row-normalize W such that $\sum_{j=1}^{N} w_{ij} = 1 \; \forall \; i \in N$, where w_{ij} is the i,j-th element from W.

The parameter ρ captures the spatially weighted effects of neighboring values $W\mathbf{y}_t$ on $\boldsymbol{\lambda}_t$. In this simple framework, nonlinear feedback effects across entries can be captured, shown by rewriting the model as:

$$\mathbf{y}_t = H^{-1}\Big(c + \sum_{p=1}^{P} \mathbf{y}_{t-p}\phi_p + \sum_{k=0}^{K} \mathbf{X}_{t-k}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_t + \sum_{q=1}^{Q} \boldsymbol{\varepsilon}_{t-q}\mu_q\Big), \quad (4.2)$$

$$\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}} \sim \mathcal{N}ID(0, \sigma_{\boldsymbol{\varepsilon}}^2),$$

where $H := I_N - \rho W$ and I_N denotes the $N \times N$ identity matrix. Following LeSage (2008) we obtain the following infinite power series expansion

$$\mathbf{y}_t = (I_N + \rho W + \rho^2 W^2 + ...)\Big(c + \sum_{p=1}^{P} \mathbf{y}_{t-p}\phi_p + \sum_{k=0}^{K} \mathbf{X}_{t-k}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_t + \sum_{q=1}^{Q} \boldsymbol{\varepsilon}_{t-q}\mu_q\Big). \tag{4.3}$$

Equation (4.3) reveals that when $\rho > 0$, effects spill over to other regions $j \neq i$ with a rate that declines as proximity to i increases, via the structure imposed by W. Feedback occurs for positive w_{ij} and w_{ji} and mutual neighbors i and j, as by construction of the matrix W, every observation is a second order neighbor of itself. A stable process therefore requires exogenous shocks to die out over space, which for the linear spatial process without time dynamics is guaranteed if $\rho \in [-1/|\omega_{min}|, 1/\omega_{max}]$, where $\omega_{min}/\omega_{max}$ are the smallest/largest eigenvalues of W (Lee, 2004), or equivalently $|\rho| < 1$, if the rows of W sum up to one.[3]

The endogenous nature of this model causes inconsistencies in the least squares estimator that increase with N. However, we can consistently estimate SAR models by Quasi Maximum Likelihood methods (Q)ML or Generalized Methods of Moments (GMM), e.g. Kelejian and Prucha (2010). ML estimation of SAR models is pioneered in Ord (1975) and the asymptotics of the QML estimator are derived in Lee (2004). Finite sample distributions are investigated by Das et al. (2003); Bao and Ullah (2007).

4.2.2 The Smooth Transition Spatial Autoregressive model

The linearity of the SAR model imposes the crucial simplifying assumption that the spatial dependence is fixed for any levels of both $\mathbf{y}_{t-p}$ and $\mathbf{X}_{t-k}$. Anselin (1995) argues that spatial heterogeneity can complicate

[3] As we shall see, stability is understood in terms of bound on $||\rho W||$ and depending on the configuration of W, alternative lower-level parameter restrictions can be obtained, see also Elhorst (2010a) for a discussion. In the nonlinear setting, these do not apply as ρ becomes non-scalar, several useful results on the stability of nonlinear spatial systems can be found in the appendix.

the analysis. His argument is centered on the notion that geographic phenomena often do not deviate around a constant mean, but likely move from one local average to another. For this reason, the simple SAR model may be a problematic model for describing the very phenomenon that Cliff and Ord (1969) are trying to model; see Fotheringham (2009) for a discussion. As we shall see in Section 4.5, the linearity assumption is not supported by the data.

In what follows we allow the spatial dependence parameter ρ to change as a function of a set of variables $\mathbf{Z}_t$ that may include (spatial lags of) $\mathbf{y}_{t-p}$ or $\boldsymbol{\varepsilon}_{t-q}$ for any $(p, q) \geq 1$ and or $\mathbf{X}_{t-k}$ for any $k \geq 0$. In particular, we build on the popular smooth transition autoregressive (STAR) model introduced in Teräsvirta and Anderson (1992); Teräsvirta (1994).[4] The resulting Smooth Transition Spatial Autoregressive (ST-SAR) model with ARMAX terms takes the form[5]

$$\mathbf{y}_t = \rho(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t) \circ W\mathbf{y}_t + c + \sum_{p=1}^{P} \mathbf{y}_{t-p}\phi_p + \sum_{k=0}^{K} \mathbf{X}_{t-k}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_t + \sum_{q=1}^{Q} \boldsymbol{\varepsilon}_{t-q}\mu_q, \quad (4.4)$$

$$\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}} \sim p_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t, \Sigma, \lambda),$$

where the spatial dependence $\rho(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t)$ is determined by

$$\rho(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t) = \kappa + \frac{\delta}{1 + \exp(-\gamma(\mathbf{Z}_t - \tau(\boldsymbol{\theta}^{\tau}; \mathbf{Z}_t)))}, \quad (4.5)$$

$$\text{and} \quad \tau(\boldsymbol{\theta}^{\tau}; \mathbf{Z}_t) = \alpha + \mathbf{Z}_t\boldsymbol{\varphi}, \quad (4.6)$$

where $\circ$ denotes element-by-row multiplication, $\boldsymbol{\theta}^{\rho}$ denotes the vector of unknown parameters $\boldsymbol{\theta}^{\rho} := (\kappa, \delta, \gamma, \boldsymbol{\theta}^{\tau})$, and $\boldsymbol{\theta}^{\tau} := (\alpha, \boldsymbol{\varphi})$ is a parameter vector of possible additional parameters within the threshold function that may include linear coefficients w.r.t. any of the ARMAX terms

[4] The STAR model is well known in the time-series literature for modeling nonlinear dynamics with thresholds; see Granger and Teräsvirta (1993) for a literature review of nonlinear time-series models. For a comprehensive review of STAR models, the reader is referred to Dijk et al. (2002).

[5] We discuss the nonlinear model within an ARMAX framework because, as we shall see in our applications, all of terms can affect the data both directly and through the spatial dependence parameters. Allowing the terms to explicitly effect $\mathbf{y}_t$ directly and through $\rho(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t)$ simultaneously, is crucial to determine effect channels.

contained in $\mathbf{Z}_t$. Note that we use $\mathbf{Z}_t$ to refer to any variable which may be an endogenous lag, moving average or exogenous variable, and we allow it to be specified differently in Equation (4.5) and Equation (4.6).

The quantity $\mathbf{Z}_t - \tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t)$ measures deviations of $\mathbf{Z}_t$ from a possibly time-varying quantity $\tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t)$. In general, we allow $\tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t)$ to be any function of the data $\mathbf{Z}_t$. In this paper we consider first order polynomials around three important alternatives, the cross-sectional average $\tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t) = \alpha + \varphi N^{-1} \sum_{i=1}^{N} z_{it}$, the local average $\tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t) = \alpha + \varphi W \mathbf{Z}_t$, and local observations $\tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t) = \alpha + \mathbf{Z}_t \varphi$. Other options include modeling $\tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t)$ as a constant α only, or using wider regional averages $W_l \mathbf{Z}_t$ with W_l as a spatial weights matrix that includes up to l higher order spatial lags.

Note that Equation (4.5)-Equation (4.6) allow the spatial dependence to change smoothly between regimes. The ST-SAR differs considerably from time-varying spatial parameter models such as the spatial score model proposed in Blasques et al. (2018) which attempts to filter the unobserved time-varying sequence of global spatial parameters $\{\rho_t\}_{t \in \mathbb{Z}}$ by means of a score filter. The ST-SAR explores the relation between the spatial dependence parameter ρ and variables in $\mathbf{Z}_t$, which allow it to produce time-varying local spatial parameters. The ST-SAR parameters, δ, $\tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t)$ and γ produce dynamics that cannot be reproduced by the time-varying spatial parameter model of Blasques et al. (2018).

It is also worth noting that the STAR dynamics nest not only the linear SAR model, but also, a threshold model (like the TAR (Tong, 2015)) with instantaneous switching between regimes. The linear SAR case is obtained when $\gamma \to 0$. In contrast, a TAR model is obtained when $\gamma \to \infty$. Depending on $\mathbf{Z}_t$, the transition mechanism may be endogenous or exogenous in nature. In the empirical section we shall consider both exogenous cases such as $\mathbf{Z}_t = \mathbf{X}_{t-p}$ and endogenous examples where we allow the nonlinearities to be driven by ARMA terms. Finally, we note

that, just as in the case of the SAR model, the ST-SAR can also be re-written as

$$\mathbf{y}_t = H(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t)^{-1}\Big(c + \sum_{p=1}^{P} \mathbf{y}_{t-p}\phi_p + \sum_{k=0}^{K} \mathbf{X}_{t-k}\beta_k + \boldsymbol{\varepsilon}_t + \sum_{q=1}^{Q} \boldsymbol{\varepsilon}_{t-q}\mu_q\Big), \quad (4.7)$$

where $H(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t) := I_N - \rho(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t) \circ W$. In SAR terminology, $(I_N - \rho W)^{-1}$ is referred to as the (global) spatial multiplier. In the ST-SAR, we highlight that $H(\boldsymbol{\theta}^{\rho}; \mathbf{Z}_t)^{-1}$ varies locally and over time. As we shall see, this calls for new generalizations of stability conditions.

Identification of the model's parameters

Just as with univariate threshold modeling, an important feature of the model is the possible failure of parameter identification (Teräsvirta et al., 2010). As pointed out, the SAR model is nested in the ST-SAR model, and letting the spatial dependence parameter $\rho(\boldsymbol{\theta}; \mathbf{Z}_t)$ be a constant introduces well-known identification problems related to the fact that nuisance parameters are present only under the alternative assumption of nonlinearity. This is discussed for example by Davies (1977, 1987). In the current case, if $\gamma = 0$ then the parameters inside $\tau(\boldsymbol{\theta}^{\tau}; \mathbf{Z}_t)$ are not identified and k and δ are not separately identified. Furthermore, if $\phi = 1$ and $\alpha = 0$ the spatial dependence may remain constant, unless different variables are used inside $\tau(\boldsymbol{\theta}^{\tau}; \mathbf{Z}_t)$. As we can see from this, the identification problem of distinguishing the SAR model from the ST-SAR model is not straightforward. For example, Likelihood Ratio testing fails because the dimensionality of the parameter space depends on the hypothesis of nonlinearity being true or false. Wald statistics can also not applied to the individual parameters that are essentially meaningless and redundant when the process is linear. To tackle the issue, the following section develops not only consistency of the MLE, but also set-consistency which allows for the failure of parameter identification. This in turn allows one to obtain unbiased estimates of expected likelihood

using out-of-sample validation, or in-sample estimates of Kullback-Leibler divergence by using information criteria. The next section details this further and develops two tools to diagnose the presence and strength of nonlinearity that are valid when one ore several parameters are not identified.

4.3 Asymptotic theory for the ST-SAR model

Estimation of the ST-SAR model's parameters is crucial to infer if nonlinear dynamics are present in the data. The estimated parameters will also inform us about the existence of threshold dynamics, the location of those thresholds, and the smoothness and speed of transitions. In this section we present and discuss the properties of both the log likelihood function and the ML estimator. We provide conditions for the existence, strong consistency and asymptotic normality of the MLE. We also highlight model selection procedures that can be applied to decide between linear and nonlinear descriptions of the data. Our results allow for failure of parameter identification and potential model misspecification. Proofs can be found in Appendix 4.7. Our asymptotic results all refer to increasing the time dimension rather than the spatial dimension since the applications we consider are such that N cannot grow. Additional observations are collected over time only.

For simplicity, we focus our attention on the ST-SAR model with autoregressive dependence of order one ($p = 1$) and deliver a simpler exposition of the theory by focusing on a contemporaneous exogenous variable ($k = 0$) and excluding MA terms ($q = 0$) from this section. In any case, the same asymptotic results for both the correctly specified and the misspecified case are easily generalized to further lags for the exogenous variables and MA terms, at the cost of heavier notation, additional assumptions, and longer proofs. It is well known that the stationarity re-

sults can be generalized to models with moving average components, and can be easily extended to accommodate for (lagged) exogenous variables as long as some data generating process is defined. The endogenous case, on the other hand, is naturally the most interesting case for a study on stationarity. In general, besides extending the stationarity and moments conditions, extra parameter restrictions would need to be put in place to ensure the invertibility of the ST-SAR model and the recovery of the error term sequence.

4.3.1 Existence and measurability of the MLE

Let $\boldsymbol{\theta}$ denote the vector of parameters of our ST-SAR model, $\boldsymbol{\theta} := (\boldsymbol{\theta}^y, \boldsymbol{\theta}^\rho)$, $\boldsymbol{\theta}^\tau \in \boldsymbol{\theta}^\rho$, $\boldsymbol{\theta} := (c, \boldsymbol{\beta}, \phi, \kappa, \delta, \gamma, \alpha, \varphi)'$. Furthermore, let $\boldsymbol{\theta}_0$ denote the parameter of interest. Naturally, the ML estimator $\hat{\boldsymbol{\theta}}_T$ is defined as

$$\hat{\boldsymbol{\theta}}_T \in \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}), \tag{4.8}$$

where

$$\ell_t(\boldsymbol{\theta}) = \ln\det H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t) + \ln p_{\boldsymbol{\varepsilon}}\Big(H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)\mathbf{y}_t - c - \mathbf{y}_{t-1}\phi - \mathbf{X}_t\boldsymbol{\beta}, \Sigma; \lambda\Big). \tag{4.9}$$

The dependence of $\ell_t(\boldsymbol{\theta})$ on the data is omitted in the notation for convenience. Equation (4.9) differs from a standard cross-section likelihood function by the log determinant $\ln\det H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)$, which accounts for the nonlinear spatial feedback (Anselin, 1988). In this paper we shall focus on innovations with density $p_{\boldsymbol{\varepsilon}}$ given by the multivariate Student's t-distribution. The t-distribution naturally generalizes the multivariate normal distribution to allow for fat tails, rendering the dynamics more robust to incidental outliers. Using the standard expression for the multivariate t-distribution with λ degrees of freedom we obtain

$$\ell_t(\boldsymbol{\theta}) = Q(\boldsymbol{\theta}^\rho; \mathbf{Z}_t) + A(\boldsymbol{\theta}) + -\frac{1}{2}(\lambda + N)F(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{X}_t, \mathbf{Z}_t),$$

where $Q(\boldsymbol{\theta}^{\rho};\mathbf{Z}_t)$ is the log determinant

$$Q(\boldsymbol{\theta}^{\rho};\mathbf{Z}_t) := \ln \det H(\boldsymbol{\theta}^{\rho};\mathbf{Z}_t),$$

$A(\boldsymbol{\theta})$ is a constant given by

$$A(\boldsymbol{\theta}) := \ln \Gamma\left((\lambda+N)/2\right)\left[\det\Sigma^{\frac{1}{2}}(\lambda\pi)^{\frac{N}{2}}\Gamma\left(\lambda/2\right)\right]^{-1},$$

and the random element $F(\boldsymbol{\theta},\mathbf{y}_t,\mathbf{X}_t,\mathbf{Z}_t)$ is naturally defined as

$$F(\boldsymbol{\theta},\mathbf{y}_t,\mathbf{X}_t,\mathbf{Z}_t) := \ln\Big(1+\lambda^{-1}\boldsymbol{\varepsilon}_t'\Sigma^{-1}\boldsymbol{\varepsilon}_t\Big),$$

$$\boldsymbol{\varepsilon}_t = H(\boldsymbol{\theta}^{\rho};\mathbf{Z}_t)\mathbf{y}_t - c - \mathbf{y}_{t-1}\phi - \mathbf{X}_t\boldsymbol{\beta}.$$

We first establish the existence and measurability of the MLE $\hat{\boldsymbol{\theta}}_T$. This ensures that the arg max set in Equation (4.8) is not empty and that $\hat{\boldsymbol{\theta}}_T$ is a random variable.

ASSUMPTION. 7 (*Compactness of* Θ). $(\Theta,\mathfrak{B}(\Theta))$ *is a measurable space and* Θ *is a compact subset of* $\mathbb{R}^{d_{\boldsymbol{\theta}}}$.

THEOREM. 9 (*Existence and Measurability*). *Let* ASSUMPTION. *7 hold. Then there exists a.s. an* $\mathfrak{F}/\mathfrak{B}(\Theta)$*-measurable map* $\hat{\boldsymbol{\theta}}_T : \Omega \to \Theta$ *satisfying Equation* (4.4) *for all* $T \in \mathbb{N}$.

4.3.2 Consistency and of the MLE

The consistency of the MLE $\hat{\boldsymbol{\theta}}_T$ w.r.t. the parameter of interest $\boldsymbol{\theta}_0 \in \Theta$ can be obtained under standard regularity conditions. Assumptions 8-9 impose the SE (Stationary and Ergodic) nature of the data and a bounded moment for $Q(\boldsymbol{\theta}^{\rho};\mathbf{Z}_t)$ and $F(\boldsymbol{\theta},\mathbf{y}_t,\mathbf{X}_t,\mathbf{Z}_t)$. ASSUMPTION. 10 ensures that $\boldsymbol{\theta}_0$ is identified.

ASSUMPTION. 8. *The random sequence* $\{\mathbf{y}_t,\mathbf{X}_t\}_{t\in\mathbb{Z}}$ *is SE.*

ASSUMPTION. 9. *The following moment conditions are satisfied:*

i $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|Q(\boldsymbol{\theta}^{\rho};\mathbf{Z}_t)| < \infty$;

ii $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|F(\boldsymbol{\theta},\mathbf{y}_t,\mathbf{X}_t,\mathbf{Z}_t)| < \infty$.

ASSUMPTION. 10. $\boldsymbol{\theta}_0 \in \Theta$ *is the unique maximizer of the limit likelihood;*

$$\mathbb{E}\ell_t(\boldsymbol{\theta}_0) > \mathbb{E}\ell_t(\boldsymbol{\theta}) \ \forall \ (\boldsymbol{\theta}, \boldsymbol{\theta}_0) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

More primitive moment conditions can also be given. For example, we will show for $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ that, when the model is correctly specified, Assumption 8 holds within defined parameter regions. As a counterpart to Assumption 9, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|Q(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)| < \infty$ can be obtained by bounding $\rho(\boldsymbol{\theta}^p, \mathbf{Z}_t)W$ away from 1 in norm (see our Supplementary Appendix), which necessarily holds within stable parameter regions.[6] $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|F(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{X}_t, \mathbf{Z}_t)| < \infty$ is implied by logarithmic moment conditions on $\mathbf{y}_t$ and $\mathbf{X}_t$. Again, when the model is correctly specified, then within that same stable parameter region logarithmic moments of $\mathbf{y}_t$ and $\mathbf{X}_t$ follow trivially because $H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)^{-1}$ is uniformly bounded, and hence, Theorem 6.10 in Pötscher and Prucha (1997) applies, as the nonlinear ST-SAR model is bounded by a linear contracting recursion. For example, given that the innovations are Student's -t distributed, $\lambda > 0$ is needed to ensure the existence of logarithmic moments. Finally, Assumption 10 requires $\delta > 0$ and $\gamma > 0$, which holds trivially if the model is correct and not overspecified. As we shall see, even when Assumption 10 does not hold but Θ is still compact, set-consistency can be obtained and model selection can be used to drop the unidentified parameters. When the model is misspecified, moments of the data have to be assumed.

THEOREM. 10 below establishes the strong consistency of the MLE $\hat{\boldsymbol{\theta}}_T$ with respect to $\boldsymbol{\theta}_0 \in \Theta$. When the model is well specified, $\boldsymbol{\theta}_0$ corresponds naturally to the so-called *true parameter* that indexes the distribution of the data. If the model is misspecified, then $\boldsymbol{\theta}_0$ is often called a *pseudo-true parameter* that, by construction, is the minimizer of the expected Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) between

[6]The moment condition $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|Q(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)| < \infty$ is implied by positive definiteness of $\det H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)^{-1}$ which for the SAR with a row-normalized W follows from $|\rho| < 1$. We note that the nonlinear case $|\rho(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)| < 1$ is not a necessary condition for $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|Q(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)| < \infty$; see LEMMA. 5 in the Supplementary Appendix.

the true conditional density of the data $p^0(\mathbf{y}_t|\mathbf{y}^{t-1})$ and the parametric conditional density implied by the ST-SAR model $p(\mathbf{y}_t|\mathbf{y}^{t-1}, \boldsymbol{\theta})$; see e.g. White (1994) for details. In this sense, under model misspecification, the MLE converges at least to the parameter that delivers the best approximation to the true distribution of the data. The economic interpretation of empirical evidence is then as follows. When the model is correctly specified, the estimated parameters can directly be used as evidence for nonlinearity in an economic process. When the model is misspecified, then the parameters converge to the values for which the model best describes the data features, and as such we may conclude that the evidence for the existence of nonlinear regime-dependence in the observed data is stronger than the evidence for linear dependence which instead describes the data poorly.

THEOREM. 10 (*Strong consistency under possible misspecification*). *Let Assumptions 7-10 hold. Furthermore, let* Θ *be such that* Σ *is positive definite for every* $\boldsymbol{\theta} \in \Theta$. *Then the MLE satisfies* $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ *as* $T \to \infty$ *where*

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\ \mathrm{KL}\Big(\ p^0(\mathbf{y}_t|\mathbf{y}^{t-1})\ ,\ p(\mathbf{y}_t|\mathbf{y}^{t-1}, \boldsymbol{\theta})\ \Big).$$

Propositions 1 and 2 give sufficient conditions for the geometric ergodicity of data generated by the ST-SAR. This allows us to impose conditions on the ST-SAR data generating process that ensure Assumptions 8-9 for the endogenous parts of the model. Propositions 1 and 2 can be easily extended to accommodate for exogenous variables $\mathbf{X}_t$ as long as some data generating process is also defined for $\mathbf{X}_t$.

PROPOSITION. 1. *Let* $\{\mathbf{y}_t\}_{t\in\mathbb{N}}$ *be generated by the ST-SAR model in (4.4) with* $\beta = 0$, $\boldsymbol{\varepsilon}_t$ *iid with full support, and* $\sup_{\mathbf{y}\in\mathbb{R}^N} \|\rho(\boldsymbol{\theta}_0^\rho; \mathbf{y}) \circ W\| < 1$. *Then* $\{\mathbf{y}_t\}_{t\in\mathbb{N}}$ *is an aperiodic,* ψ*-irreducible,* T*-Chain.*

PROPOSITION. 2. *Let the conditions of Proposition 1 hold. Assume further that* $\|\boldsymbol{\varepsilon}_t\|^r < \infty$ *for some* $r > 0$, *and* $\lim_{\|\mathbf{y}\|\to\infty} H(\mathbf{y})^{-1} = H_\infty \in \mathbb{R}^{p\times p}$ *with* $\|H_\infty\||\phi| < 1$. *Then* $\{\mathbf{y}_t\}_{t\in\mathbb{N}}$ *is geometrically ergodic.*

$\sup_{\mathbf{y}\in\mathbb{R}^N} \|\rho(\boldsymbol{\theta}_0^\rho;\mathbf{y})\circ W\| < 1$ in Proposition 1 imposes a stability condition that ensures invertibility and a uniform bound of the spatial multiplier process $H(\mathbf{y})^{-1}$. In the Supplementary Appendix, we show that this condition follows when the spectral radius of $\rho(\boldsymbol{\theta}_0^\rho;\mathbf{y})\circ W$ stays strictly below 1 at $\sup_{\mathbf{y}\in\mathbb{R}^N}$. $\|H_\infty\||\phi| < 1$ in Proposition 2 imposes a stricter contraction condition in the time dimension. COROLLARY. 5 makes use of PROPOSITION. 1 and PROPOSITION. 2 to obtain the consistency of the MLE $\hat{\boldsymbol{\theta}}_T$ with respect to $\boldsymbol{\theta}_0$. Note that, this time, the parameter $\boldsymbol{\theta}_0$ does indeed correspond to the *true parameter* that defines the *true distribution* of the data.

COROLLARY. 5 (Consistency under correct specification). *Let* $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ *be generated by the ST-SAR model Equation* (4.4) *under some* $\boldsymbol{\theta}_0 \in \Theta$. *Suppose that Assumptions 7 and 10 hold, and let the conditions of Propositions 1-2 be satisfied. Finally, let* Σ *be positive definite for every* $\boldsymbol{\theta} \in \Theta$. *Then the MLE satisfies* $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ *as* $T \to \infty$.

Theorem 10 and Corollary 5 rely on the uniqueness of the maximizer $\boldsymbol{\theta}_0$. This assumption may however fail to hold. For example, if the model is misspecified, then several parameter values might provide an equally good approximation to the unknown data generating process in Kullback-Leibler divergence. In particular, we might have a non singleton set

$$\Theta_0^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\ KL\Big(\ p^0(\mathbf{y}_t|\mathbf{y}^{t-1})\ ,\ p(\mathbf{y}_t|\mathbf{y}^{t-1},\boldsymbol{\theta})\ \Big).$$

where Θ_0^* is now the *argmin set* composed of more than one element of Θ. Alternatively, if the model is correctly specified, then the uniqueness assumption may fail if the true unknown data generating process is given exactly by a linear SAR since some parameters (e.g. γ, α and ϕ) are unidentified when $\delta = 0$. In this case, there exists a set

$$\Theta_0 = \Big\{\boldsymbol{\theta} \in \Theta\ :\ KL\big(\ p^0(\mathbf{y}_t|\mathbf{y}^{t-1})\ ,\ p(\mathbf{y}_t|\mathbf{y}^{t-1},\boldsymbol{\theta})\ \big) = 0\Big\}$$

of points that deliver a correct description of the distribution of the data.[7]

[7]Examples of the failure of the uniqueness assumption in other econometric settings can be found

4.3.3 Set-consistency of the MLE allowing for possible parameter identification failure

Below, we highlight that if the restrictive uniqueness condition fails, we can still show that the MLE $\hat{\boldsymbol{\theta}}_T$ converges to the set of maximizers of the limit log likelihood function. A simple regularity condition is required which states that the *level sets* of the limit log likelihood function ℓ_∞ are *regular* (see Definition 4.1 Pötscher and Prucha (1997)). The following theorem is obtained directly by application of Lemma 4.2 in Pötscher and Prucha (1997) to a time-invariant continuous limit criterion $\mathbb{E}\ell_t : \Theta \to \mathbb{R}$ defined on a compact parameter space Θ. This theorem holds for possibly misspecified models and ensures set consistency of the MLE $\hat{\boldsymbol{\theta}}_T$ to the set of pseudo-true parameters Θ_0^* of our ST-SAR model. Below, we let $d(\cdot,\cdot)$ denote the usual metric distance from a point to a set, whereby $d(\boldsymbol{\theta}, \Theta^*) = \inf\{\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \ , \ \boldsymbol{\theta}^* \in \Theta^*\}$ for any $\boldsymbol{\theta} \in \Theta$ and $\Theta^* \subseteq \Theta$.

THEOREM. 11. (Set consistency of MLE under possible misspecification and parameter identification failure) *Let Assumptions 7-9 hold and let* Θ *be such that* Σ *is positive definite for every* $\boldsymbol{\theta} \in \Theta$. *Then the MLE* $\hat{\boldsymbol{\theta}}_T$ *is set consistent as* $T \to \infty$,

$$d(\hat{\boldsymbol{\theta}}_T, \Theta_0^*) \xrightarrow{a.s.} 0 \quad \text{as } T \to \infty$$

where Θ_0^* *is the argmin set*

$$\Theta_0^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\ \mathrm{KL}\Big(\ p^0(\mathbf{y}_t|\mathbf{y}^{t-1}) \ , \ p(\mathbf{y}_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}) \ \Big).$$

Theorem 12 obtains the same type of set consistency of the MLE $\hat{\boldsymbol{\theta}}_T$ applied to the setting of Corollary 5, but this time, it is stated for the case of an overspecified ST-SAR model. This is particularly relevant when the true process in fact linear (SAR). In this case, the MLE is shown to be consistent to the set of true parameters $\Theta_0 \subseteq \Theta$ that deliver an *equivalent*, *correct* and *exact* description of the distributional properties

e.g. in (Freedman and Diaconis, 1982) which addresses a simple location problem with *iid* data and (Kabaila, 1983) in the context of time-series models.

of the data.

THEOREM. 12. (Set consistency of MLE under correct specification and parameter identification failure) *Let* $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ *be generated by the ST-SAR model Equation* (4.4)*. Suppose that Assumption 7 holds, and let the conditions of Propositions 1-2 be satisfied. Finally, let* Σ *be positive definite for every* $\boldsymbol{\theta} \in \Theta$*. Then the MLE satisfies* $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \Theta_0$ *as* $T \to \infty$ *where* Θ_0 *is the set of points deliver an equivalent and correct distribution of the data*

$$\Theta_0 = \Big\{\boldsymbol{\theta} \in \Theta : \mathrm{KL}\Big(\ p^0(\mathbf{y}_t|\mathbf{y}^{t-1}) \ , \ p(\mathbf{y}_t|\mathbf{y}^{t-1}, \boldsymbol{\theta})\Big) = 0\Big\}.$$

4.3.4 Asymptotic normality of the MLE

THEOREM. 13 below obtains the asymptotic normality of the MLE. Once again we allow the ST-SAR model to be well specified or misspecified. ASSUMPTION. 11 assumes that the score is either a martingale difference sequence (mds) or, alternatively, that it is near epoch dependent (NED) of size -1 on an underlying α-mixing sequence of appropriate size. It is well known that, if the model is well specified, then the score is a martingale difference sequence (mds). As such, we obtain the desired asymptotic normality application of Billyingsley's central limit theorem (CLT) for an SE martingale difference sequence (mds); see Billingsley (1961). The mds assumption is also appropriate for mild forms of misspecification; see White (1994). Under strong model misspecification, the asymptotic Gaussianity of the score may still be obtained by application of a central limit for processes that are NED on an α-mixing process; see e.g. Theorem 10.2 in Pötscher and Prucha (1997). The verification of the NED property can be easily achieved by appealing to preservation theorems such as Theorem 6.6 in (Pötscher and Prucha, 1997), for example in Corollary 6.8 therein it is obtained if the score is Lipschitz on some transformation of the data which is itself NED of the desired size.

In Assumption 11 the α-mixing sequence is of size $2r/(r-2)$, for some

$r > 2$. As we shall see, we will also require r bounded moments from the score to obtain a CLT. Finally, we note that the CLT could also be obtained for a ϕ-mixing sequence of size $r/(r-1)$.

ASSUMPTION. 11. *The score $\{\nabla \ell_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is either a martingale difference sequence or it is near epoch dependent of size -1 on an underlying α-mixing sequence of size $2r/(r-2)$, for some $r > 2$.*

ASSUMPTION. 12 imposes additional moment conditions that ensure the application of a CLT to the score and a uniform law of large numbers to the second derivative of the log likelihood function. Below we let $\nabla^i Q(\boldsymbol{\theta}_0^\rho; \mathbf{Z}_t)$ and $\nabla^i F(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{X}_t, \mathbf{Z}_t)$ denote the ith derivative of $Q(\boldsymbol{\theta}_0^\rho; \mathbf{Z}_t)$ and $F(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{X}_t, \mathbf{Z}_t)$ with respect to the vector $\boldsymbol{\theta}$. The moment conditions are imposed on each element of the resulting vectors and matrices.

ASSUMPTION. 12. *The following moment conditions are satisfied:*

i $\mathbb{E}|\nabla Q(\boldsymbol{\theta}_0^\rho; \mathbf{Z}_t)|^r < \infty$;

ii $\mathbb{E}|\nabla F(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{X}_t, \mathbf{Z}_t)|^r < \infty$;

iii $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\nabla^2 Q(\boldsymbol{\theta}_0^\rho; \mathbf{Z}_t)| < \infty$;

iv $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\nabla^2 F(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{X}_t, \mathbf{Z}_t)| < \infty$.

If the score is an mds, then conditions (i) and (ii) hold with $r = 2$. If the score is NED, then r is the same as in Assumption 11.

The moment bounds stated in ASSUMPTION. 12, will be satisfied when the data $\mathbf{y}$ and $\mathbf{X}$ have bounded moments of appropriate order. Again, just as for the proof of consistency, when the model is correctly specified, bounded moments for $\mathbf{y}$ and $\mathbf{X}$ can be obtained by applying the theorem 6.10 in Pötscher and Prucha (1997) to the dynamic model stated in Equation (4.7). In particular, when the contraction condition holds $H(\boldsymbol{\theta}^\rho; \mathbf{y}_{t-1})^{-1}$ is bounded see LEMMA. 3 in the Appendix, and the ST-SAR is bounded by a linear recursion, and hence, m moments for $\mathbf{y}$ can be obtained when $\mathbf{X}$ and innovations have m moments.

THEOREM. 13 now delivers the asymptotic Gaussianity of the standardized MLE by imposing the further regularity condition that $\boldsymbol{\theta}_0$ lies in the interior of the parameter space $\text{int}(\Theta)$. This theorem also assumes that $\boldsymbol{\theta}_0$ is well identified. This is reflected in the invertibility of the limit Hessian $\mathbb{E}\ell_t''(\boldsymbol{\theta}_0)$.

THEOREM. 13 (Asymptotic normality of the identified parameters). *Let assumptions 1-6 hold with Σ positive definite for every $\boldsymbol{\theta} \in \Theta$ and invertible Hessian $\mathbb{E}\ell_t''(\boldsymbol{\theta}_0)$. If $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$, then the MLE satisfies*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\mathcal{J}(\boldsymbol{\theta}_0)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\right) \text{ as } T \to \infty,$$

where $\mathcal{J}(\boldsymbol{\theta}_0) := \mathbb{E}\ell_t'(\boldsymbol{\theta}_0)\ell_t'(\boldsymbol{\theta}_0)^\intercal$ is the expectation of the outer product of the score, and $\mathcal{I}(\boldsymbol{\theta}_0) := -\mathbb{E}\ell_t''(\boldsymbol{\theta}_0)$ denotes the Fisher information matrix.

As we shall see, the Monte Carlo simulation developed in Section 4.4 provides evidence of both the consistency and normality claims made in THEOREM. 10 and THEOREM. 13 in the correct and misspecified case.

4.3.5 Model selection under possible parameter identification failure

It is well known that threshold parameters are not identified under the null (Teräsvirta et al., 2010). In univariate literature, nonlinearity tests are often based on auxiliary regressions (Dijk et al., 1999). In the ST-SAR, the expansion approach results in many components as nonlinear feedback extends both in space and time. Auxiliary statistics therefore lead to inefficient results. As an alternative, we explore model selection based on information criteria following Granger et al. (1995); Sin and White (1996). We highlight that information criteria consistently rank the models asymptotically according to Kullback-Leibler divergence, even if parameters are unidentified. To address possible other sources of bias, we also provide a theoretical argument for model selection based on a validation-sample estimate of Kullback-Leibler divergence which

is again valid when one ore several parameters of the model are not identified. Simulations show support the use of information criteria for model selection when the true process is linear and parameters of the model are unidentified, and when the process is nonlinear and the MLE of identified parameters is in fact well-behaved. These results are shown to be robust to additive outliers and fat-tailed errors.

We conclude this section with details on the model selection adopted in the empirical section of this paper. We will consider both in-sample and out-of-sample model selection criteria. Furthermore, we pay special attention to selection criteria that provide an asymptotically consistent ranking of competing models even in the presence of identification issues.

Let $L_T(\boldsymbol{\theta}) = \sum_{t=2}^{T} \ell_t(\boldsymbol{\theta})$ denote the sample log likelihood at $\boldsymbol{\theta} \in \Theta$. It is well know that model selection based on the KL divergence can be achieved by selecting the model with highest expected log likelihood $\mathbb{E}L_T(\boldsymbol{\theta}_0^*)$ evaluated at the best (pseudo-true or true) parameter $\boldsymbol{\theta}_0^* \in \Theta$. Unfortunately, the sample log likelihood $L_T(\hat{\boldsymbol{\theta}}_T)$ that is available in practice is an asymptotically biased estimator of the expected log likelihood $\mathbb{E}L_T(\boldsymbol{\theta}_0^*)$. This is easily shown by using a simple quadratic expansion

$$\lim_{T\to\infty} \mathbb{E}\Big(L_T(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}L_T(\boldsymbol{\theta}_0^*)\Big) = \lim_{T\to\infty} \mathbb{E}\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*)' \frac{1}{T} L_T''(\boldsymbol{\theta}_T^*) \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*) \neq 0.$$

Under considerably restrictive conditions, Akaike (1973, 1974) showed originally that for a model with k parameters,

$$\lim_{T\to\infty} \mathbb{E}\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*)' \frac{1}{T} L_T''(\boldsymbol{\theta}_T^*) \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*) \approx k,$$

and hence, an asymptotically unbiased estimator of $\mathbb{E}\ell_t(\boldsymbol{\theta}_0^*)$ is given by $\frac{1}{T}\sum_{t=2}^{T} \ell_t(\hat{\boldsymbol{\theta}}_T) - k$,

$$\lim_{T\to\infty} \mathbb{E}\Big(\frac{1}{T}\sum_{t=2}^{T} \ell_t(\hat{\boldsymbol{\theta}}_T) - k\Big) = \mathbb{E}\ell_t(\boldsymbol{\theta}_0^*).$$

This follows easily for an asymptotically normal MLE of a correctly specified model since then the information equality holds $\mathcal{J}(\boldsymbol{\theta}_0) = \mathcal{I}(\boldsymbol{\theta}_0)$, and hence

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\mathcal{J}(\boldsymbol{\theta}_0)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\right) = \mathcal{N}\left(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\right),$$

which implies that $\lim_{T\to\infty} \mathbb{E}\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*)' \frac{1}{T} L_T''(\boldsymbol{\theta}_T^*) \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*) = \mathrm{tr}(I_k) = k$. Akaike also proposed the well known AIC information criteria based on the unbiased estimator $\frac{1}{T}\sum_{t=2}^{T} \ell_t(\hat{\boldsymbol{\theta}}_T) - k$ given by AIC$= 2T\big(k - \frac{1}{T}\sum_{t=2}^{T} \ell_t(\hat{\boldsymbol{\theta}}_T)\big)$. Since then, several authors have shown that the AIC can also be used to consistently rank models according to the KL divergence in considerably more general settings (Konishi and Kitagawa, 2008)[8]. The AIC and its variations can be used for consistent in-sample model selection under wider forms of misspecification, for nested or non-nested models, and, most importantly, when test statistics fail, for example because of parameter identification problems; see e.g. Granger et al. (1995); Sin and White (1996); Konishi and Kitagawa (2008).

Importantly, as k are the parameters of the model and independent from the data generating process being linear or nonlinear, it is easy to see that when the model includes unidentified parameters, they penalize the likelihood and increase the AIC while their contribution to the likelihood can be expected to remain low. The small contribution to the likelihood of unidentified parameters is eventually implied for growing data by the result of THEOREM. 12. The AIC therefore favors dropping unidentified parameters in the same way that it favors, for example, dropping autoregressive lags that do not meaningfully contribute to the implied density of a model. Simply put, in the special case that the data is linear, the linear SAR model and the larger nesting ST-SAR model that includes non-meaningful parameters attain very similar log likelihoods.

[8]See pages 61-64 for the Takeuchi Information Criterion. The original reference of Takeuchi 1976 is in Japanese and difficult to find.

At this point, the linear model can be selected on the basis that relative parsimony is favored by the AIC.

For this reason, the use of the AIC for model selection in the context of threshold models has been suggested already by Tong (1983); Li (1988); Tong (1990). Furthermore, Wong and Li (1998) showed that the AICc is an asymptotically unbiased estimator of the expected Kullback–Leibler information for SETAR models and analyzed the finite sample properties of AIC, AICc and BIC by simulation. Theoretical and simulated results on the consistency of information criteria in selecting the lag order of linear autoregressive models have been extended to the case of threshold models by Kapetanios (2001). Finally, Psaradakis et al. (2009) perform an extensive simulation study on the usefulness of the information criteria in selecting between alternative nonlinear time series models and concludes that they are effective even in small samples given that nonlinearity is substantial but that the criteria, particularly the ones with higher penalties, often favor linear models when the data do not have prominent nonlinear characteristics.

REMARK. 2. *We focus on the AIC, the corrected AIC (AICc), and a modified AIC (mAIC). The AICc, introduced by Hurvich and Tsai (1989), improves on the finite sample properties; see Brockwell and Davis (1991); McQuarrie and Tsai (1998); Burnham and Anderson (2004). The mAIC is based on the general setting put forward by Sin and White (1996).*

Unfortunately, specification issues can still influence the in-sample performance of information criteria, for example because the nonlinear model overfits linear data. For this reason, we also consider criteria based on a *validation sample*. In particular, we obtain the sample log likelihood $\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T) = \sum_{t=2}^{\tilde{T}} \tilde{\ell}_t(\hat{\boldsymbol{\theta}}_T)$ based on a validation sample of size $\tilde{T}$, where $\hat{\boldsymbol{\theta}}_T$ is obtained using the estimation sample of size T. The tilde is used in $\tilde{L}$ to emphasize that this log likelihood is calculated using the validation sample.

Lemma 1 states that, when using an (approximately) independent val-

idation sample, the sample log likelihood $\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T)$ is immediately an asymptotically unbiased estimator of $\mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0^*)$. This can be shown using the same quadratic expansion argument as used to derive the AIC, and then letting both T and $\tilde{T}$ diverge to infinity sequentially. In practice, for time-series data with some form of fading memory (e.g. mixing, near epoch dependence, L_p-approximability, etc), a burn-in period of T^* observations between the estimation sample $\mathbf{y}_1, ..., \mathbf{y}_T$ and the validation sample $\mathbf{y}_{T+T^*}+1, ..., \mathbf{y}_{T+T^*+\tilde{T}}$ is needed to ensure the assumption of *approximate independence* of the validation sample.

REMARK. 3. *Unbiased estimates of the out-of-sample likelihood differential can in principle be cross-validated rather than calculated over a single holdout. However, while the approximate independence of the holdout was trivially satisfied by the use of a burn-in that separates it from the estimation sample, leave-one-out or other repeated validation strategies require a correct-specification assumption on both competing SAR and ST-SAR models in order to maintain the required approximate independence of the residuals or need to implement sophisticated strategies that ensure the independence is satisfied in other ways, see Gao et al. (2016); Bergmeir et al. (2018).*

LEMMA. 1. *Let ℓ be twice continuously differentiable, suppose that $\hat{\boldsymbol{\theta}}_T \xrightarrow{as} \boldsymbol{\theta}_0^*$ as $T \to \infty$ and assume that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{L}'_T(\boldsymbol{\theta})| < \infty$ hold. Then $\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T)$ is an asymptotically unbiased estimator of $\mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0^*)$,*

$$\lim_{T,\tilde{T}\to\infty} \mathbb{E}\Big(\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0^*)\Big) = 0$$

Lemma 1 tells us that we can rank models consistently according to the KL divergence without the need to impose penalties whose magnitude rely on intricate assumptions. Lemma 2 below highlights that the ranking is consistent regardless of potential identification issues. In particular, it shows that the models are asymptotically well ranked according to the KL divergence even in the case of a set consistent MLE for the parameters of a well-specified or misspecified ST-SAR model.

LEMMA. 2. *Let ℓ be twice continuously differentiable, suppose that*

$d(\hat{\boldsymbol{\theta}}_T, \Theta_0^*) \xrightarrow{as} 0$ as $T \to \infty$ and assume that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{L}'_T(\boldsymbol{\theta})| < \infty$ hold. Then $\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T)$ is an asymptotically unbiased estimator of $\mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0^*)$ for all $\boldsymbol{\theta}_0^* \in \Theta_0^*$,

$$\lim_{T,\tilde{T}\to\infty} \mathbb{E}\Big(\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0^*)\Big) = 0 \quad \forall\ \boldsymbol{\theta}_0^* \in \Theta_0^*.$$

REMARK. 4. *Let the data be generated by a linear SAR model under some $\boldsymbol{\theta}_0 \in \Theta_0 c \subseteq \Theta$. Let ℓ be twice continuously differentiable, suppose that $d(\hat{\boldsymbol{\theta}}_T, \Theta_0) \xrightarrow{as} 0$ as $T \to \infty$ and assume that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{L}'_T(\boldsymbol{\theta})| < \infty$. Then $\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T)$ is an asymptotically unbiased estimator of the expected log likelihood $\mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0)$ at the true parameter, i.e. $\lim_{T,\tilde{T}\to\infty} \mathbb{E}\big(\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0)\big) = 0$.*

In the special case that the linear SAR model is correctly specified, Remark 4 tells us that the linear SAR model will attain the same zero KL divergence as any larger nesting ST-SAR model. The same holds true for any other model that nests the SAR, as the larger model is also correctly specified. At this point, the linear model can be selected on the basis of being the most parsimonious model that is correctly specified.

In practice, a situation of this type will lead to very similar log likelihoods for the competing models over the validation sample. In Proposition 3 we highlight that the differences in these log likelihood values can be tested for statistical significance using the Diebold-Mariano test statistic (Diebold and Mariano, 1995). Specifically, we can test the rank position of any two models by testing if the difference in log likelihoods in the validation sample is statistically significant or not. This test is also known as a logarithmic scoring rule, see e.g. Diks et al. (2011); Amisano and Giacomini (2007); Bao et al. (2007). Below, we consider two competing models, A and B, and let $\tilde{\ell}_{\tilde{t}}^{\mathrm{A}}(\hat{\boldsymbol{\theta}}_T^{\mathrm{A}})$ and $\tilde{\ell}_{\tilde{t}}^{\mathrm{B}}(\hat{\boldsymbol{\theta}}_T^{\mathrm{B}})$ denote their respective log likelihood contributions at a certain time $T + T^* + 1 < t \leq T + T^* + \tilde{T}$ in the validation sample. Furthermore, we let $\tilde{\Delta}_t(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)$ denote the log likelihood differences $\tilde{\Delta}_t(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B) := \tilde{\ell}_{\tilde{T}}^{\mathrm{A}}(\hat{\boldsymbol{\theta}}_T^{\mathrm{A}}) - \tilde{\ell}_{\tilde{T}}^{\mathrm{B}}(\hat{\boldsymbol{\theta}}_T^{\mathrm{B}})$ evaluated at the point estimates $\hat{\boldsymbol{\theta}}_T^A$ and $\hat{\boldsymbol{\theta}}_T^B$) respectively, and $\tilde{\Delta}_t(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B})$ denote the log

likelihood differences evaluated at each model's pseudo-true parameter. Finally, we let $\tilde{\sigma}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)$ be a consistent estimator of the standard deviation of $\tilde{\Delta}_t(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)$.

PROPOSITION. 3. (Diebold-Mariano test statistic: logarithmic scoring rule) *Let* $\hat{\boldsymbol{\theta}}_T^A \xrightarrow{as} \boldsymbol{\theta}_0^{*A}$ *and* $\hat{\boldsymbol{\theta}}_T^B \xrightarrow{as} \boldsymbol{\theta}_0^{*B}$ *as* $T \to \infty$. *Suppose that the data is strictly stationary and ergodic. Then under the null hypothesis that model A and B fit the data equally well* $\mathrm{H}_0: \ \mathbb{E}\tilde{\Delta}_t(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B}) = 0$, *it follows that*

$$\mathrm{DM}_{\tilde{T},T} := \tilde{T}^{\frac{1}{2}} \left(\tilde{T}^{-1} \sum_{t=T+T^*+1}^{\tilde{T}} \frac{\tilde{\Delta}_t(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)}{\tilde{\sigma}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)} \right) \xrightarrow{d} \mathcal{N}(0,1) \quad as \ \ T, \tilde{T} \to \infty.$$

If instead we have $\mathbb{E}\tilde{\Delta}_t(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B}) > 0$ *(model A is best) then* $\mathrm{DM}_{\tilde{T},T} \to \infty$ *as* $T, \tilde{T} \to \infty$. *Finally, if* $\mathbb{E}\tilde{\Delta}_t(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B}) < 0$ *(model B is best) then* $\mathrm{DM}_{\tilde{T},T} \to -\infty$.

REMARK. 5. *In Section 4.5, a more conservative finite sample correction of the statistic following a Student's-t distribution is also used, see Harvey et al. (1997).*

Just as the AIC, out-of-sample model performance evaluation has been applied in the context of threshold models in earlier literature. See for example Clements et al. (2003) who investigates out-of-sample comparison of the Mean Squared Error and concludes that, in line with to the conclusions around the use of the AIC detailed by Psaradakis et al. (2009), data need to exhibit a substantial degree of non-linearity before the SETAR model is favored over a linear model. For these reasons, we can expect both approaches to favor the ST-SAR only when the true nonlinearity is strong in the data. This is a useful feature because we would only want to accept the alternative assumption of nonlinearity over the null assumption of linearity in an empirical application if the evidence is substantial. Finally, it is important to stress that the DM-type test developed here imposes assumptions directly on the forecast errors, in particular that the likelihood differential is covariance stationary, and can therefore work in the case of unidentified parameters or even in a

model-free environments, see Diebold (2015) for reflections on this.

4.4 Monte Carlo study

To evaluate the empirical relevance of our estimation theory, we conduct a Monte Carlo study. Importantly, we investigate size and power of model selection based on standard information criteria. Foremost, we explore how well popular information criteria are able to distinguish between linearity and nonlinearity when the data is generated by a linear model and the ST-SAR contains unidentified nuisance parameters. We also explore how well the criteria recognize the nonlinear features of data when the true process in nonlinear.

In the following numerical investigation we focus on selection frequencies based on standard information criteria. Recall that evidence exists that information criteria perform well in small samples in the context of univariate threshold models when nonlinearity is strong but favor linear models when nonlinearity is weak (Psaradakis et al., 2009). For this reason, we simulate from a linear model to explore how well the information criteria perform when the data is linear, and only simulate from a relatively flat nonlinear dependence signal when we explore the suitability of information criteria to detect nonlinearity when the data indeed is nonlinear. The data generating process is of the general form:

$$\mathbf{y}_t = H(\boldsymbol{\theta}^{\rho}; \mathbf{y}_{t-1})^{-1}(\boldsymbol{\varepsilon}_t), \quad \boldsymbol{\varepsilon}_t \sim TID(1, I_N; 5), \tag{4.10}$$

We keep the ratio of distant and close-by neighbors comparable across experiments by allowing the network density of the weights matrix to increase with N. In each draw we generate a random zero diagonal row-normalized weights matrix with $N/10$ neighbors for each observation. The process is initialized with $H_1 = I_N$, and the first 50 steps of the sequence are discarded to avoid dependence on the initialization. We simulate

1000 datasets and estimate parameters with Student's-t likelihood.

We focus on an ST-SAR process driven by local averages as the local average should be more sensitive to additive outliers than, say, the cross-sectional mean. We simulate the linear datasets according to a linear SAR process with:

$$\rho = 0.5$$

and simulate the nonlinear data sets according to the nonlinear ST-SAR process:

$$\delta = .4,\ \gamma = 1.05,\ \alpha = -.2,\ \varphi = 1.4,\ \kappa = -.4,$$

$$\mathbf{Z}_t = \mathbf{y}_{t-1},\ \tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t) = \alpha + \varphi W \mathbf{y}_{t-1}.$$

We also consider the effect of additive outliers, similar to Dijk et al. (1999), by simulating contaminated sequences (+ AO) according to the following replacement process:

$$\mathbf{y}_t^* = \mathbf{y}_t + 1.[\boldsymbol{\zeta}_t > 0.5]\psi\boldsymbol{\epsilon}_t, \tag{4.11}$$

$$\{\boldsymbol{\zeta}_t\} \sim UID(0, I_N), \quad \{\boldsymbol{\epsilon}_t\} \sim BID(-I_N, I_N; \pi),$$

with $\pi = 0.05$ and ψ set to the sample equivalents of $\sqrt{\mathbb{E}\mathbf{y}_t^2 - (\mathbb{E}\mathbf{y}_t)^2}$, and estimating on $\mathbf{y}_t^*$.

In table 4.1, we pit the results of the ST-SAR with all its parameters (ST-SAR 2) against SAR estimates and focus on selection between the SAR and the ST-SAR when the process is linear (Size). The selection frequencies are also provided for contaminated data generated from the SAR (right). Both SAR and ST-SAR model are correctly specified with regard to the non-contaminated process, but the SAR is more parsimonious while the ST-SAR has additional parameters to over fit the data and possibly the outliers. The results indicate information criteria can be used to distinguish between linearity and nonlinearity with performance improving as the dimensions of the data grow.

Table 4.1: Size: selection frequencies for (contaminated) data generated from the SAR (right). The results indicate information criteria can be used to distinguish between linearity and nonlinearity with performance improving as the dimensions of the data grow. The ST-SAR is robust to over fitting outliers.

SAR DGP (AO right columns)		ST-SAR 2 vs. SAR			ST-SAR 2 vs. SAR		
		AIC	AICc	mAIC	AIC	AICc	mAIC
N=30	T=10	46	44	45	44	42	44
	T=25	33	32	32	30	29	30
	T=50	27	27	27	26	25	26
	T=100	22	22	22	23	23	23
	T=250	22	22	22	20	20	20
N=40	T=10	52	51	51	54	52	53
	T=25	32	31	31	33	33	33
	T=50	24	24	24	25	25	25
	T=100	22	22	22	23	23	23
	T=250	23	23	23	18	18	18
N=50	T=10	41	39	40	43	42	42
	T=25	25	25	25	27	26	26
	T=50	24	24	24	23	23	23
	T=100	20	19	19	19	19	19
	T=250	18	18	18	18	18	18
N=60	T=10	32	31	32	33	32	32
	T=25	24	23	24	27	27	27
	T=50	24	23	24	26	26	26
	T=100	17	17	17	16	16	16
	T=250	17	17	17	17	17	17

The results of table 4.1 show that the AIC has empirically relevant size. We have discussed that the ST-SAR converges to the set of points that deliver an equivalent and correct distribution of the data, see THEOREM. 12 and that the SAR should thus be selected on the basis of parsimony. The simulation evidence is in support of this notion. For data simulated from the linear SAR, we see that the SAR is indeed selected over the ST-SAR 2 with increasing frequency as the sample size increases. However, as data grows, the larger ST-SAR 2 is still incorrectly selected over the nested SAR with nonzero frequency. At $T = 250,\ N = 60$ we select the ST-SAR 2 in 17% of the cases. This suggests that in practice, one may want the improvement in AICc to be relatively large or prefer to keep

Table 4.2: Power: selection frequencies for data generated from the ST-SAR. The results indicate information criteria can be used to distinguish between linearity and nonlinearity with performance improving as the dimensions of the data grow.

ST-SAR DGP		ST-SAR 1 vs. SAR			ST-SAR 2 vs. SAR			ST-SAR 2 vs. ST-SAR 1		
		AIC	AICc	mAIC	AIC	AICc	mAIC	AIC	AICc	mAIC
N=30	T=10	38	38	38	45	41	43	46	45	46
	T=25	62	61	62	63	62	63	50	48	49
	T=50	80	79	80	83	82	82	57	57	57
	T=100	85	85	85	97	97	97	80	80	80
	T=250	100	100	100	100	100	100	96	96	96
N=40	T=10	51	49	50	52	50	51	44	43	44
	T=25	72	72	72	73	72	72	51	50	50
	T=50	93	93	93	91	91	91	59	59	59
	T=100	92	92	92	100	100	100	84	84	84
	T=250	100	100	100	100	100	100	99	99	99
N=50	T=10	53	52	53	54	52	53	45	43	45
	T=25	84	84	84	85	84	85	55	55	55
	T=50	98	98	98	98	98	98	66	66	66
	T=100	99	99	99	100	100	100	89	89	89
	T=250	100	100	100	100	100	100	100	100	100
N=60	T=10	63	62	63	59	58	59	45	43	44
	T=25	88	88	88	87	87	87	57	56	57
	T=50	99	99	99	99	99	99	71	71	71
	T=100	99	99	99	100	100	100	92	92	92
	T=250	100	100	100	100	100	100	100	100	100

the SAR when the improvement is modest and the data is small. In our empirical applications we find, however, very substantial improvements in the AICc while working with considerable numbers of observations. In our our first empirical application we shall focus on T close to 10 but use a cross-section that is roughly 12 times that of the largest experiment covered by our simulations, while in our second application, T increases beyond what is considered here. The robustness to contamination of the process can again be seen, this time by the fact that selection rates of the ST-SAR do not inflate when additive outliers enter the process.

In table 4.2 we estimate two versions of the ST-SAR; a restricted model that is underspecified – φ and κ are fixed at 0 – (ST-SAR 1) and the correctly specified ST-SAR with all its parameters (ST-SAR 2). As before, we also estimate the SAR. Again, we find evidence that selection

frequencies, now for data generated from the ST-SAR, support the use of information criteria to distinguish between linearity and nonlinearity. In particular, while table 4.1 highlighted the ability of information criteria to correctly favor the SAR when the data is linear, table 4.2 highlights that the criteria favor the ST-SAR when the data is nonlinear. As with size, power improves as the dimensions of the data grow.

The results of table 4.2 show that the AIC has good power if the process is nonlinear. Both the misspecified ST-SAR 1 and correctly specified 2 are selected over the underspecified SAR with increasing frequency as the sample size increases. Furthermore, as data grows the larger and correct ST-SAR 2 is selected over the nested ST-SAR 1 with probability 1. We again see improvements both as T and N increase. Table 4.7 in the Appendix provides additional power results for a contaminated process. Overall, the presence of additive outliers has a small effect on power. For very small samples $T = 10, N \leq 60$, we observe some increase in power indicating slightly increased over fitting. However, for $T > 10, N \leq 60$, the outliers negatively impact power. While we reach a frequency of 92% for $(N, T) = (60, 100)$ without contamination, we obtain only a rate of 80% for distorted data. The reduction in power contrasts the univariate STAR framework in which additive outliers can trick the threshold into fitting the contamination as a nonlinear process (Dijk et al., 1999). We find that in the cross-sectional case, the results mirrors the conclusions of the errors in variables literature. Finally, note that, as in the distribution case, the results are dependent on the strength of the nonlinear signal. In our empirical application we find strong nonlinearities.

The simulations presented here confirm the appropriateness of standard information criteria to decide between different descriptions of spatial spillover processes. Importantly, the evidence indicates that that, not only do information criteria distinguish well between linearity and nonlinearity, they also distinguish between alternative nonlinearities. The AICc comes

forward as the most conservative measure, and therefore we apply it as our primary choice criterion in the empirical section. Additional simulation results in the Appendix, fig. 4.7 in particular, further highlight that the MLE of the identified parameters is well-behaved in empirically relevant sample sizes.

4.5 The empirics of nonlinear spatial dependencies

This section presents two empirical cases. In our first study we use a panel of short T and large N. Our second study focuses on the opposite case of large T and small N. This allows us to explore nonlinearities both from a cross-sectional perspective, as well as from a time-varying perspective.

4.5.1 Application I: Dutch residential densities

The first application evaluates nonlinear spatial dynamics in the clustering of Dutch residential densities at the district level over a period of ten years. The primary focus is on the advantages of the ST-SAR compared to its linear counterpart. We investigate spatially varying features of the dependence structure, particularly in relation to a number of spatially explicit socio-economic variables. Steering urban development and preserving open, green spaces is a major policy concern in the Netherlands Koomen et al. (2008). Understanding the drivers influencing the balance between agglomeration and dispersion is essential to help define policies. These policies have a strong spatial dimension, which can be difficult to disentangle. Panel and cross-sectional methods are essential analysis tools, and we shall focus on the role of cross-sectional nonlinearities in obtaining accurate estimates.

Economic rationale for ST-SAR dynamics in residential densities

The dependent variable is urban density measured as addresses per hectare. We investigate two types on nonlinearities. First, we model nonlinear spatial autocorrelation to allow for differential strength in clustering. In line with the decay in agglomeration forces along the urban gradient (Fotheringham, 1981; Rosenthal and Strange, 2003), we expect autoregressive spatial dependence to fluctuate along clusters of population densities. The linearity of the SAR on the other hand assumes away any variation in autocorrelation along the urban gradient. The second nonlinearity is in the relationship between local densities and the surrounding household composition. This choice is particularly interesting because dense urban centers accommodate different households than spacious low density neighborhoods. Literature on sorting has made empirically tested predictions about the equilibrium distribution of household types across different neighborhoods (Epple and Sieg, 1999). The demand patterns for housing rooted in preference heterogeneity produces a heterogeneous relationship between concentrations in density and household composition. We focus particularly on the share of population under 14 years in surrounding areas, which proxies a mixture of social and demographic characteristics. As households with children locate in low density neighborhoods outside the city center, we can expect that dense urban cores have a positive correlation with the presence of children in surrounding areas. On the other hand, the low density areas outside main urban cores follow the inverse. A linear spatial lag forces the two opposite relationships to average out, which falsely leads to the conclusion that surrounding households are not related to urban densities, contradicting the sorting theory (Epple and Sieg, 1999). The ST-SAR specification allows us to capture the theorized positive and negative relationships simultaneously.

Data for Dutch residential densities

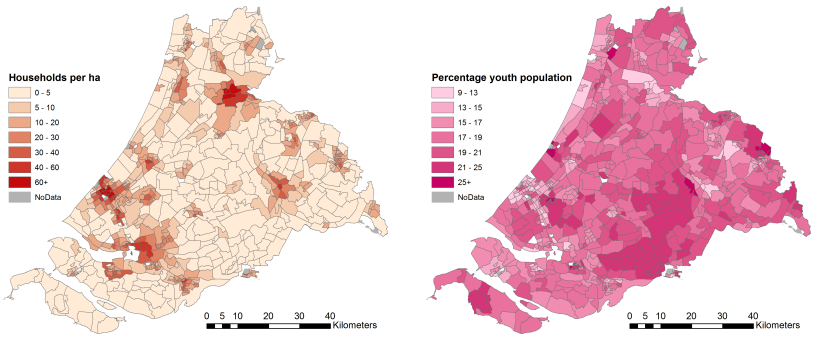

Figure 4.1: Time average spatial distribution of (left) household density and (right) population under 14.

Table 4.3: Overview of explanatory variables and parameter symbols.

Parameter	Interacting variable	Units	Range	Mean
β_{cdens}	Log company density	Continuous	-3.93 to 3.42	-0.14
β_{wcdens}	Spatial log company densities	Continuous	-2.62 to 2.30	-0.21
$\beta_{\%shh}$	Percentage of single households	Continuous	5 to 75.22	32.47
$\beta_{w\%shh}$	Spatial percentage single households	Continuous	0 to 59.75	32.15
$\beta_{w\%hhkids}$	Spatial percentage households with children	Continuous	0 to 24.11	8.75
$\beta_{w\%wim}$	Spatial percentage western immigrants	Continuous	0 to 59.03	36.78
$\beta_{\%nwim}$	Percentage non-western immigrants	Continuous	0 to 67.92	9.90
$\beta_{\%>65}$	Percentage elderly over 65	Continuous	1 to 43.23	15.09
$\beta_{w\%<14}$	Spatial percentage children	Continuous	0 to 25.38	17.81
ρ	Second order queen contiguity matrix	Standardized	0 to 46*	18.91**

Transition function parameters are indexed by the variables they interact with. *The range of the spatial weights matrix is the minimum to maximum number of connections. **Average number of connections.

The time series covers observations of 717 districts from 2005 to 2014 obtained from the Dutch Central Bureaus of Statistics.[9] Figure 4.1 shows the concentrations of urban densities and young population outside urban areas. The other regressors, that control for a variety of local demographic and economic characteristics, are taken from the same dataset. Local

[9]The data is available for download from the Dutch Central Bureau of Statistics: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data.

values and spatial averages have been selected based on the AICc. The regressors are lagged by one time period.

Results for Dutch residential densities

Table 4.4 presents the results. The static estimates provide strong evidence for clustering in household densities indicated by the high estimate of ρ and the high t-value. As theorized, we find weak evidence for a relationship with the surrounding household compositions indicated by the small estimate of $\beta_{w\%14}$ and its low t-value. Household densities are strongly linked to company densities, other controls have dubious signs. The negative effect of single person households is not as expected as small households should consume little space.

The second model allows for smooth transition nonlinearities in the dependence on surrounding households. The negative value of the constant exogenous spatial lag ($\beta_{w\%14-_{t-1}}$) combined with the positive value of the upper threshold parameter ($\delta_{w\%14-_{t-1}}$) indicates that the dependencies run from negative to positive as densities increase, in line with the theory. The parameters of the transition function strongly improve the AICc (by -11939 points). The nonlinear model also improves the estimates of the control variables, both local and surrounding single person households now correlate positively with densities. The effect of company densities is substantially smaller in magnitude, indicating that the impact may easily be overestimated by the SAR. The spatial autocorrelation parameter is significant but reduced drastically in magnitude. This suggests that the nonlinearities in the relationship with spatial averages may also partially capture nonlinear spatial autocorrelations.

Model (3) controls for additional nonlinear spatial autocorrelation, further improving the AICc (-1799 points). The maps in fig. 4.2 show that spatial autocorrelation is high in the urban clusters and decays outwards, in line with theory.

Table 4.4: Estimation results for Dutch residential densities from 2005-2014. Significance at 90, 95 and 99% level are, respectively, indicated as *, ** and ***. t-values in parenthesis.

	(1) SAR + WX	(2) SAR + ST-WX	(3) ST-SAR + ST-WX
β_{const}	2.309***	0.970***	0.974***
	(26.479)	(36.626)	(41.905)
$\beta_{cdens_{t-1}}$	1.043***	0.108***	0.109***
	(201.116)	(21.832)	(28.378)
$\beta_{wcdens_{t-1}}$	-0.825***	-0.101***	-0.147***
	(-58.556)	(-13.821)	(-31.429)
$\beta_{\%shh_{t-1}}$	-0.006***	0.009***	0.009***
	(-9.319)	(34.352)	(38.678)
$\beta_{w\%shh_{t-1}}$	-0.025***	0.013***	0.009***
	(-17.731)	(23.705)	(19.950)
$\beta_{w\%hhkids_{t-1}}$	-0.032***	0.018***	0.015***
	(-12.350)	(17.606)	(17.875)
$\beta_{\%nwim_{t-1}}$	0.019***	0.003***	0.001***
	(35.607)	(12.336)	(3.162)
$\beta_{w\%14-_{t-1}}$	0.009*	-0.432***	-0.385***
	(1.987)	(-27.409)	(29.141)
$\delta_{w\%14-_{t-1}}$		1.008	0.540
$\gamma_{w\%14-_{t-1}}$		0.209	0.362
$\phi_{w\%14-_{t-1}}$		1.742	0.276
δ_{ρ}	0.779***(69.605)	0.048***(7.106)	0.368
γ_{ρ}			1.235
ϕ_{ρ}			1.358
λ	3.013	2.508	2.559
LL	-2078.771	3893.733	4795.443
AICc	4179.58	-7759.405	-9558.810

The test proposed in Proposition 3, is valid only for large $\tilde{T}$. However, the AICc provides ample evidence supporting the nonlinearities. In particular, the AICc improves by 13738.39 points when the nonlinearities are allows in both the spatial lags of the exogenous and endogenous regressors. To understand how the ST-SAR improves this much, we re-fitted the models excluding the last year and compared te 1-step ahead forecast errors of the SAR+WX and the ST-SAR+ST-WX. Figure 4.3 shows that the Squared Forecast Errors (SFE) from the linear model contain a consistent mismatch in major urban areas. The SFE of the nonlinear model, however, balance evenly. This shows that the nonlinear model is better at fitting both rural and urban density regimes within one framework. Apart from the clustering of prediction errors, the predictive

power across all regions is tremendously improved by the nonlinear model as seen by the magnitude of the SFE.

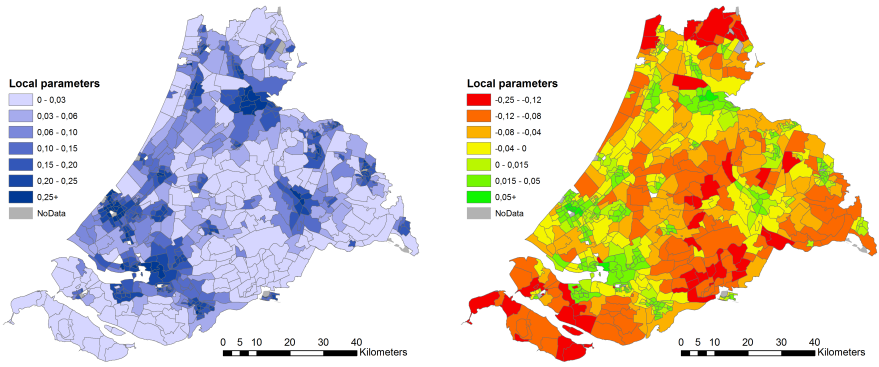

Figure 4.2: Left): time average of estimated autocorrelation parameters. Right): time average of estimated dependence on the share on population under 14 years in surrounding neighborhoods. The estimation results provide convincing evidence for weak/negative and strong/positive dependence regimes with smooth transitions in between.

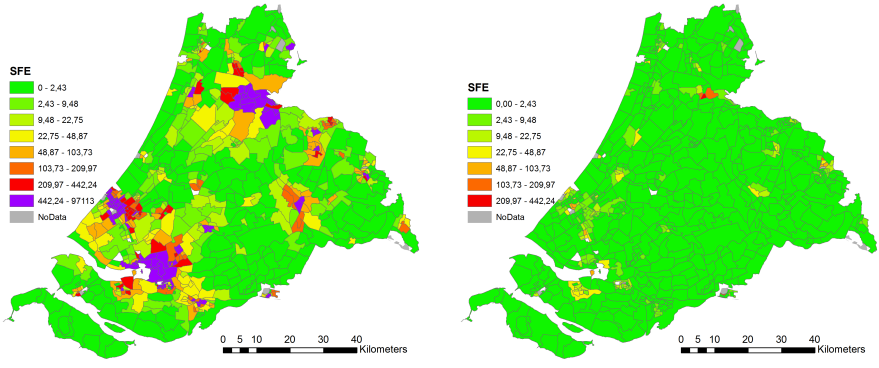

Figure 4.3: Left): SFE of the SAR (2014 as holdout data). Right): SFE of the ST-SAR. Legends are based on natural breaks of the errors of the ST-SAR. The residuals provide convincing evidence for the ability of the ST-SAR to neutralize residual clustering while the SAR does not perform well in this regard. The reduction in SFE also suggests that the ST-SAR provides better 1-step ahead forecasts.

4.5.2 Application II: interest rates in the Euro region

In this second empirical study we evaluate the evolution of monthly interest rates on government bonds maturing in ten years for 15 European sovereigns. The study tracks the European sovereigns over a period of 26 years, that spans the time before the European Union, the expansion of the EU, the Great Recession, and the Greek sovereign debt crisis. The primary focus is on detailing time-varying dynamics in convergence and dispersion in rates that cannot be fitted by a linear model. this application differs from the previous on in the sense that the temporal dimension is much larger. Again, we find strong evidence that favors the ST-SAR over the SAR.

Economic rationale for ST-SAR dynamics in long term interest rates

The Economic and Monetary Union (EMU) comprises a set of policies that aims at converging the economies of the member states of the European Union. The EMU prescribes euro convergence criteria, the prerequisites for a nation to join the Eurozone. Co-movement in the long term interest rates is essential to the monetary stability of the Euro region. Before the European Union, the European Economic Community relied heavily on the European Exchange Rate Mechanism (ERM) to regulate variability in exchange rates of different sovereigns as a way to achieve monetary stability. The ERM played a central role in the preparations for the Economic and Monetary Union and the subsequent introduction of the euro in 1999. The primary goal of the ERM has been to prevent large fluctuations in currency values relative to those of other European sovereigns. Empirical evidence suggests that only few, large industrial countries have some ability to choose their interests rates (Frankel et al., 2004). Interest rates are strongly effected by those of other countries (Frankel et al., 2004; Caceres et al., 2016; Kharroubi et al., 2016), but there are policy opportunities to adjust national rates. For

example, target levels and transfers of reserves (Pina, 2017), programs that increase foreign bond buying (Carvalho and Fidora, 2015), or lending rate policies (von Borstel et al., 2016). Adjustments in national interest rates have been at the center of monetary policy used as part of the European Monetary System (EMS) to lower or increase currency value such that the different currencies remained within a narrow range of one another.

Replacement of the actual currencies of all participating member states by a common currency mandates that the economies of all member states are in par with one another. After introduction of the euro, national interest rates thus still play an essential role in ensuring that fluctuations in the economies of member states remain within a narrow range. A strong adjustment in long term interest rate of a particular sovereign with respect to the common European average, signals that the underlying economy has difficulty in following the common trend. On the other hand, if all interest rates closely follow a common stochastic trend, it signals that economies are in par with on another. This can also be understood in the conventional framework where fixed or pegged interest rates are seen as a way to establish a credible nominal anchor for monetary policy, while flexible exchange rates are seen as a way to allow countries to pursue independent monetary policy (Frankel et al., 2004). Integration of financial systems and co-movements are further discussed by Caceres et al. (2016).

The cross-sectional dependencies in the de-trended changes signal the strength of commonalities in the fluctuations in the economies of member states such as in Caceres et al. (2016). Estimating spatial dependence parameters using ST-SAR has the obvious advantage that it does non only provide information on the average strength in co-movement, but it allows to study also the time-varying features in strength as well as heterogeneity across member states. The average cross-sectional averages

of the dependence parameters signal overall strength of convergence, while the standard deviations indicate an overall dispersion measure that is independent of the scale of change. In stable times, the average contraction should be high and the variance in spatial parameters should be low. Under financial instability we may expect the opposite. The parameters of the ST-SAR therefore not only provide means to filter dynamic dependencies, but also provide information on the functioning of the EMS in this specific application. Specifically, the ST-SAR provides a way to analyze whether the economies of member states are relatively in par with one another, as prescribed by the EMU's common currency mandates.

To do so, we view the interest rates as generated by the model:

$$\mathbf{v}_t = c_t + \mathbf{y}_t,$$

where $\mathbf{v}_t$ is the observed data vector, c_t is the common stochastic trend, and $\mathbf{y}_t$ is a vector of dynamics around the common stochastic trend. We are interested it analyzing $\mathbf{y}_t$, which contains the contraction and dispersion dynamics around the common stochastics. By de-trending using a common stochastic trend, synchronization due to common business cycles or seasonality is controlled for. We assume c_t to follow a random walk with $c_t = c_{t-1} + v_t$, and $\{v_t\}_{t\in\mathbb{Z}} \sim p_v(v_t, \Sigma, \lambda)$. Therefore our best expectation of c_t is $c_t \sim \mathbb{E}^N(\mathbf{v}_t|\mathbf{v}_{t-1})$, and the dynamics of particular interest are:

$$\mathbf{y}_t = \mathbf{v}_t - \mathbb{E}^N(\mathbf{v}_t|\mathbf{v}_{t-1}) = \mathbf{v}_t - \mathbb{E}^N(\mathbf{v}_{t-1}) \sim \mathbf{v}_t - N^{-1}\sum_{1}^{N}(\mathbf{v}_{t-1}),$$

hence we use $\mathbf{y}_t = \mathbf{v}_t - N^{-1}\sum_1^N(\mathbf{v}_{t-1})$ as our dependent variable. We refer to $\mathbf{y}_t$ as the de-trended data. We are interested in a description of the convergence and dispersion dynamics contained in $\mathbf{y}_t$ as a nonlinear cross-sectional dependence process, possibly driven by the past states of $\mathbf{y}_t$ and moving average affects.

Data for long term interest rates across the Euro region

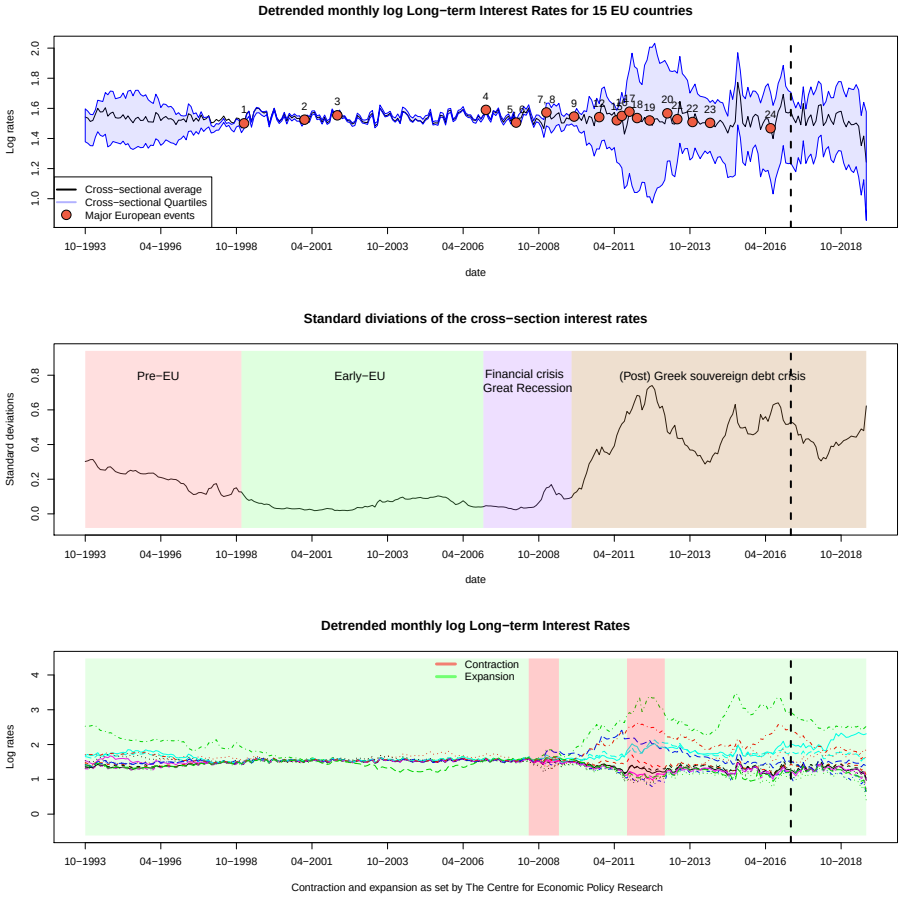

Figure 4.4: Data on monthly long term interest rates for bonds of 10-year maturity. An overview of te labeled events is contained in the Supplementary Appendix. The vertical dashed line indicates the split between training data and validation data used for our DM tests. Colors correspond to individual countries, see section 4.7.4.

The data was obtained from the European Central Bank for 311 months starting October 1993 and running through August 2019.[10] This period includes the formation of the European Union, its expansion, the Great Recession and the eventual Greek sovereign debt crisis. We model log

[10] http://sdw.ecb.europa.eu/browseChart.do?node=bbn4864

de-trended rates. The de-trended data is visualized in fig. 4.4, the raw data including a list of labeled events and color codes is provided in the Supplementary Appendix. The time series reveal clear common patterns, especially between 1998 and 2008. Before 1998 and after 2008 there are commonalities but specifically the stressed Eurozone sovereigns (Greece, Portugal, Ireland and to some extent Spain and Italy), seem to follow a separate pattern. Our network structure is based on the correlation matrix of the de-trended data. We assign each sovereign three neighbors based on the strongest correlation. This number was determined by the AICc. The approach allows for differences in the centrality of the sovereigns within the network, and for entanglement between sovereigns that are distant from each other in a purely geographic sense. The resulting network is fully connected. We explore time lags up to order 4, and apply further restrictions guided by the AICc. As we shall see in our final model, allowing 4 lags in the ST-SAR is sufficient to render the residuals approximately free from correlations.

Results for European long term interest rates

As a first exploration we regress models of the type:[11]

$$\mathbf{v}_t - N^{-1}\sum_1^N(\mathbf{v}_{t-1}) = \mathbf{y}_t = H(\boldsymbol{\theta}^\rho;(\mathbf{y}_t,\boldsymbol{\varepsilon}_t))^{-1}(\boldsymbol{\varepsilon}_t).$$

on the entire dataset. We calculate DM and mDM, respectively one-sided Pr(>|z|) and Pr(>|t|) against the null hypothesis that the SAR attains higher log likelihood, based on model fits on training data log likelihood evaluated on the validation sample depicted in fig. 4.4. We reserved the final 36 observations for this validation purpose, of which the first 6 observations are discarded as a burn-in.

[11] The exact threshold $\rho(\boldsymbol{\theta}^\rho;\mathbf{y}_t,\boldsymbol{\varepsilon}_t) = \dfrac{\delta}{1+\exp(-\gamma(W\mathbf{y}_{t-1}-(\alpha+\sum_{p=1}^P\mathbf{y}_{t-p}\varphi_{\phi,p}+\sum_{q=1}^Q\boldsymbol{\varepsilon}_{t-q}\varphi_{\mu,q})))}+\kappa$.

Table 4.5: Estimation results for the different spatial models on the full dataset. λ fixed at 2.5.

	SAR	ST-SAR (AR)	ST-SAR (MA)	ST-SAR (ARMA)
β_{const}	0.608***	-0.234***	0.310***	0.273***
	(56.670)	(-30.703)	(37.294)	(34.646)
κ_ρ	0.590***	0.762***	0.414***	0.366***
	(82.600)	(121.424)	(46.118)	(36.576)
δ_ρ		1.057	1.175	1.198
γ_ρ		-2.991	-0.240	-1.662
α_ρ		0.043	-1.641	-0.665
$\varphi_{\phi,t-1}$		0.911		1.069
$\varphi_{\phi,t-2}$				0.022
$\varphi_{\phi,t-3}$		-0.055		0.043
$\varphi_{\phi,t-4}$				0.108
$\varphi_{\mu,t-1}$			14.337	0.719
$\varphi_{\mu,t-2}$			13.776	0.496
$\varphi_{\mu,t-3}$			9.801	0.440
$\varphi_{\mu,t-4}$			3.756	0.213
LL	2314.54	8385.81	7823.34	9520.93
AICc	-4623.08	-16755.59	-15626.64	-19013.76
DM		0.00	0.00	0.00
mDM		0.00	0.00	0.00

Table 4.5 presents the estimation results from both the static and nonlinear spatial models for different specifications of the threshold. In the static model, we find strong evidence for spatial dependence indicated by the high estimate for ρ together with a high t-statistic. The three nonlinear specifications, respectively the ST-SAR driven by past observations, moving averages, and both ARMA dynamics, all improve the AICc values by several thousand points compared to the SAR. The most elaborate ST-SAR improves the AICc by an overwhelming 14390.68 points against the SAR. The significant evidence for nonlinearity is confirmed by the finding that the DM-type tests overwhelmingly reject the null of linearity, even for the most parsimonious ST-SAR. The residuals are in strong support of the choice to allow for fat tails. As an example, the kurtosis of residuals from the linear SAR is over 14 and a Jarque-Bera tests reject Gaussianity in favor of fatter tails with a p-value of ~ 0 for all four

models. Evidence for nonlinearities in the convergence and dispersion process persists across the different ST-SAR specifications. However, the SAR contains no time dynamics. We therefore extend our analysis to control for additional ARMA dynamics.

Extensions

In our extended results, we allow for additional flexibility and explore

$$\mathbf{v}_t - N^{-1}\sum_1^N(\mathbf{v}_{t-1}) = \mathbf{y}_t = H(\boldsymbol{\theta}^{\rho};(\mathbf{y}_t,\boldsymbol{\varepsilon}_t))^{-1}ARMA(\boldsymbol{\theta}^{\phi,\mu};\mathbf{y}_t,\boldsymbol{\varepsilon}_t).$$

Table 4.6: Estimation results for the extended spatial models on the full dataset. λ fixed at 2.5, constant omitted from table.

	SAR + ARMA	ST-SAR (ARMA) + ARMA
ϕ_{t-1}	-0.202***	0.110***
	(-15.200)	(2.904)
ϕ_{t-2}	0.063***	
	(8.050)	
ϕ_{t-3}	0.263***	0.423***
	(19.080)	(6.049)
ϕ_{t-4}	0.268***	0.258***
	(24.140)	(5.455)
μ_{t-1}	1.610***	0.819***
	(73.140)	(11.793)
μ_{t-2}	1.517***	0.612***
	(55.100)	(8.873)
μ_{t-3}	0.908***	0.314***
	(37.270)	(5.033)
μ_{t-4}	0.223***	-0.140***
	(15.870)	(-10.410)
κ_ρ	0.625***	
	(59.770)	0.404
δ_ρ		5.405
γ_ρ		-1.524
α_ρ		-0.602
$\varphi_{\phi,t-1}$		0.969
$\varphi_{\phi,t-3}$		-0.459
$\varphi_{\phi,t-4}$		-0.260
$\varphi_{\mu,t-1}$		-0.671
$\varphi_{\mu,t-2}$		-0.569
$\varphi_{\mu,t-3}$		-0.331
LL	8375.408	10146.85
AICc	-16728.76	-20255.54
DM		0.002
mDM		0.004

Table 4.6 presents the results. The ST-SAR dynamics remain significant as judged by the various diagnostics even when additional ARMA dynamics are added to the conditional mean equation. Importantly, the AICc of the ST-SAR improves over that of the SAR by a very significant amount, a 3526.78 point improvement. The out-of-sample validation test further confirms the evidence for nonlinearity. The estimated probability that the linear spatial model attains lower KL is below 0.002, and 0.004 for the more conservative modified test.

We also find that the residuals of the nonlinear model, similarly to our first application, are smaller and better centered at zero. This can be seen in fig. 4.9. The residuals of the SAR remain respectively below and above zero for prolonged periods and contain significant remaining correlation patterns while the ST-SAR approximately neutralizes the dynamics as revealed by the residual ACF in fig. 4.10. Jarque-Bera tests again reject Gaussianity in favor of fatter tails, supporting again the choice for the Student's-t specification, with a p-value of ~ 0 for both models, with the residuals of the ST-SAR (ARMA) + ARMA reaching a kurtosis of 22.

Figure 4.5 displays the evolution of the fitted spatial dependence parameters. A first striking feature is the convergence of the parameters in anticipation of the Union, continuing till around 2000. In the pre-EU period we observe separate regimes. Ireland, Portugal, Italy and Spain form a low-dependence group. Greece forms an exception and follows an individual trajectory. After 2000, the parameters corresponding to the different sovereigns linearize, indicating strong financial stability and near perfect co-movement. The onset of the Great Recession around 2008 marks an abrupt turn after which separation in a high and low regime recurs. Interestingly, the pattern after the recession reverts to the pre-EU behavior, with Greece returning to an individual trajectory and Ireland, Portugal, Italy and Spain forming a less integrated group. This breakaway is in sharp contrast to the increasing interdependence across

other member states. Divergence between the low and high dependence regimes has continued after the crisis, and the sustained strong variation in contraction parameters indicates that the Eurozone remains to struggle in attaining EU-wide financial stability. These results suggest that the EMS has still not fully succeeded in aligning all economies across the Eurozone. Figure 4.6 further visualizes the time-varying nature of the dependence regimes over cross-sections and time.

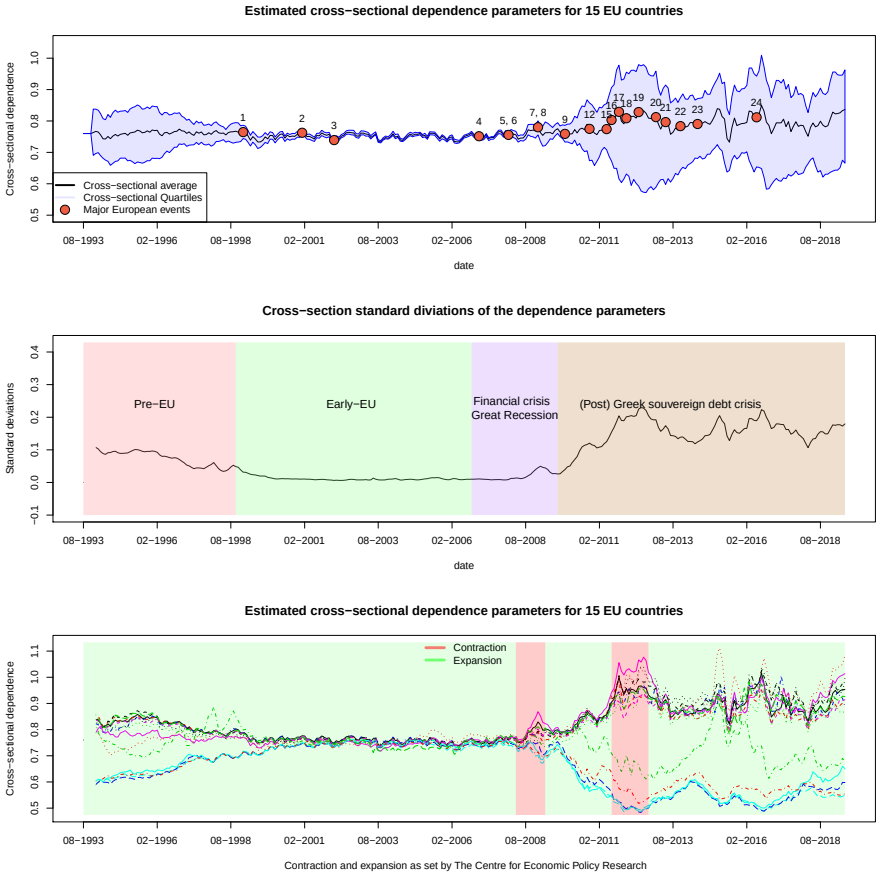

Figure 4.5: Evolution of spatial parameters estimated with the ARMA + ST-SAR (ARMA). Colors correspond to individual countries, see section 4.7.4. The estimation results highlight the nonlinear nature of dependence between sovereigns during Pre-EU times that has clearly broken into a two-regime system after the Financial Crisis.

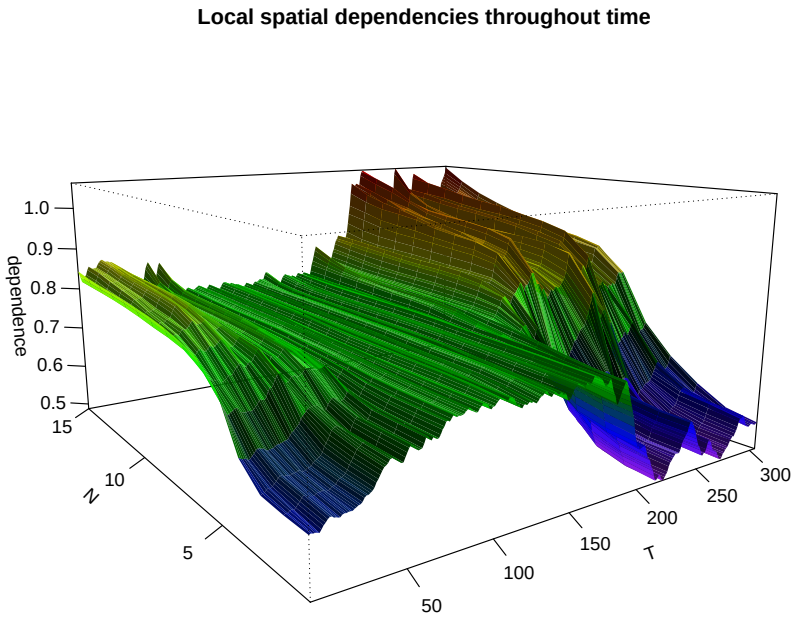

Figure 4.6: Evolution of spatial parameters estimated with the ST-SAR.

Several interesting aspects about the convergence and dispersion dynamics can be learned from our final estimates. Local dependencies are partly driven by feedback, but also impacted by moving averages that may relate to directed financial policy or shocks. Since $\hat{\gamma} = -1.524$ is negative, and $\hat{\delta} = 5.405$ is positive, the spatial parameters increase with $W\mathbf{y}_{t-1} - \hat{\tau}_{t-1}$, thus the signs of the estimated φ parameters indicate the direction of individual contributions.[12] The complex threshold equation hints at

[12]The estimated threshold is $\hat{\tau}_{t-1} = -.969\mathbf{y}_{t-1} + .459\mathbf{y}_{t-3} + .260\mathbf{y}_{t-4} + .671\boldsymbol{\varepsilon}_{t-1} + .569\boldsymbol{\varepsilon}_{t-2} + .331\boldsymbol{\varepsilon}_{t-3} - .602$. Note that the signs in the table 4.6 are opposite as they enters as $-\hat{\tau}_{t-1}$ in the likelihood function.

several subtleties. The negative signs of the moving averages suggest that positive shocks are followed by reduced dependence, while sustained exogenous policies that reduce rates result in increased contraction. If all effects are considered jointly, the following regime-dependent behavior can be distinguished:

1. $\hat{\tau}_{t-1} < W\mathbf{y}_{t-1}$ local threshold value is below average neighbor rates, followed by intensified dependence (dispersion),
2. $\hat{\tau}_{t-1} > W\mathbf{y}_{t-1}$ local threshold value is above average neighbor rates, followed by reduced dependence (convergence).

These regimes suggest cyclic behavior. First, high rates relative to neighboring sovereigns due to exogenous impacts (high $\boldsymbol{\varepsilon}_t - q$ for $q = 1, ..., 4$) is followed by reduced dependence to the group average, making isolated rate increases due to shock possible. Once assimilated, high relative rates (high $\mathbf{y}_t - p$ for $p = 1, ..., 4$ relative to $W\mathbf{y}_{t-1}$) is followed by intensified spatial dependence. Together this implies initial systemic vulnerability to exogenous shocks, but subsequent resistance to the spread of assimilated shocks. That resistance breaks when a large neighborhood is affected (W_{t-1} increases), accelerating the spread through increased feedback. Finally, the negative signs of deeper lags of $\mathbf{y}_t - p$ indicate that initial increases in contraction are followed by a return to reduced dependence, slowing feedback.

The regimes also suggest asymmetries in spillovers. If at location i rates increase, dependence to neighbor j reduces. From the perspective of location j the opposite occurs, resulting in the opposite dynamics. This means that while a local positive impulse lowers spatial dependence locally, it increases the dependence parameters of neighbors, implying that outward spillovers accelerate while inward feedback slows down. On the other hand, lowered rates are followed by intensified inward spillovers but slower outward spillovers.

4.6 Conclusion

In this paper we introduced a new model for nonlinear spatial time series in which cross-sectional dependence varies smoothly over space by means of smooth-transitions between dependence regimes. In this framework, nonlinearities in cross-sectional dynamics are modeled as a function of the data. This is an advance over existing methods. Allowing for time-variation is particularly useful when modeling spatial data for large T, nonlinearities over the cross-section are particularly useful if N is large.

We have shown that the parameters of the model can be consistently estimated by maximum likelihood under appropriate regularity conditions. In particular, we provide conditions that deliver existence, strong consistency and asymptotic normality of the MLE of all static parameters that constitute the dynamic dependence structure. The theory holds for both correctly specified and misspecified models and allows for possible identification issues of the threshold parameters. Our simulation evidence suggests that the limit theory is relevant in finite samples. Furthermore, we find that information criteria are able to distinguish between the SAR specification and ST-SAR type nonlinearities. The simulations results showed that model selection is robust to overfitting of additive outliers. We have also provided a theoretical argument for model selection based on a validation-sample estimate of the Kullback-Leibler divergence. In our empirical application, both the validation test and the information criteria support nonlinearities.

The model has been applied to study space-time dynamics in two cases. We studied clustering in urban densities in a large number of districts, and convergence and dispersion in monthly long term interest rates. We found that the ST-SAR resulted in better filtering behavior over the cross-section and time dimension, improved estimates for exogenous variables,

and improved forecasts. We also found that the nonlinearities in the spatial parameters can lead to economically relevant insights, while the SAR is often criticized for its empirical interpretation. We conclude that the ST-SAR is a powerful tool for both understanding and predicting future values in cross-sectional time series.

4.7 Appendix

4.7.1 Proofs to main theorems

Proof of Theorem 9

Proof. Note first that $L_T(\boldsymbol{\theta}) := (1/T)\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta})$ is a.s. continuous (a.s.c.) in $\boldsymbol{\theta} \in \Theta$ through continuity (c.) of each term $\ell_t(\boldsymbol{\theta}) = \ln \det H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t) + \ln p_{\boldsymbol{\varepsilon}}\Big(H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)\mathbf{y}_t - c - \mathbf{y}_{t-1}\phi - \mathbf{X}_t\boldsymbol{\beta}, \Sigma; \lambda\Big)$. Together with the compactness of Θ (Assumption 1) this implies by Weierstrass' theorem that the arg max set is non-empty a.s. and hence that $\hat{\boldsymbol{\theta}}_T$ exists a.s. $\forall T \in \mathbb{N}$. Note by a similar argument that $L_T(\boldsymbol{\theta})$ is continuous in $(\mathbf{y}_t, \mathbf{X}_t)$ $\forall\, \boldsymbol{\theta} \in \Theta$ and hence measurable w.r.t. the product Borel σ-algebra $\mathfrak{B}(\mathcal{Y}) \otimes \mathfrak{B}(\mathcal{X})$ that are, in turn, measurable maps w.r.t. $\mathcal{F}$ by Proposition 4.1.7 in Dudley (2002).[13] Finally, the measurability of $\hat{\boldsymbol{\theta}}_T$ follows from (Foland, 2009, p.24) and (White, 1994, Theorem 2.11) or (Gallant and White, 1988, Lemma 2.1, Theorem 2.2).[14] □

Proof of Theorem 10

Proof. Recall that $L_T(\boldsymbol{\theta}) := (1/T)\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta})$ and $L_\infty(\boldsymbol{\theta}) = \mathbb{E}\ell_t(\boldsymbol{\theta})$ with

$$\ell_t(\boldsymbol{\theta}) = \ln \det H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t) + \ln p_{\boldsymbol{\varepsilon}}\Big(H(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)\mathbf{y}_t - c - \mathbf{y}_{t-1}\phi - \mathbf{X}_t\boldsymbol{\beta}, \Sigma; \lambda\Big).$$

Following the usual consistency argument (found e.g. in (White, 1994, Theorem 3.4) or Theorem 3.3 in Gallant and White (1988)) we obtain

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

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

$\hat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ from the uniform convergence of the criterion function

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_T(\boldsymbol{\theta}) - L_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0 \ \forall \ f_1 \in \mathcal{F} \ \text{ as } \ T \to \infty \tag{4.12}$$

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

$$\sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\boldsymbol{\varepsilon}} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0) \ \forall \ \boldsymbol{\varepsilon} > 0. \tag{4.13}$$

The uniform convergence is obtained by application of the ergodic theorem for separable Banach spaces in Rao (1962), as in (Straumann and Mikosch, 2006, Theorem 2.7), to the sequence $\{L_T(\cdot)\}$ with elements taking values in $\mathbb{C}(\Theta, \mathbb{R})$. This uniform law of large numbers $\sup_{\boldsymbol{\theta}\in\Theta} |L_T(\boldsymbol{\theta}) - \mathbb{E}\ell_t(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$ as $T \to \infty$ follows, under a uniform moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| < \infty$, by the SE nature of $\{L_T\}_{t\in\mathbb{Z}}$ which is implied by continuity of ℓ on the SE sequence $\{(\mathbf{y}_t, \mathbf{X}_t)\}_{t\in\mathbb{Z}}$ (Assumption 2) and Proposition 4.3 in Krengel (1985). The uniform moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| < \infty$ follows immediately from Assumption 9 since

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_t(\boldsymbol{\theta})| \leq \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |Q(\boldsymbol{\theta}^\rho; \mathbf{Z}_t)| + |A(\boldsymbol{\theta})|$$

$$+\frac{1}{2}(\lambda + N)\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |F(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{X}_t, \mathbf{Z}_t)| < \infty.$$

Finally, the identifiable uniqueness (see e.g. White (1994)) of $\boldsymbol{\theta}_0 \in \Theta$ in (4.13) follows from the assumed uniqueness (Assumption 10), the compactness of Θ, and the continuity of the limit $\mathbb{E}\ell_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$ which is implied by the continuity of L_T in $\boldsymbol{\theta} \in \Theta \ \forall \ T \in \mathbb{N}$ and the uniform convergence in (4.12). □

Proof of Theorem 11

Proof. The proof follows by the same argument as laid down in the proof of Theorem 2. Only now the assumption, that $\boldsymbol{\theta}_0$ is the unique maximizer, is missing (Assumption 10). Without uniqueness, we obtain the desired set consistency result by application of Lemma 4.3 in (Pötscher and Prucha, 1997), after noting that the continuity of the limit criterion $L_\infty(\boldsymbol{\theta}) = \mathbb{E}\ell_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$ and the compactness of Θ ensure that the

levels sets of L_∞ are regular (see Definition 4.1 in Pötscher and Prucha (1997)). The continuity of L_∞ is obtained directly from the continuity of $\ell_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$ for every t, and the uniform convergence of the sample criterion $\frac{1}{T}\sum_{t=1}^{T} \ell_t$ to the limit L_∞. □

Proof of Theorem 12

Proof. The desired result follows immediately by application of Theorem 11 after noting that the data generated by the ST-SAR model converges to a unique SE solution by Propositions 1 and 2. □

Proof of Theorem 13

Proof. We obtain the asymptotic Gaussianity of the MLE immediately from (i) the strong consistency of $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$; (ii) the a.s. twice continuous differentiability of $\ell_T(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$; (iii) the asymptotic normality of the score

$$\sqrt{T} L_T'(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{J}(\boldsymbol{\theta}_0)), \qquad \mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}\big(\ell_t'(\boldsymbol{\theta}_0)\ell_t'(\boldsymbol{\theta}_0)^\intercal\big); \tag{4.14}$$

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

$$\sup_{\boldsymbol{\theta}\in\Theta} \big\| L_T''(\boldsymbol{\theta}) - L_\infty''(\boldsymbol{\theta}) \big\| \xrightarrow{a.s.} 0; \tag{4.15}$$

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

The consistency condition $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ in (i) follows by Theorem 2 and the additional assumption that $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$.

The smoothness condition in (ii) is trivially satisfied for the student's-t density.

The asymptotic normality of the score in (4.16) follows by Theorem 18.10[iv] in van der Vaart (2000) by an application of the CLT for SE martingales in Billingsley (1961) or NED processes in Pötscher and Prucha (1997) Theorem 10.2, to obtain

$$\sqrt{T} L_T'(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{J}(\boldsymbol{\theta}_0)) \;\text{ as }\; T \to \infty, \tag{4.16}$$

where $\mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}(\ell_t'(\boldsymbol{\theta}_0)]\ell_t'(\boldsymbol{\theta}_0)^\intercal) < \infty$. The SE nature of $\{L_T'(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$

follows by continuity of L'_T on the SE sequence $\{(\mathbf{y}_t, \mathbf{X}_t)\}_{t\in\mathbb{Z}}$; see Proposition 4.3 in Krengel (1985). Assumption 5 imposes the mds or NED nature of the score sequence $\{\ell'_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$. The finite (co)variances follow from the first two moments bounds of Assumption 6.

The uniform convergence in (iv) is obtained under the moment bound

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\ell''_t(\boldsymbol{\theta})\| < \infty$$

and by the SE nature of $\{\ell''_T\}_{t\in\mathbb{Z}}$. The moment bound is ensured by Assumption 6. The SE nature is implied by continuity of ℓ'' on the SE sequence $\{\mathbf{y}_t, \mathbf{X}_t\}_{t\in\mathbb{Z}}$.

Finally, the non-singularity of the limit $L''_\infty(\boldsymbol{\theta}) = \mathbb{E}\ell''_t(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$ in (v) is implied by the uniqueness of $\boldsymbol{\theta}_0$ as a maximum of $L''_\infty(\boldsymbol{\theta})$ in Θ. □

Proof of Lemma 1

Proof. Expand $\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T)$ at $\boldsymbol{\theta}_0^*$ to obtain

$$\lim_{T,\tilde{T}\to\infty}\mathbb{E}\Big(\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0^*)\Big) = \lim_{T,\tilde{T}\to\infty}\mathbb{E}\tilde{L}'_{\tilde{T}}(\boldsymbol{\theta}_T^*)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*)$$

Next, use the uniform moment $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{L}'_T(\boldsymbol{\theta})| < \infty$ to interchange the limit and expectation by appealing to a dominated convergence theorem, and use Slutsky's theorem to obtain,

$$\lim_{T,\tilde{T}\to\infty}\mathbb{E}\tilde{L}'_{\tilde{T}}(\boldsymbol{\theta}_T^*)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*) = \lim_{\tilde{T}\to\infty}\mathbb{E}\lim_{T\to\infty}\tilde{L}'_{\tilde{T}}(\boldsymbol{\theta}_T^*)\lim_{T\to\infty}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*)$$

Finally, use the continuity of ℓ, a continuous mapping theorem, and the consistency of the MLE to obtain the desired result

$$\lim_{T,\tilde{T}\to\infty}\mathbb{E}\Big(\tilde{L}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}\tilde{L}_{\tilde{T}}(\boldsymbol{\theta}_0^*)\Big) = \lim_{\tilde{T}\to\infty}\mathbb{E}\tilde{L}'_{\tilde{T}}(\boldsymbol{\theta}_0^*)\times 0 = 0 \quad \text{a.s.}$$

□

Proof of Lemma 2

Proof. Obtained immediately by the same argument as that of Lemma 1. □

Proof of Proposition 3

Proof. Follows immediately by noting that under the null H_0 : $\mathbb{E}\tilde{\Delta}_t(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B}) = 0$, we have

$$\lim_{T,\tilde{T}\to\infty} \tilde{T}^{\frac{1}{2}} \left(\tilde{T}^{-1} \sum_t^{\tilde{T}} \frac{\tilde{\Delta}_t(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)}{\tilde{\sigma}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)} \right) = \lim_{\tilde{T}\to\infty} \tilde{T}^{\frac{1}{2}} \left(\tilde{T}^{-1} \sum_t^{\tilde{T}} \frac{\lim_{T\to\infty} \tilde{\Delta}_t(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)}{\lim_{T\to\infty} \tilde{\sigma}_{\tilde{T}}(\hat{\boldsymbol{\theta}}_T^A, \hat{\boldsymbol{\theta}}_T^B)} \right)$$

$$= \lim_{\tilde{T}\to\infty} \tilde{T}^{\frac{1}{2}} \left(\tilde{T}^{-1} \sum_t^{\tilde{T}} \frac{\tilde{\Delta}_t(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B})}{\tilde{\sigma}_{\tilde{T}}(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B})} - \tilde{\Delta}_t(\boldsymbol{\theta}_0^{*A}, \boldsymbol{\theta}_0^{*B}) \right) \overset{d}{=} \mathcal{N}(0,1).$$

The first equality is obtained by Slutsky's Theorem. The second equality by consistency $\hat{\boldsymbol{\theta}}_T^A \overset{as}{\to} \boldsymbol{\theta}_0^{*A}$ and $\hat{\boldsymbol{\theta}}_T^B \overset{as}{\to} \boldsymbol{\theta}_0^{*B}$ and the a.s. continuity of $\tilde{\Delta}_t$ and $\tilde{\sigma}_{\tilde{T}}$ on $\Theta \times \Theta$. The last equality, in distribution, is obtained by application of the CLT for strictly stationary and ergodic martingale difference sequences in Billingsley (1961). □

4.7.2 Additional results

LEMMA. 3. *Let A be an arbitrary finite-dimensional matrix. For an induced matrix norm $\|A\| < 1$ the following inequality is implied:*

$$(1 + \|A\|)^{-1} \leq \|(I_N - A)^{-1}\| \leq (1 - \|A\|)^{-1},$$

with $0 < (1 + \|A\|)^{-1}$ and $(1 - \|A\|)^{-1} < \infty$. By the finite-dimensionality we can also write

$$0 < c \leq \|(I_N - A)^{-1}\|_\infty \leq C < \infty,$$

for some positive constants c and C.

For a matrix H defined by $H = (I_N - A)$, LEMMA. 3 provides existence, non-negativity, and boundedness of the inverse H^{-1} for finite dimensional H. This is useful since throughout our theory as we always work with a fixed N and let only T tend to infinity.

LEMMA. 4. *Let A be an arbitrary matrix with eigenvalues $\omega_1, ..., \omega_n \in \mathbb{C}^{n \times n}$, real or complex, and $r(A) = \max\{|\omega_1|, ..., |\omega_n|\}$ be its spectral radius. If $r(A) < 1$ there exists $\|A\| < 1$ for some induced matrix norm.*

LEMMA. 4 allows the condition $\|A\| < 1$ in LEMMA. 3 to be replaced by $r(A) < 1$ if no suitable norm can be found. In what follows we will continue stating $\|\cdot\|$, but remind the reader that in practice one may focus on sample estimates of $r(\cdot)$ as a rule of thumb.

LEMMA. 5. *For any $H^{-1} \in \mathbb{R}^{n \times n}$ defined as $H^{-1} = (I_N - A)^{-1}$ with $N < \infty$ and $r(A) < 1$, we have that the following is implied*

i $\det(H^{-1}) > 0$,

ii $\log \tau(H^{-1})^N \leq \log \det(H^{-1}) \leq \log r(H^{-1})^N < \infty$,

iii $|\log \det(H^{-1})| < \infty$,

Claim *iii* in LEMMA. 5 is particularly useful in establishing that ASSUMPTION. 9 holds under correct specification.

4.7.3 Proofs for additional results

Proof of Lemma 3

Proof. The result follows by taking norms in $I_N = (I_N - A)(I_N - A)^{-1}$, which gives[15]

$$1 \leq \|I_N - A\|\|(I_N - A)^{-1}\|.$$

This can be rearranged to obtain

$$1 \leq (1 + \|A\|)\|(I_N - A)^{-1}\|.$$

Multiplying by $\|(I_N - A)^{-1}\|^{-1}$ gives

$$\|(I_N - A)^{-1}\|^{-1} \;\leq\; \|I_N - A\| \;\leq\; (1 + \|A\|),$$

thus

$$(\|(I_N - A)^{-1}\|)^{-1} \;\leq\; (1 + \|A\|),$$

$$(1 + \|A\|)^{-1} \;\leq\; \|(I_N - A)^{-1}\|,$$

providing the first inequality.

The second inequality follows immediately by the fact that the operator norm is sub-multiplicative. In particular, $\|I\| = \|B \cdot B^{-1}\| \leq \|B\| \cdot \|B^{-1}\|$ implies that $\|B\|^{-1} \leq \|B^{-1}\|$. Hence $(1 + \|A\|)^{-1} < \|(I - A)^{-1}\|$.

Finiteness of $\|(I_n - A)^{-1}\|$ follows trivially from

$$(1 - \|A\|)^{-1} \;< \infty. \;\text{ since } \|A\| < 1.$$

Non-negativity of $(I_n - A)^{-1}$ follows by noting that all its eigenvalues are non-zero. The minimum eigenvalue of a non-singular matrix is equal to the inverse of the spectral radius of the inverse matrix, thus in this case $\tau((I_n - A)^{-1}) = r(I - A)^{-1}$. Having just established that $(1+\|A\|)^{-1} \leq \|(I_n-A)^{-1}\| \leq (1-\|A\|)^{-1}$ it follows trivially that like-wise

$$(1 + \|A\|) \geq \|(I_n - A)\| \geq (1 - \|A\|),$$

[15]The result is similar to Proposition 6.4.1. in Lange (1999), but reworked here because both the proof and the final result are partial.

which delivers the upper bounds of $r(I_n - A)$ by noting that $r(I_n - A) \leq \|I_n - A\|$, hence $r(I - A)^{-1} > 0$, and equally so $\tau((I_n - A)^{-1}) > 0$.

Finally, by noting that any two norms in finite dimension $n < \infty$ are always within a constant factor of one another, such that we can write for some real numbers $0 < c_1 \leq c_1 \leq c_2$ the inequality

$$c_1\|(I_n - A)^{-1}\|_\infty \leq \|(I_n - A)^{-1}\| \leq c_2\|(I_n - A)^{-1}\|_\infty,$$

proves the second claim by setting $c = c_1\|(I_n - A)^{-1}\|_\infty$ and $C = c_2\|(I_n - A)^{-1}\|_\infty$. □

Proof of Lemma 4

Proof. This follows from by noting that for any matrix A and any positive number $e > 0$, there exists an induced matrix norm $\|A\|$ such that

$$r(A) \leq \|A\| < r(A) + e.$$

See Proposition 6.3.2. Lange (1999). Trivially,

$$r(A) < 1 \implies 1 - r(A) > 0.$$

Choose $e = 1 - r(A)$, the proof is completed by noting that we can now write

$$\begin{array}{rcccl} r(A) & \leq & \|A\| & < & r(A) + 1 - r(A), \\ r(A) & \leq & \|A\| & < & 1. \end{array}$$

□

Proof of Lemma 5

Proof. The proof of all three claims starts by noting that by definition (slight abuse of notation: reintroducing p and k)

$$\det(H^{-1}) = \left(\Pi_{i=1}^{k}\omega_i\right)\left(\Pi_{i=k+1}^{p}\omega_i\bar{\omega}_i\right) = \left(\Pi_{i=1}^{k}\omega_i\right)\left(\Pi_{i=k+1}^{p}|\omega_i|^2\right),$$

with $\omega_1, ..., \omega_N \in \mathbb{C}^{N\times N}$, real or complex, as the eigenvalues of H^{-1}. Hence the first claim follows by showing that

$$\left(\Pi_{i=1}^{k}\omega_i\right)\left(\Pi_{i=k+1}^{p}|\omega_i|^2\right) > 0.$$

Thus we need to show that $\omega_i > 0 \; \forall \; i \in 1, ..., N$, since then $(|\omega_i|^2)^{p-k} > 0$, and $\omega_i^{N-p} > 0$, hence the left side of the second equality is strictly positive. Note that LEMMA. 3 and LEMMA. 4 deliver the following inequality under assumptions of LEMMA. 5,

$$(1 + \|A\|)^{-1} \leq \|(I_N - A)^{-1}\| \leq (1 - \|A\|)^{-1},$$

which we can also write as

$$(1 - \|A\|) \leq \|(I_N - A)\| \leq (1 + \|A\|).$$

The desirable result follows by proving that $\tau(H^{-1}) > 0$, where $\tau(H^{-1}) = \tau((I_N - A)^{-1}) = \min\{|\omega_1|, ..., |\omega_N|\}$. Applying the useful identity $\tau(A) = (r(A^{-1}))^{-1}$, we have

$$\tau(H^{-1}) = (r(H))^{-1},$$

hence showing that $\tau(H^{-1}) > 0$ equals showing that $(r(H))^{-1} > 0$, which follows from $r(H) < \infty$. Using the general inequality $r(H) \leq \|H\|$ we can write $r(H) \leq \|(I_N - A)\| \leq (1 + \|A\|)$ thus proving $\tau(H^{-1}) > 0$.

Using the definition of $\det(H^{-1})$, and the bounds of H^{-1} we obtain the range of the determinant by allowing the finite number of N eigenvalues to be either strictly minima or maxima

$$0 < (\tau(H^{-1})^N \leq \det(H^{-1}) \leq r(H^{-1})^N < \infty.$$

The second claim follows easily now by taking logs and applying Jensen's inequality.

Finally, the third claim follows by noting that $0 < \det(H^{-1})$ implies that the log is defined, hence its absolute value is finite.

□

Proof of Proposition 1

Proof. The result follows by Theorem 2.2 and Example 2.1 in Cline and Pu (1998). In particular, we note first that LEMMA. 3 provides the uniform bound of $H(\mathbf{y})^{-1}$ by noting that if $\sup_{\mathbf{y} \in \mathbb{R}^N} \|\rho(\boldsymbol{\theta}_0^\rho; \mathbf{y}) \circ W\| < 1$

we have

$$0 < \bar{h} \leq \sup_{\mathbf{y}\in\mathbb{R}^N} \|H(\mathbf{y})^{-1}\| \leq \bar{H} < \infty$$

with

$$\bar{h} = \left(1 + \sup_{\mathbf{y}\in\mathbb{R}^N} \|\rho(\boldsymbol{\theta}_0^\rho; \mathbf{y}) \circ W\|\right)^{-1}, \quad \bar{H} = \left(1 - \sup_{\mathbf{y}\in\mathbb{R}^N} \|\rho(\boldsymbol{\theta}_0^\rho; \mathbf{y}) \circ W\|\right)^{-1}.$$

Having just established that $\sup_{\mathbf{y}\in\mathbb{R}^N} \|H(\mathbf{y})^{-1}\|$ is bounded away from zero by some constant $\bar{h}$ and from infinity by some constant $\bar{H} < \infty$, and that $H^{-1}(\mathbf{y})$ is invertible, we can now verify that the assumptions in Theorem 2.2 and Example 2.1 in Cline and Pu (1998) hold. First we note that $H(\mathbf{y})$ and $H(\mathbf{y})^{-1}$ are both trivially locally bounded, and that $\boldsymbol{\varepsilon}_t$ has full support. Finally, we note that $H(\mathbf{y})^{-1}\mathbf{y}\phi$ is also locally bounded since

$$\sup_{\|\mathbf{y}\|\leq M} \|H(\mathbf{y})^{-1}\mathbf{y}\phi\| \leq \sup_{\mathbf{y}} \|H(\mathbf{y})^{-1}\||\phi| \sup_{\|\mathbf{y}\|\leq M} \|\mathbf{y}\| \leq BM\phi < \infty \quad \forall\, M > 0.$$

□

Proof of Proposition 2

Proof. We recall that $\sup_{\mathbf{y}\in\mathbb{R}^N} \|H(\mathbf{y})^{-1}\| \leq B < \infty$ under the assumptions of Proposition 1. Next we obtain the desired result from Theorem 3.1 of Cline and Pu (1999). First, we note that $H(\mathbf{y})^{-1}\mathbf{y}$ is trivially unbounded in $\mathbb{R}^N$. Second, we have that $H(\mathbf{y})^{-1}\mathbf{y}/(1 + \|\mathbf{y}\|)$ is bounded in $\mathbb{R}^N$ since

$$\|H(\mathbf{y})^{-1}\mathbf{y}\phi/(1 + \|\mathbf{y}\|)\|$$
$$\leq \|H(\mathbf{y})^{-1}\|\|\mathbf{y}\||\phi|/(1 + \|\mathbf{y}\|) \leq B\|\mathbf{y}\||\phi|/(1 + \|\mathbf{y}\|) \leq B|\phi|.$$

Next, we note that

$$\sup_{\|\mathbf{y}\|\leq M} \mathbb{E}\|H(\mathbf{y})^{-1}\boldsymbol{\varepsilon}_t\|^r \leq \sup_{\|\mathbf{y}\|\in\mathbb{R}^N} \|H(\mathbf{y})^{-1}\|\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^r \leq B\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^r < \infty$$

$$\text{for every } M > 0.$$

Additionally, it holds trivially true that

$$\lim_{\|\mathbf{y}\|\to\infty} \mathbb{E}\frac{\|H(\mathbf{y})^{-1}\boldsymbol{\varepsilon}_t\|^r}{\|\mathbf{y}\|^r} \leq \lim_{\|\mathbf{y}\|\to\infty} \frac{B\|\boldsymbol{\varepsilon}_t\|^r}{\|\mathbf{y}\|^r} \to 0.$$

Furthermore, it holds true that

$$\lim_{\substack{\|H(\mathbf{y})^{-1}\mathbf{y}\phi\|\to\infty \\ \|\mathbf{y}-\mathbf{y}'\|/\|\mathbf{y}\|\to 0}} \left\| \frac{H(\mathbf{y})^{-1}\mathbf{y}\phi}{1+\|\mathbf{y}\|} - \frac{H(\mathbf{y}')^{-1}\mathbf{y}'\phi}{1+\|\mathbf{y}'\|} \right\|$$

$$= \lim_{\substack{\|\mathbf{y}\|\to\infty,\ \|\mathbf{y}'\|\to\infty \\ \|\mathbf{y}-\mathbf{y}'\|/\|\mathbf{y}\|\to 0}} \left\| \frac{H(\mathbf{y})^{-1}\mathbf{y}\phi}{1+\|\mathbf{y}\|} - \frac{H(\mathbf{y}')^{-1}\mathbf{y}'\phi}{1+\|\mathbf{y}'\|} \right\|$$

since $H(\mathbf{y})^{-1}$ is uniformly bounded in $\mathbf{y}$, and hence,

$$\|H(\mathbf{y})^{-1}\mathbf{y}\phi\| \to \infty \quad \Leftrightarrow \quad \|\mathbf{y}\| \to \infty \ ,$$

$$\text{and} \quad \Big\{ \ \|\mathbf{y}\| \to \infty \quad \wedge \quad \|\mathbf{y}-\mathbf{y}'\|/\|\mathbf{y}\| \to 0 \ \Big\} \quad \Leftrightarrow \quad \|\mathbf{y}'\| \to \infty.$$

As a result

$$\lim_{\substack{\|\mathbf{y}\|\to\infty,\ \|\mathbf{y}'\|\to\infty \\ \|\mathbf{y}-\mathbf{y}'\|/\|\mathbf{y}\|\to 0}} \left\| \frac{H(\mathbf{y})^{-1}\mathbf{y}\phi}{1+\|\mathbf{y}\|} - \frac{H(\mathbf{y}')^{-1}\mathbf{y}'\phi}{1+\|\mathbf{y}'\|} \right\| = \|H_\infty\phi - H_\infty\phi\| = 0.$$

Finally, we also have

$$\limsup_{\|\mathbf{y}\|\to\infty} \frac{\|H(\mathbf{y})^{-1}\mathbf{y}\phi\|}{\|\mathbf{y}\|} \leq \limsup_{\|\mathbf{y}\|\to\infty} \frac{\|H(\mathbf{y})^{-1}\|\|\mathbf{y}\|\|\phi|}{\|\mathbf{y}\|} = \|H_\infty\||\phi| < 1.$$

□

4.7.4 Additional Monte Carlo results and figures

In this additional experiment, we investigate whether the MLE is well-behaved and approximately normal for increasing sample sizes in the case of identified parameters. This itself is not the most interesting result to study, but it confirms that our theory is correct. The data generating process is of the form:

$$\mathbf{y}_t = H(\boldsymbol{\theta}^\rho; \mathbf{y}_{t-1})^{-1}(\boldsymbol{\varepsilon}_t), \quad \boldsymbol{\varepsilon}_t \sim TID(1, I_N; 5), \tag{4.17}$$

We set the parameters values to

$$\delta = 1.35,\ \gamma = 1.05,\ \alpha = -.2,\ \varphi = 1.4,\ \kappa = -.4,$$

$$\mathbf{Z}_t = \mathbf{y}_{t-1},\ \tau(\boldsymbol{\theta}^\tau; \mathbf{Z}_t) = \alpha + \varphi/N \textstyle\sum_1^N (\mathbf{y}_{t-1}),$$

which satisfies the conditions for geometric ergodicity and allows for local positive and negative clustering.

We keep the ratio of distant and close-by neighbors comparable across experiments by allowing the network density of the weights matrix to increase with N. In each draw we generate a random zero diagonal row-normalized weights matrix with $N/10$ neighbors for each observation. The process is initialized with $H_1 = I_N$, and the first 50 steps of the sequence are discarded to avoid dependence on the initialization. We simulate 1000 datasets and estimate the parameters of the ST-SAR with Student's-t likelihood. We consider samples of size $T = 25,\ 100,\ 250$ for $N = 30$. Figure 4.7 presents kernel density estimates of the distribution of the MLE for the different sample sizes.

Figure 4.7 presents the results and shows that for small sample sizes the estimators are not perfectly normal. For larger sample sizes, we see a fast convergence towards the limiting result. A second experiment with $N = 60$ was also performed, we noticed improvements in the distributions for small T as N grows. The results indicate that for an empirically relevant signal and sample size the MLE is well-behaved. Note that these results do not directly generalize to any empirical setting. Specifically, (near)-linear signals will cause identification problems even in larger samples that break the uniqueness assumption required for normality. However, our main simulation results show that information criteria can be used to assess the presence and significance of nonlinearity.

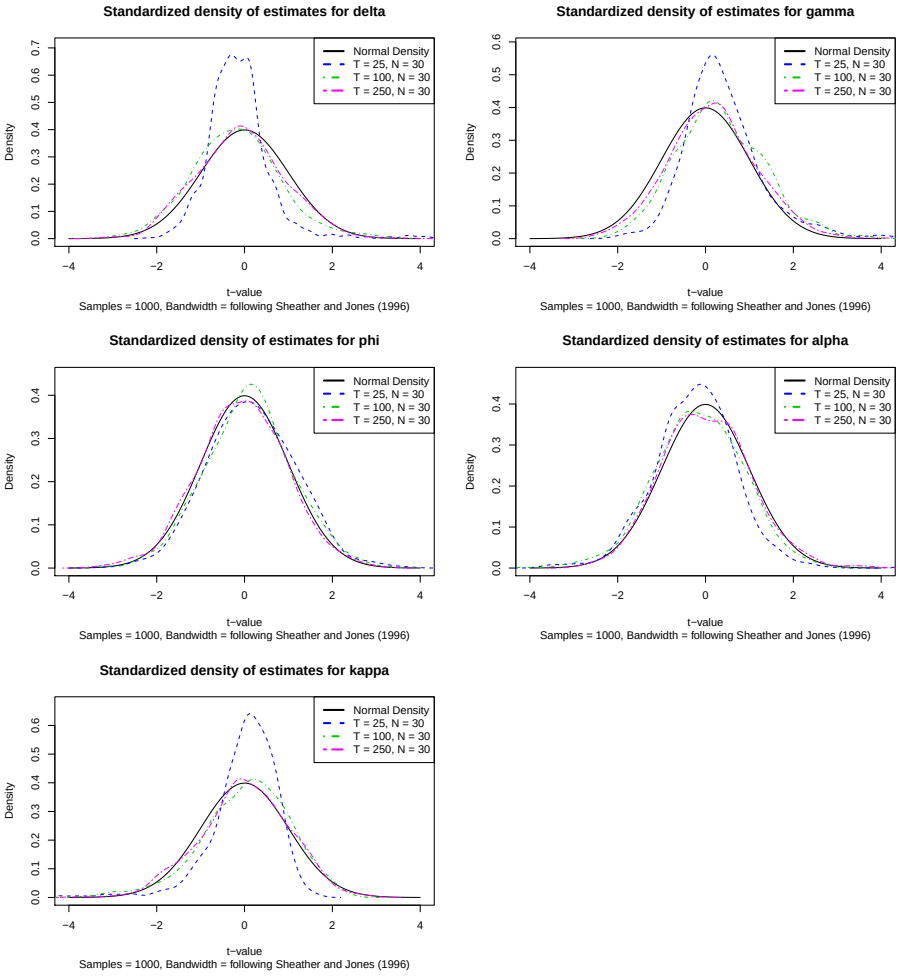

Figure 4.7: Kernel density estimates of estimated parameters from 1000 simulation replications for $N = 30$ indicating that parameters are approximately well-behaved when identified. Note that this does not permit the use of t-statistics to test for significance, evidence for non-linearity can be obtained from AIC and DM-type tests as in our empirical analyses.

Table 4.7: Power: Selection frequencies for contaminated data generated from the ST-SAR.

DGP: ST-SAR + AO		ST-SAR 1 vs. SAR			ST-SAR 2 vs. SAR			ST-SAR 2 vs. ST-SAR 1		
		AIC	AICc	mAIC	AIC	AICc	mAIC	AIC	AICc	mAIC
N=30	T=10	47	44	45	45	41	43	42	40	41
	T=25	59	58	59	61	60	60	48	48	48
	T=50	84	84	84	81	81	81	52	52	52
	T=100	97	97	97	96	96	96	62	61	62
	T=250	100	100	100	100	100	100	95	95	95
N=40	T=10	52	51	51	55	53	54	47	45	46
	T=25	73	72	72	76	76	76	54	54	54
	T=50	93	93	93	91	91	91	54	53	53
	T=100	93	93	93	99	99	99	71	71	71
	T=250	98	98	98	100	100	100	99	99	99
N=50	T=10	57	55	56	63	62	63	54	53	54
	T=25	83	83	83	80	80	80	51	51	51
	T=50	97	97	97	96	96	96	61	61	61
	T=100	99	99	99	100	100	100	77	77	77
	T=250	100	100	100	100	100	100	99	99	99
N=60	T=10	65	64	65	63	62	63	49	48	49
	T=25	89	89	89	88	88	88	49	49	49
	T=50	99	99	99	98	98	98	56	56	56
	T=100	99	99	99	100	100	100	80	80	80
	T=250	100	100	100	100	100	100	100	100	100

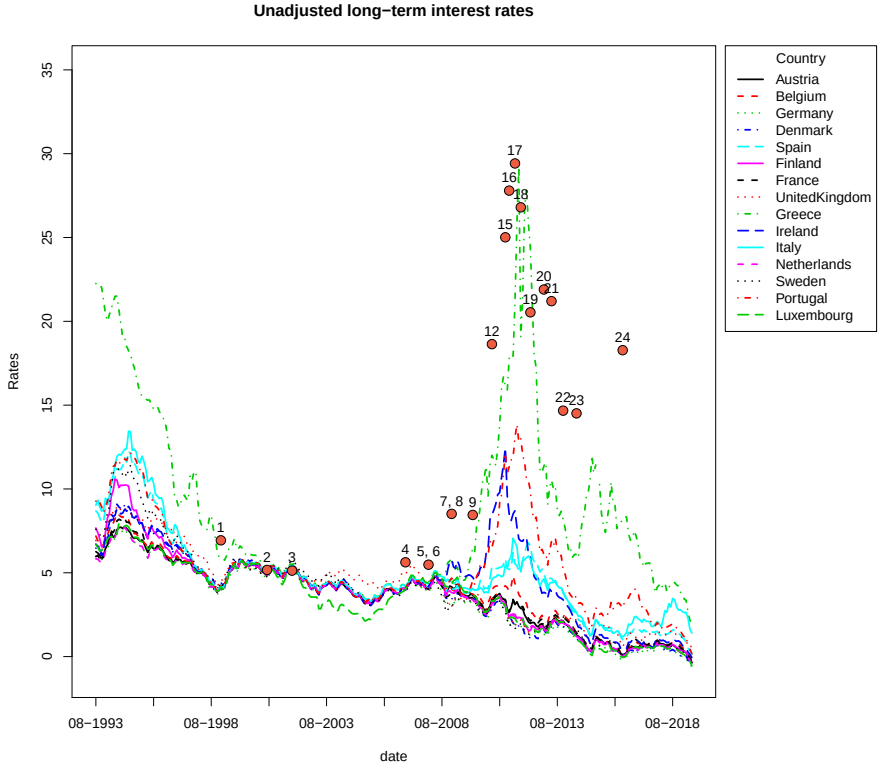

Figure 4.8: Raw data and sovereign colors used in application II.

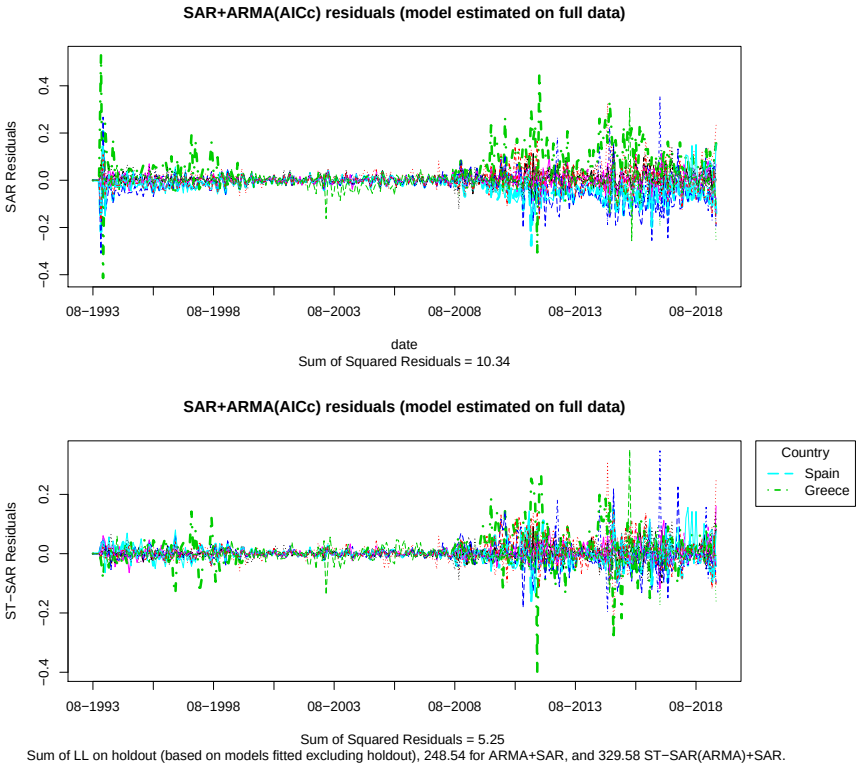

Figure 4.9: Residuals of the final SAR and ST-SAR showing that after filtering out linear spatial dynamics, the residuals of Spain and Greece are not properly centered on zero.

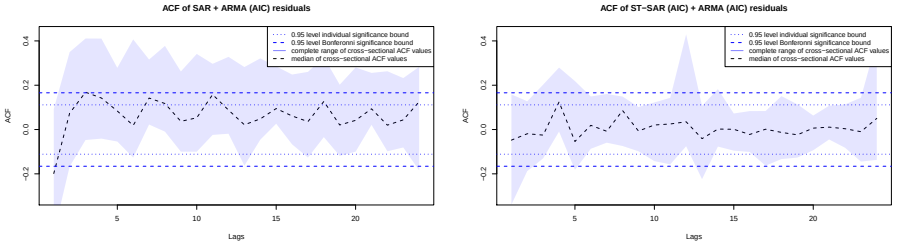

Figure 4.10: Residual correlations of the SAR and ST-SAR estimated on full data, highlighting the improved filtering of the ST-SAR.

4.7.5 Time-line of events related to European Long term Interest Rates

1. January 1999, start of the Euro;
2. January 2001, Greece joins the Euro;
3. January 2002, Euro coins and notes are introduced;
4. January 2007, Slovenia joins the Euro;
5. January 2008, Malta and Cyprus join the Euro;
6. November 26, 2008, 200bn European Economic Recovery Plan;
7. January 2009, Slovakia joins the Euro;
8. January 2009, Estonia, Denmark, Latvia and Lithuania join the ERM;
9. December 17, 2009, Greece hits deficit record;
10. April 19, 2010, Greece hits borrowing cost record;
11. May 2, 2010, Greece accepts 110bn bailout package;
12. November 28, 2010, Ireland accepts 85bn bailout package;
13. January 2011, Estonia joins the Euro;
14. February 14, 2011, agreement of 500bn ESM bailout fund;
15. May 3 , 2011, agreement over 78bn bailout package for Portugal;
16. July 21, 2011, agreement over additional 109bn bailout package for Greece;
17. October 6, 2011, Bank of England injects additional 75bn pounds into the economy;
18. January 2012, major downgrade wave including nine Eurozone nations by S&P;
19. June 2012, Spain and Cyprus request assistance from the ESM;
20. January 23 2013, England threats to leave the European Union;
21. May 2, 2013, ECB cuts the rate on its benchmark refinancing facility to 0.50%;
22. November 7, 2013, ECB cuts the rate on its benchmark refinancing facility to 0.25%;
23. June 2014, first negative interest rates by the ECB;
24. June 23, 2016, Brexit.