Consistency and asymptotic normality of least squares estimators in generalized STAR models

Svetlana Borovkova, Hendrik P. Lopuhaä, and Budi Nurani
VU Amsterdam, TU Delft, and UNPAD Bandung

February 8, 2008

Abstract

Space-time autoregressive (STAR) models, introduced by CLIFF and ORD (1973) are successfully applied in many areas of science, particularly when there is prior information about spatial dependence. These models have significantly fewer parameters than vector autoregressive models, where all information about spatial and time dependence is deduced from the data. A more flexible class of models: generalized STAR models - has been introduced in BOROVKOVA *et al.* (2002), where the model parameters are allowed to vary per location.

This paper establishes strong consistency and asymptotic normality of the least squares estimator in generalized STAR models. These results are obtained under minimal conditions on the sequence of innovations, which are assumed to form a martingale difference array. We investigate the quality of the normal approximation for finite samples by means of a numerical simulation study, and apply a generalized STAR model to a multivariate time series of monthly tea production in West-Java, Indonesia.

Keywords: Space-time autoregressive models, least squares estimator, law of large numbers for dependent sequences, central limit theorem, multivariate time series.

Corresponding address:
H.P. Lopuhaä
Delft Institute of Applied Mathematics
Faculty of EEMCS
Delft University of Technology
Mekelweg 4
2628 CD Delft
Netherlands
Email: H.P.Lopuhaa@ewi.tudelft.nl

Running head: Least squares estimation in STAR models

1 Introduction

In many applications, several time series are recorded simultaneously at a number of locations. Examples are daily oil production at a number of wells, monthly crime rates in city neighborhoods, monthly tea production at several sites, hourly air pollution measurements at various city locations, gene frequencies among sub-populations. These examples give rise to spatial time series, i.e., data indexed by time and location.

The time series from nearby locations are likely to be related, i.e., dependent. The classical way to model such data, is the vector autoregressive (VAR) model (HANNAN (1970)). Although very flexible, the VAR model has too many unknown parameters that must be estimated from only limited amount of data. Moreover, in spatial applications such as geology or ecology, the exact location of sites where measurements are collected and the properties of the surroundings bring extra information about possible interdependencies. In the oil production example, the permeability of the subsurface between two wells is a good indicator of how strongly the production of the two wells are related. In the example of air pollution, the prevailing wind direction and the distances between locations are the main determinants of the measurement dependencies.

1.1 STAR models

Models that explicitly take into account spatial dependence are referred to as *space-time models*. A class of such models – the *space-time autoregressive (STAR)* and *space-time autoregressive moving average (STARMA) models* – were introduced in the 70's by CLIFF and ORD (1973) and MARTIN and OEPPEN (1975), and further studied in CLIFF and ORD (1975ab, 1981), PFEIFER and DEUTSCH (1980abc), STOFFER (1986). As VAR models, STAR models are characterized by linear dependencies in both space and time. The essential difference with VAR models is that, in STAR models, spatial dependencies are imposed by a model builder by means of a *weight matrix*. Spatial features such as distances between locations, neighboring sites or external characteristics such as subsurface permeability or prevailing wind directions are incorporated in such a weight matrix.

Let $\{\mathbf{Z}(t) : t = 0, \pm 1, \pm 2, \ldots\}$ be a multivariate time series of N components. STAR models require the definition of a hierarchical ordering of "neighbors" of each site. For instance, on a regular grid, one can define neighbors of a site "first order neighbors", neighbors of the first order neighbors - "second order neighbors" and so on. The next observation at each site is then modelled as a linear function of the previous observations at the same site and of the weighted previous observations at the neighboring sites of each order. The weights are incorporated in weight matrices $\mathbf{W}^{(k)}$ for spatial order k. Formally, the space-time autoregressive model of autoregressive order p and spatial orders $\lambda_1, \ldots, \lambda_p$ (STAR($p_{\lambda_1,\ldots,\lambda_p}$)) is defined as:

$$\mathbf{Z}(t) = \sum_{s=1}^{p} \sum_{k=0}^{\lambda_s} \phi_{sk} \mathbf{W}^{(k)} \mathbf{Z}(t-s) + \mathbf{e}(t) \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

where λ_s is the spatial order of the sth autoregressive term, ϕ_{sk} is the autoregressive parameter at time lag s and spatial lag k, $\mathbf{W}^{(k)}$ is the $N \times N$-weight matrix for spatial lag $k = 0, 1, \ldots, \lambda_s$, and $\mathbf{e}(t)$ are random error disturbances with mean zero. Because there are no neighbors at spatial lag zero, $\mathbf{W}^{(0)} = \mathbf{I}_N$. For spatial order one or higher we only have first order or higher order neighbors, so that for $k \geq 1$ the weight matrix should satisfy $w_{ii}^{(k)} = 0$. Finally, weights are usually normalized to sum up to one for each location: $\sum_{j=1}^{N} w_{ij}^{(k)} = 1$ for all i, k. A STARMA model is an extension of the STAR model, which includes the moving average of the innovations. PFEIFER and DEUTSCH (1980abc) give a thorough analysis of STARMA models, and describe a three-stage modelling procedure in the spirit of BOX and JENKINS (1976). They describe the parameter regions where STAR models of lower order represent a stationary process and discuss the estimation of the STAR model parameters by conditional maximum likelihood.

1.2 Applications and choice of weights

STAR models have been successfully applied in many areas of science. Geographical research initially inspired the development of these models by CLIFF and ORD (1975ab). Later, EPPERSON (1993) applied STAR models to gene frequencies among populations, thus modelling genetic variation over space and time. PFEIFER and DEUTSCH (1980a) and SARTORIS (2005) applied STAR models to crime rates in Boston and Sao Paulo areas; in the last decade applications of STARMA models in criminology have become increasingly popular (LIU and BROWN (1998)). STAR models are also well-known in economics (GIACOMINI and GRANGER (2004)) and have been applied to forecasting of regional employment data (HERNANDEZ-MURILLO and OWYANG (2004)) and traffic flow (GARRIDO (2000), KAMARIANAKIS and PRASTACOS (2005)). Finally, STAR models are well-known and widely applied in geology and ecology (see e.g., KYRIAKIDIS and JOURNEL (1999) and references therein).

The definition of weights is left to the model builder, An example for sites in a regularly spaced system are *uniform* weights $w_{ij}^{(k)} = 1/n_{ki}$, if the sites i and j are kth order neighbors and zero otherwise, where n_{ki} is the number of kth order neighbors of the site i (BESAG, 1974). In this example, weight matrices are defined simultaneously for all spatial lags; this is the only weighting scheme where this is quite trivial. Another weighting scheme, usually assigned to *areas* (such as in the crime rate example), are *binary* weights (BENNET, 1979). In such a scheme, the weights of the spatial lag one matrix are set to 0 or 1 depending on whether areas have a common boundary or not. Higher spatial lag matrices are then derived from the first lag matrix according to the graph-theoretical approach (see e.g., ANSELIN and SMIRNOV (1996)). The most widely used weights are based on the inverse of the Euclidean distance between locations, such as $c(d_{ij} + 1)^{-\alpha}$ or $c \exp(-\alpha d_{ij})$, where d_{ij} is the distance between sites i and j and c, α are some positive constants (see, e.g., CLIFF and ORD, 1981). In this case, again, defining weight matrices of higher spatial lag is non-trivial and can be rather arbitrary (just as defining neighbors of various orders). Because of these difficulties, models of spatial lag one are often used in practice and these usually provide reasonable description of the data. For example, one can consider all the sites as the first order neighbors and choose the weights according to distances between the sites. This approach is especially useful if the sites are not on a regular grid.

1.3 Generalized STAR models

The main restriction on the STAR model defined in (1) is that the autoregressive parameters ϕ_{sk} are assumed to be the same for all locations. However, there is no apriori justification for this assumption; in fact these parameters are most likely to be different for different locations. A GSTAR (generalized STAR) model is a natural generalization of STAR models, allowing the autoregressive parameters to vary per location: $\phi_{sk}^{(i)}$, $i = 1, 2, \ldots, N$. Note that STAR models are a subclass of GSTAR models. In this paper we shall consider GSTAR models; naturally all the results will continue to hold for STAR models. A GSTAR($p_{\lambda_1,\lambda_2,\ldots,\lambda_p}$) model of autoregressive order p and spatial order λ_s of the sth autoregressive term, is given by

$$\mathbf{Z}(t) = \sum_{s=1}^{p} \left[\mathbf{\Phi}_{s0}\mathbf{Z}(t-s) + \sum_{k=1}^{\lambda_s} \mathbf{\Phi}_{sk}\mathbf{W}^{(k)}\mathbf{Z}(t-s) \right] + \mathbf{e}(t) \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{2}$$

where $\mathbf{\Phi}_{sk} = \mathrm{diag}(\phi_{sk}^{(1)}, \ldots, \phi_{sk}^{(N)})$ and $\mathbf{W}^{(k)}$ satisfies the same conditions as for the model (1). GSTAR models have been considered by BOROVKOVA *et al.* (2002) and form a subclass of the GSTARMA models examined by DI GIACINTO (2006). The term GSTAR has also been adopted by TERZI (1995) in a different context, who considers STAR models with contemporaneous spatial correlation, but still imposes the same parameter for each site.

1.4 Estimation and asymptotics

A GSTAR model can be represented as a linear model and its autoregressive parameters can be estimated by the method of least squares. The conditional maximum likelihood estimators of PFEIFER and DEUTSCH (1980a) are in fact least squares estimators, but the authors point out that one should be cautious in using their approximate confidence regions since these are based on the usual standard linear regression assumptions that do not hold in a time series setup. However, next to nothing is proved so far about the asymptotic properties of the least squares estimator for STAR and GSTAR models, and the conditions for the consistency and asymptotic normality of this estimator are nowhere specified. An aim of this paper is to fill this substantial gap, by establishing consistency and asymptotic normality of the least squares estimator in the model (2) (and hence, also (1)).

Linear regression model with random covariates

The existing results on least squares estimators in related models do not apply to the present setup. The model (2) can be seen as a multiple linear regression model with random covariates:

$$y_t = \mathbf{x}_t\beta + \epsilon_t \qquad t = 1, 2, \ldots, n, \tag{3}$$

where β is a $k \times 1$ vector of parameters, $\mathbf{x}_t = (x_{t1}, \ldots, x_{tk})$ is a $1 \times k$ vector of explanatory variables, and $\epsilon_1, \epsilon_2 \ldots, \epsilon_n$ are random variables. The behavior of the least squares estimator in such models, in particular the behavior of

$$\sum_{t=1}^{n} x_{ti} x_{tj} \quad \text{and} \quad \sum_{t=1}^{n} x_{ti} \epsilon_t \qquad \text{for } i, j = 1, 2, \ldots, n, \tag{4}$$

has been of interest to many authors, e.g., see CRISTOPEIT and HELMES (1980) and LAI and WEI (1982), who also provide further references to the statistical and engineering literature; see also WHITE (1984) for references to the econometrics literature. Strong consistency and asymptotic normality of the least squares estimator is established by these authors under relatively mild conditions. However, their results do not directly apply to our situation. LAI and WEI (1982), consider the general situation where the sequence $\epsilon_1, \epsilon_2 \ldots, \epsilon_n$ forms a martingale difference array. This is unrealistic in our situation, where this sequence is formed by all $e_i(t)$'s, for $i = 1, 2, \ldots, N$ and $t = 0, 1, 2, \ldots$, for which at a fixed time t, the elements $e_1(t), \ldots, e_N(t)$ of the error vector $\mathbf{e}(t)$ still can be correlated. We will assume a martingale difference array structure for the *vectors* $\mathbf{e}(t)$, but this does not imply a similar structure for the entire sequence of *individual components* $e_i(t)$. WHITE (1984) treats linear regression models with matrix-valued $\mathbf{x}_t$ and vector-valued ϵ_t, but requires either mixing, stationary ergodicity, or asymptotically independence of the sequences $\{\mathbf{x}_t\}$ and $\{\epsilon_t\}$. This is stronger than necessary in our specific setup, where we only need moment conditions on the errors vectors.

Vector autoregressive models

Alternatively, one may attempt to apply existing results for vector autoregressive models, since the GSTAR model (2) can be written as a VAR model. For instance, the GSTAR(1_1) model can be written in the form $\mathbf{Z}(t) = \mathbf{A}\mathbf{Z}(t-1) + \mathbf{e}(t)$. Consistency and asymptotic normality of the least squares estimator for $\mathbf{A}$ in such models has been investigated by ANDERSON and TAYLOR (1979), DUFLO *et al.* (1991), and ANDERSON and KUNITOMO (1992). Here, again, the available results are not directly applicable. This is because the specific structure of the matrix $\mathbf{A} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{W}$ implies that we are dealing with a *restricted* matrix-valued least squares estimator. The restricted least squares estimator is related to the unrestricted one via a linear mapping (see e.g., RAO (1968)), but it is somewhat tedious to determine this mapping, and it is even more tedious to determine the limiting covariance matrix of the vector $(\widehat{\phi}_{01}, \widehat{\phi}_{11}, \widehat{\phi}_{02}, \widehat{\phi}_{12}, \ldots, \widehat{\phi}_{0N}, \widehat{\phi}_{1N})'$ from the limit covariance of the restricted LS estimator.

Approach in this paper

We use the specific structure of the matrix $\mathbf{A} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{W}$ directly and formulate the GSTAR model as a linear model in such a way as to avoid the 'restricted LS reasoning', see Section 2. In this way we obtain the main equations (10) and (11), from which the asymptotic properties of the estimator will be derived. We establish strong consistency and asymptotic normality of the least squares estimator in GSTAR models (2). These results are obtained under *minimal* conditions on the sequence of innovations $\mathbf{e}(t)$, which are assumed to form a martingale difference array. An important special case to which our results apply, is the situation where one has independent homoscedastic innovations with mean zero and finite covariance. We derive the covariance matrix of the estimator and show how it can be estimated from the data. This enables constructing confidence regions for the model parameters. For clarity of the exposition, we first concentrate on GSTAR models of autoregressive and spatial order one. The linear model form of GSTAR(1_1) models is presented in Section 2. Consistency and asymptotic normality are established for GSTAR(1_1) models and extended to higher order models in Section 3. The same results for ordinary STAR models are obtained as a corollary. The details of the proofs are in the appendix. In Section 4 we describe a numerical simulation study, performed to investigate how well the behavior at finite samples matches the limiting theory, i.e., the rate of convergence and the quality of the normal approximation. This approximation is remarkably close already for small samples, e.g., for time series of length 50. Finally, the GSTAR model is fitted to a 24-variate time series of monthly tea production in West-Java, Indonesia. For this dataset, the model parameters are significantly different for different locations, which gives a justification for using GSTAR rather than STAR models.

2 LS estimators in GSTAR models of order one

Consider the GSTAR(1_1) model defined through (2), where we write $\phi_{ki} = \phi_{1k}^{(i)}$ for $k = 0, 1$, i.e.,

$$Z_i(t) = \phi_{0i} Z_i(t-1) + \phi_{1i} \sum_{j=1}^{N} w_{ij} Z_j(t-1) + e_i(t). \tag{5}$$

Suppose we have observations $Z_i(t)$, $t = 0, 1, \ldots, T$, for locations $i = 1, 2, \ldots, N$. Then, with $V_i(t) = \sum_{j=1}^{N} w_{ij} Z_j(t)$, the model equations for the ith location can be written as $\mathbf{Y}_i = \mathbf{X}_i \beta_i + \mathbf{u}_i$, where $\beta_i = (\phi_{0i}, \phi_{1i})'$,

$$\mathbf{Y}_i = \begin{bmatrix} Z_i(1) \\ Z_i(2) \\ \vdots \\ Z_i(T) \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} Z_i(0) & V_i(0) \\ Z_i(1) & V_i(1) \\ \vdots & \vdots \\ Z_i(T-1) & V_i(T-1) \end{bmatrix}, \quad \mathbf{u}_i = \begin{bmatrix} e_i(1) \\ e_i(2) \\ \vdots \\ e_i(T) \end{bmatrix}. \tag{6}$$

Consequently, the model equations for all locations simultaneously follow the linear model structure $\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$, with $\mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_N')'$, $\mathbf{X} = \text{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_N)$, $\beta = (\beta_1', \ldots, \beta_N')'$, and $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_N')'$. Note that for each site $i = 1, 2, \ldots, N$, we have a separate linear model $\mathbf{Y}_i = \mathbf{X}_i \beta_i + \mathbf{u}_i$, which means that for each site the least squares estimator for β_i can be computed separately. However, the value of the estimator *does depend* on the values of $\mathbf{Z}(t)$ at other sites, since $V_i(t) = \sum_{j \neq i} w_{ij} Z_j(t)$.

For later theoretical purposes we bring in some additional structure to separate the deterministic weights w_{ij} from the random variables $Z_i(t)$. If, for each $i = 1, 2, \ldots, N$, we define

$$\mathbf{M}_i = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ w_{i1} & \cdots & w_{i,i-1} & 0 & w_{i,i+1} & \cdots & w_{iN} \end{bmatrix}, \tag{7}$$

then $\mathbf{X}_i$ can be written as

$$\mathbf{X}_i' = \mathbf{M}_i \left[\begin{array}{cccc} \mathbf{Z}(0) & \mathbf{Z}(1) & \cdots & \mathbf{Z}(T-1) \end{array} \right], \tag{8}$$

and likewise

$$\mathbf{X}' = \mathbf{M} \left(\mathbf{I} \otimes \left[\begin{array}{cccc} \mathbf{Z}(0) & \mathbf{Z}(1) & \cdots & \mathbf{Z}(T-1) \end{array} \right] \right), \tag{9}$$

where $\mathbf{M} = \mathrm{diag}(\mathbf{M}_1, \ldots, \mathbf{M}_N)$. Here $\mathbf{A} \otimes \mathbf{B}$ denotes the block matrix with blocks $a_{ij}\mathbf{B}$. We conclude that the least squares estimator $\widehat{\beta}_T$ for $\beta = (\phi_{01}, \phi_{11}, \phi_{02}, \phi_{12}, \ldots, \phi_{0N}, \phi_{1N})'$ satisfies the usual normal equations $\mathbf{X}'\mathbf{X}\widehat{\beta}_T = \mathbf{X}'\mathbf{u}$, with $\mathbf{X}$ and $\mathbf{u}$ described above, and can be determined uniquely as soon as the matrix $\mathbf{X}'\mathbf{X}$ is nonsingular. One can easily check that from (9) it follows that

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \mathbf{M} \left(\mathbf{I} \otimes \sum_{t=1}^{T} \mathbf{Z}(t-1)\mathbf{Z}(t-1)' \right) \mathbf{M}', & (10) \\ \mathbf{X}'\mathbf{u} &= \mathbf{M} \left(\sum_{t=1}^{T} \mathrm{vec}\left(\mathbf{Z}(t-1)\mathbf{e}(t)'\right) \right), & (11) \end{aligned}$$

where the operator $\mathrm{vec}(\cdot)$ stacks the columns of a matrix. This shows that the limiting behavior of $\widehat{\beta}_T$ is completely determined by the behavior of $\sum_{t=1}^{T} \mathbf{Z}(t-1)\mathbf{Z}(t-1)'$ and $\sum_{t=1}^{T} \mathbf{Z}(t-1)\mathbf{e}(t)'$.

3 Consistency and asymptotic normality

We will assume that the observations are centered, i.e., $E[\mathbf{Z}(t)] = \mathbf{0}$ and that

(C1) the characteristic roots of the matrix $\mathbf{\Phi}_0 + \mathbf{\Phi}_1\mathbf{W}$ are less than one in absolute value.

Assumption (C1) assures that the GSTAR(1_1) model

$$\mathbf{Z}(t) = \left(\mathbf{\Phi}_0 + \mathbf{\Phi}_1\mathbf{W}\right)\mathbf{Z}(t-1) + \mathbf{e}(t) \qquad t = 0, \pm1, \pm2, \ldots \tag{12}$$

has a causal stationary solution. In applications, the random disturbances $\mathbf{e}(t)$ are often not normally distributed, are not homoscedastic, and may not be independent over observations. We establish the statistical properties of the least squares estimator under the assumption that the sequence $\{\mathbf{e}(t)\}$ forms a vector valued martingale difference array with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$, i.e.,

(C2) $\mathbf{e}(t)$ is $\mathcal{F}_t$-measurable and $E\left[\mathbf{e}(t) \mid \mathcal{F}_{t-1}\right] = \mathbf{0}$.

A similar setup is considered by LAI and WEI(1982) and WHITE(1984) in linear regression models, and by ANDERSON and TAYLOR(1979), DUFLO *et al.* (1991), and ANDERSON and KUNITOMO (1992) in VAR models.

3.1 Weak and strong consistency in GSTAR(1_1) models

The behavior of $\widehat{\beta}_T$ can be obtained from the identity $\mathbf{X}'\mathbf{X}(\widehat{\beta}_T - \beta) = \mathbf{X}'\mathbf{u}$ together with the limiting behavior of $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{u}$ as T goes to infinity. For the latter we essentially have to apply the law of large numbers to the sums appearing in (10) and (11), which requires moment conditions on $\mathbf{Z}(t)$ and $\mathbf{Z}(t-1)\mathbf{e}(t)'$. Because condition (C1) implies causality of the solution of (12), i.e., $\mathbf{Z}(t)$ is a linear combination of the current and past disturbances $\mathbf{e}(s)$, $s \leq t$, it suffices to assume moment conditions on the $\mathbf{e}(t)$ only. We assume

- (C3) $E\left[\mathbf{e}(t)\mathbf{e}(t)' \mid \mathcal{F}_{t-1}\right] = \mathbf{\Sigma}_t$, a.s., where $T^{-1}\sum_{t=1}^{T}\mathbf{\Sigma}_t$ converges in probability to a constant matrix $\mathbf{\Sigma}$.
- (C4) $\sup_{t\geq 1} E\left[\mathbf{e}(t)'\mathbf{e}(t)\left\{\mathbf{e}(t)'\mathbf{e}(t) > a\right\} \mid \mathcal{F}_{t-1}\right]$ converges to zero in probability, as a goes to infinity.

Finally, define

$$\mathbf{\Gamma} = \sum_{j=0}^{\infty} \mathbf{A}^j \mathbf{\Sigma} (\mathbf{A}')^j \quad \text{where } \mathbf{A} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{W}. \tag{13}$$

In the case that (C3) holds with $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ constant, it is easily seen that the matrix $\mathbf{\Gamma}$ is the covariance matrix of $\mathbf{Z}(t)$. The following theorem establishes weak consistency of the least squares estimator.

Theorem 3.1 *Let $\beta = (\phi_{01}, \phi_{11}, \phi_{02}, \phi_{12}, \ldots, \phi_{0N}, \phi_{1N})'$ be the vector of parameters in the GSTAR(1_1) model and let $\mathbf{\Gamma}$ be defined in (13). If $\mathbf{\Gamma}$ is nonsingular, then under conditions (C1)-(C4), the least squares estimator $\widehat{\beta}_T$ converges to β in probability.*

To establish strong consistency we need conditions slightly stronger than (C3) and (C4):

- (C3*) $E\left[\mathbf{e}(t)\mathbf{e}(t)' \mid \mathcal{F}_{t-1}\right] = \mathbf{\Sigma}_t$, a.s., where $T^{-1}\sum_{t=1}^{T}\mathbf{\Sigma}_t$ converges to a constant matrix $\mathbf{\Sigma}$ with probability one.
- (C4*) $\displaystyle\sum_{t=1}^{\infty} t^{-m} E\left[(\mathbf{e}(t)'\mathbf{e}(t))^m \mid \mathcal{F}_{t-1}\right] < \infty$, a.s., for some $1 \leq m \leq 2$.

This enables us to prove the following result (note that this result is similar to Lemma 2 in ANDERSON and TAYLOR(1979), but is proven under weaker conditions).

Lemma 3.1 *Under conditions (C1),(C2),(C3*), and (C4*), $T^{-1}\sum_{t=1}^{T}\mathbf{Z}(t)\mathbf{Z}(t)' \to \mathbf{\Gamma}$ and $T^{-1}\sum_{t=1}^{T}\mathbf{Z}(t-1)\mathbf{e}(t)' \to \mathbf{0}$ with probability one as T goes to infinity, where $\mathbf{\Gamma}$ is defined in (13).*

As a consequence of Lemma 3.1, it follows immediately from (10) and (11) that $\mathbf{X}'\mathbf{X}/T$ converges to $\mathbf{M}(\mathbf{I}\otimes\mathbf{\Gamma})\mathbf{M}'$ and $\mathbf{X}'\mathbf{u}/T$ to zero with probability one, as T goes to infinity. Hence, if $\mathbf{\Gamma}$ is nonsingular, then also $\mathbf{X}'\mathbf{X}/T$ is nonsingular eventually. Together with the identity $\mathbf{X}'\mathbf{X}(\widehat{\beta}_T - \beta) = \mathbf{X}'\mathbf{u}$, similar to the proof of Theorem 3.1, this establishes the following theorem.

Theorem 3.2 *Let* $\beta = (\phi_{01}, \phi_{11}, \phi_{02}, \phi_{12}, \ldots, \phi_{0N}, \phi_{1N})'$ *be the vector of parameters in the GSTAR*(1_1) *model and let* $\mathbf{\Gamma}$ *be defined in (13). If* $\mathbf{\Gamma}$ *is nonsingular, then under conditions (C1),(C2),(C3*), and (C4*), the least squares estimator* $\widehat{\beta}_T$ *converges to* β *with probability one.*

Note that an important example, to which the above conditions apply, is the situation where one has independent homoscedastic error disturbances with mean zero and finite nonsingular covariance matrix.

3.2 Asymptotic normality in GSTAR(1_1) models

Asymptotic normality can be established under conditions (C1)-(C4) and

(C5) For all $r, s = 1, 2, \ldots, N$,

$$\frac{1}{T} \sum_{t=\max\{r,s\}+2}^{T} \mathbf{\Sigma}_t \otimes \mathbf{e}(t-1-r)\mathbf{e}(t-1-s)'$$

converges to $\mathbf{\Sigma} \otimes \mathbf{\Sigma}$ in probability if $r = s$, and to zero otherwise.

Note that if condition (C3) holds with $\mathbf{\Sigma}_t = \mathbf{\Sigma}$, then (C5) is trivially satisfied; see, e.g., Lemma 5.1 in Section 5. The following theorem establishes asymptotic normality of $\widehat{\beta}_T$ and provides an explicit formula for its asymptotic covariance matrix.

Theorem 3.3 *Let* $\beta = (\phi_{01}, \phi_{11}, \phi_{02}, \phi_{12}, \ldots, \phi_{0N}, \phi_{1N})'$ *be the vector of parameters in the GSTAR*(1_1) *model, let* $\widehat{\beta}_T$ *be the corresponding least squares estimator, and let* $\mathbf{\Gamma}$ *be defined in (13). If* $\mathbf{\Gamma}$ *is nonsingular, then under conditions (C1)-(C5),* $\sqrt{T}(\widehat{\beta}_T - \beta)$ *converges in distribution to a* $2N$*-variate normal distribution with mean zero and covariance matrix*

$$\left(\mathbf{M}(\mathbf{I} \otimes \mathbf{\Gamma})\mathbf{M}'\right)^{-1} \mathbf{M}\left(\mathbf{\Sigma} \otimes \mathbf{\Gamma}\right)\mathbf{M}' \left(\mathbf{M}(\mathbf{I} \otimes \mathbf{\Gamma})\mathbf{M}'\right)^{-1} \tag{14}$$

where $\mathbf{M} = \mathrm{diag}(\mathbf{M}_1, \ldots, \mathbf{M}_N)$*, with* $\mathbf{M}_i$ *defined in (7).*

From this theorem the variance of any linear combination $\mathbf{c}'\widehat{\beta}_T$ can be obtained, which enables one to construct confidence intervals for the corresponding parameter $\mathbf{c}'\beta$, such as individual parameters ϕ_{0i}, ϕ_{1i}, or differences $\phi_{0i} - \phi_{0j}$. To determine standard errors, for practical purposes it is useful to say a bit more about the structure of the limiting covariance matrix (14). Since $\mathbf{M}$ is a diagonal $2N \times N^2$ block matrix with N blocks $\mathbf{M}_i$ of size $2 \times N$ on the diagonal, one can deduce that $\mathbf{M}(\mathbf{I} \otimes \mathbf{\Gamma})\mathbf{M}' = \mathrm{diag}(\mathbf{D}_{11}, \ldots, \mathbf{D}_{NN})$ and $\mathbf{M}\left(\mathbf{\Sigma} \otimes \mathbf{\Gamma}\right)\mathbf{M}'$ is a full $2N \times 2N$ block matrix with the (i,j)th block equal to the 2×2 matrix $\sigma_{ij}\mathbf{D}_{ij}$, where

$$\mathbf{D}_{ij} = \mathbf{M}_i \mathbf{\Gamma} \mathbf{M}_j' = \begin{bmatrix} \gamma_{ij} & [\mathbf{\Gamma}\mathbf{W}']_{ij} \\ [\mathbf{W}\mathbf{\Gamma}]_{ij} & [\mathbf{W}\mathbf{\Gamma}\mathbf{W}']_{ij} \end{bmatrix} \tag{15}$$

for $i, j = 1, 2 \ldots, N$. This means that (14) is a full $2N \times 2N$ block matrix with (i,j)th block equal to $\sigma_{ij}\mathbf{D}_{ii}^{-1}\mathbf{D}_{ij}\mathbf{D}_{jj}^{-1}$. Note that the latter matrix represents the limiting covariance between the least squares estimators for the parameters corresponding to sites i and j.

To estimate (14) we simply replace $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$ by estimates $\widehat{\mathbf{\Gamma}}_T$ and $\widehat{\mathbf{\Sigma}}_T$, respectively. One possibility is

$$\widehat{\mathbf{\Gamma}}_T = \frac{1}{T+1}\sum_{t=0}^{T}\mathbf{Z}(t)\mathbf{Z}(t)' \tag{16}$$

$$\widehat{\mathbf{\Sigma}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}(t)\mathbf{R}(t)', \tag{17}$$

where $\mathbf{R}(t) = \mathbf{Z}(t) - (\widehat{\mathbf{\Psi}}_{01} + \widehat{\mathbf{\Psi}}_{11}\mathbf{W})\mathbf{Z}(t-1)$ are the residual vectors. When dividing by $T-2$ in (17), the diagonal element $\widehat{\sigma}_{ii}$ of $\widehat{\mathbf{\Sigma}}_T$ is equal to usual estimator for the error variance in the ith linear model, i.e., the residual sum of squares divided by $T-2$. Furthermore, up to $\mathbf{Z}(T)\mathbf{Z}(T)'$ and a factor $T/(T+1)$, the matrix $\widehat{\mathbf{D}}_{ii} = \mathbf{M}_i\widehat{\mathbf{\Gamma}}_T\mathbf{M}_i'$ is the same as $\mathbf{X}_i'\mathbf{X}_i$. This means that the standard errors of $\widehat{\phi}_{0i}$ and $\widehat{\phi}_{1i}$, obtained from Theorem 3.3 as the square root of the diagonal elements of $\widehat{\sigma}_{ii}\widehat{\mathbf{D}}_{ii}^{-1}$, are almost the same as the ordinary standard errors, that are obtained from fitting the ith linear model $\mathbf{Y}_i = \mathbf{X}_i\beta_i + \mathbf{u}_i$ and which are provided by most statistical packages. However, Theorem 3.3 is necessary to obtain the covariances between parameter estimates, and hence, simultaneous confidence regions. The following theorem states that the estimators $\widehat{\mathbf{\Gamma}}_T$ and $\widehat{\mathbf{\Sigma}}_T$ are consistent for $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$.

Theorem 3.4 *Let $\widehat{\mathbf{\Gamma}}_T$ and $\widehat{\mathbf{\Sigma}}_T$ be defined by (16) and (17). Then under the conditions of Theorem 3.1, $\widehat{\mathbf{\Gamma}}_T$ and $\widehat{\mathbf{\Sigma}}_T$ converge in probability to $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$, respectively. Under the conditions of Theorem 3.2, the convergence is with probability one.*

The ordinary STAR(1_1) model

$$\mathbf{Z}(t) = \phi_0\mathbf{Z}(t-1) + \phi_1\mathbf{W}\mathbf{Z}(t-1) + \mathbf{e}(t) \qquad t = 0, \pm 1, \pm 2, \ldots \tag{18}$$

is a special case of (12). Consistency and asymptotic normality of the least squares estimators in this model are therefore corollaries of Theorems 3.1, 3.2, and 3.3.

Theorem 3.5 *Let $(\widehat{\phi}_{T0}, \widehat{\phi}_{T1})$ be the least squares estimator for the parameters in STAR(1_1) model (18). Let $\mathbf{\Gamma}$ be defined in (13) with $\mathbf{A} = \phi_0\mathbf{I} + \phi_1\mathbf{W}$ and suppose that $\mathbf{\Gamma}$ is nonsingular.*

1. *Under conditions (C1)-(C4), $(\widehat{\phi}_{T0}, \widehat{\phi}_{T1})$ converges to (ϕ_0, ϕ_1) in probability. Moreover, if we replace (C3)-(C4) by (C3*)-(C4*), then the convergence is with probability one.*

2. *Under conditions (C1)-(C5), $\sqrt{T}(\widehat{\phi}_{T0} - \phi_0, \widehat{\phi}_{T1} - \phi_1)'$ converges in distribution to a bivariate normal distribution with mean zero and covariance matrix*

$$\left(\sum_{k=1}^{N}\mathbf{D}_{kk}\right)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}\mathbf{D}_{ij}\left(\sum_{k=1}^{N}\mathbf{D}_{kk}\right)^{-1}$$

where $\mathbf{D}_{ij}$ is defined in (15).

3.3 GSTAR models of higher order

Consider the GSTAR($p_{\lambda_1,\lambda_2,\ldots,\lambda_p}$) model from (2),

$$Z_i(t) = \sum_{s=1}^{p}\sum_{k=0}^{\lambda_s} \phi_{sk}^{(i)} \left(w_{i1}^{(k)} Z_1(t-s) + w_{i2}^{(k)} Z_2(t-s) + \cdots + w_{iN}^{(k)} Z_N(t-s) \right) + e_i(t), \quad (19)$$

for $t = p, p+1, \ldots, T$, where $w_{ij}^{(0)} = 1$ for $i = j$ and zero otherwise. This model can also be formulated in terms of a linear model. Write $V_i^{(k)}(t) = \sum_{j=1}^{N} w_{ij}^{(k)} Z_j(t)$ for $k \geq 1$ and for convenience write $V_i^{(0)}(t) = Z_i(t)$. Then the model equations for the ith location can be written as $\mathbf{Y}_i = \mathbf{X}_i\beta_i + \mathbf{u}_i$, with $\mathbf{Y}_i = (Z_i(p), \ldots, Z_i(T))'$, $\mathbf{u}_i = (e_i(p), \ldots, e_i(T))'$,

$$\mathbf{X}_i = \begin{pmatrix} V_i^{(0)}(p-1) & \cdots & V_i^{(\lambda_1)}(p-1) & \cdots & V_i^{(0)}(0) & \cdots & V_i^{(\lambda_p)}(0) \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ V_i^{(0)}(T-1) & \cdots & V_i^{(\lambda_1)}(T-1) & \cdots & V_i^{(0)}(T-p) & \cdots & V_i^{(\lambda_p)}(T-p) \end{pmatrix},$$

and $\beta_i = (\phi_{10}^{(i)}, \ldots, \phi_{1\lambda_1}^{(i)}, \phi_{20}^{(i)}, \ldots, \phi_{2\lambda_2}^{(i)}, \ldots, \phi_{p0}^{(i)}, \ldots, \phi_{p\lambda_p}^{(i)})'$. As before we have a separate linear model for each location, and the model equations for all locations simultaneously follow a linear model structure $\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$, with $\mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_N')'$, $\mathbf{X} = \mathrm{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_N)$, $\beta = (\beta_1', \ldots, \beta_N')'$, and $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_N')'$.

The previous results can be extended to models of higher order without much effort. The reason is that the GSTAR($p_{\lambda_1,\lambda_2,\ldots,\lambda_p}$) model (2) can also be turned into a autoregressive model of order one. Let $\mathbf{Z}^{(p)}(t) = (\mathbf{Z}(t)', \mathbf{Z}(t-1)', \ldots, \mathbf{Z}(t-p+1)')'$, $\mathbf{e}^{(p)}(t) = (\mathbf{e}(t)', \mathbf{0}', \ldots, \mathbf{0}')'$, and

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_{p-1} & \mathbf{B}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \qquad \text{where } \mathbf{B}_s = \mathbf{\Phi}_{s0} + \sum_{k=1}^{\lambda_s} \mathbf{W}^{(k)}. \quad (20)$$

Then the GSTAR($p_{\lambda_1,\ldots,\lambda_p}$) model (2) can be written as a VAR(1) model:

$$\mathbf{Z}^{(p)}(t) = \mathbf{B}\mathbf{Z}^{(p)}(t-1) + \mathbf{e}^{(p)}(t). \quad (21)$$

Similar to (13), let $\mathbf{\Gamma}^{(p)}$ be defined by

$$\mathbf{\Gamma}^{(p)} = \sum_{j=0}^{\infty} \mathbf{B}^j \mathbf{\Sigma}^{(p)} (\mathbf{B}')^j \qquad \text{with } \mathbf{\Sigma}^{(p)} = \mathrm{diag}(\mathbf{\Sigma}, \mathbf{0}, \ldots, \mathbf{0}), \quad (22)$$

where $\mathbf{B}$ is defined in (20) and $\mathbf{\Sigma}$ is the constant matrix from condition (C3) or (C3*). As before, in the case that (C3) holds for $\mathbf{e}(t)$ with $\mathbf{\Sigma}_t = \mathbf{\Sigma}$ constant, which implies that

(C3) holds for $\mathbf{e}^{(p)}(t)$ with $\mathbf{\Sigma}_t = \mathbf{\Sigma}^{(p)}$ constant, it is easily seen that the matrix $\mathbf{\Gamma}^{(p)}$ is the covariance matrix of $\mathbf{Z}^{(p)}(t)$:

$$\mathbf{\Gamma}^{(p)} = \begin{bmatrix} \mathbf{\Gamma}(0) & \mathbf{\Gamma}(-1) & \cdots & \mathbf{\Gamma}(-p+1) \\ \mathbf{\Gamma}(1) & \mathbf{\Gamma}(0) & \cdots & \mathbf{\Gamma}(-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Gamma}(p-1) & \mathbf{\Gamma}(p-2) & \cdots & \mathbf{\Gamma}(0) \end{bmatrix}, \tag{23}$$

where $\mathbf{\Gamma}(s) = E[\mathbf{Z}(t)\mathbf{Z}(t+s)']$. First, we impose condition (C1) on the matrix $\mathbf{B}$:

(C1*) all roots of $|\lambda^p \mathbf{I} - \lambda^{p-1}\mathbf{B}_1 - \cdots - \lambda \mathbf{B}_{p-1} - \mathbf{B}_p| = 0$ satisfy $|\lambda| < 1$.

Note that this condition is equivalent to saying that all eigenvalues of $\mathbf{B}$ are less than one in absolute value. We then have the following theorem.

Theorem 3.6 *Let $\beta = (\beta_1', \beta_2', \ldots, \beta_N')'$, with $\beta_i = \left(\phi_{10}^{(i)}, \ldots, \phi_{1\lambda_1}^{(i)}, \ldots, \phi_{p0}^{(i)}, \ldots, \phi_{p\lambda_p}^{(i)}\right)'$ be the vector of parameters in the GSTAR($p_{\lambda_1,\ldots,\lambda_p}$) model and let $\widehat{\beta}_T$ be the corresponding least squares estimator. Let $\mathbf{\Gamma}^{(p)}$ be defined in (22) and suppose that $\mathbf{\Sigma}$ is nonsingular. Then*

1. *under conditions (C1*), (C2)-(C4), $\widehat{\beta}_T$ converges to β in probability. Moreover, if we replace (C3)-(C4) by (C3*)-(C4*), then the convergence is with probability one.*

2. *under conditions (C1*), (C2)-(C5), $\sqrt{T}(\widehat{\beta}_T - \beta)$ converges in distribution to a d-variate normal distribution, where $d = (p + \lambda_1 + \cdots + \lambda_p)N$, with zero mean and covariance matrix*

$$\left(\mathbf{N}\left(\mathbf{I} \otimes \mathbf{\Gamma}^{(p)}\right)\mathbf{N}'\right)^{-1} \mathbf{N}\left(\mathbf{\Sigma} \otimes \mathbf{\Gamma}^{(p)}\right)\mathbf{N}'\left(\mathbf{N}\left(\mathbf{I} \otimes \mathbf{\Gamma}^{(p)}\right)\mathbf{N}'\right)^{-1} \tag{24}$$

where $\mathbf{N} = diag(\mathbf{N}_1, \ldots, \mathbf{N}_N)$ with $\mathbf{N}_i$ defined by (31) in Section 5.

The parameters $\mathbf{\Gamma}^{(p)}$ and $\mathbf{\Sigma}$ that appear in (24) can be estimated as before:

$$\widehat{\mathbf{\Sigma}}_T = \frac{1}{T-p+1} \sum_{t=p}^{T} \left(\mathbf{Z}(t) - \widehat{\mathbf{Z}}(t)\right)\left(\mathbf{Z}(t) - \widehat{\mathbf{Z}}(t)\right)',$$

where $\widehat{\mathbf{Z}}(t) = \sum_{s=1}^{p} [\widehat{\mathbf{\Phi}}_{s0}\mathbf{Z}(t-s) + \sum_{k=1}^{\lambda_s} \widehat{\mathbf{\Phi}}_{sk}\mathbf{W}^{(k)}\mathbf{Z}(t-s)]$. The matrix $\mathbf{\Gamma}^{(p)}$ can be estimated by replacing $\mathbf{\Gamma}(s)$ in (23) by

$$\widehat{\mathbf{\Gamma}}_T(s) = \frac{1}{T-s+1} \sum_{t=0}^{T-s} \mathbf{Z}(t)\mathbf{Z}(t+s)'$$

and $\widehat{\mathbf{\Gamma}}_T(-s) = \widehat{\mathbf{\Gamma}}_T(s)'$, for $s \geq 0$. Similar to Theorem 3.4 one can show that this provides consistent estimators for $\mathbf{\Gamma}^{(p)}$ and $\mathbf{\Sigma}$.

The ordinary STAR($p_{\lambda_1,\lambda_2,\ldots,\lambda_p}$) model is a special case of (19) with $\phi_{sk}^{(i)} = \phi_{sk}$. Hence, consistency and asymptotic normality of the least squares estimator is a corollary of Theorem 3.6.

Theorem 3.7 *Let $\phi = \left(\phi_{10}, \ldots, \phi_{1\lambda_1}, \phi_{20}, \ldots, \phi_{2\lambda_2}, \ldots, \phi_{p0}, \ldots, \phi_{p\lambda_p}\right)'$ be the vector of parameters in the ordinary STAR$(p_{\lambda_1,\ldots,\lambda_p})$ model and let $\widehat{\phi}_T$ be the corresponding least squares estimator. Let $\mathbf{\Gamma}^{(p)}$ be defined in (22) and suppose that $\mathbf{\Sigma}$ is nonsingular. Then*

1. *under conditions (C1*), (C2)-(C4), $\widehat{\phi}_T$ converges to ϕ in probability. Moreover, if we replace (C3)-(C4) by (C3*)-(C4*), then the convergence is with probability one.*

2. *under conditions (C1*), (C2)-(C5), $\sqrt{T}(\widehat{\phi}_T - \phi)$ converges in distribution to a d-variate normal distribution, where $d = p+\lambda_1+\cdots+\lambda_p$, with zero mean and covariance matrix*

$$\left(\sum_{k=1}^{N} \mathbf{N}_k \mathbf{\Gamma}^{(p)} \mathbf{N}_k'\right)^{-1} \sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{ij} \mathbf{N}_i \mathbf{\Gamma}^{(p)} \mathbf{N}_j' \left(\sum_{k=1}^{N} \mathbf{N}_k \mathbf{\Gamma}^{(p)} \mathbf{N}_k'\right)^{-1}$$

where $\mathbf{\Gamma}^{(p)}$ is defined in (22) and $\mathbf{N}_i$ defined by (31) in Section 5.

4 Numerical study and data analysis

4.1 Simulation study

First, we simulated a three-variate GSTAR(1_1) time series corresponding to model (12), with independent $N(\mathbf{0}, \mathbf{\Sigma})$ distributed innovations. We used $N = 3$ locations and varying length $T = 50, 100, 500, 1000$ and $10\,000$, autoregressive parameters $\mathbf{\Phi}_0 = \text{diag}(0.3, 0.1, 0.1)$, and $\mathbf{\Phi}_1 = \text{diag}(0.4, 0.3, 0.3)$, and

$$\mathbf{W} = \begin{pmatrix} 0 & 0.4 & 0.6 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0.8 & 0 \end{pmatrix} \qquad \mathbf{\Sigma} = \begin{pmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{pmatrix}.$$

For this model, the maximal absolute eigenvalue of the matrix $\mathbf{\Phi}_0 + \mathbf{\Phi}_1\mathbf{W}$ is 0.49. Hence, all absolute eigenvalues are within the unit circle, which guarantees a causal stationary solution of (12). We also performed another simulation, of a GSTAR model with the parameters on the border of the stationarity region: $\mathbf{\Phi}_0 = \text{diag}(0.99, 0.1, 0.1)$, and $\mathbf{\Phi}_1 = \text{diag}(0.1, 0.03, 0.03)$ ($\mathbf{W}$ and $\mathbf{\Sigma}$ are the same as in previous GSTAR model). For these parameter values, the maximum absolute eigenvalue of $\mathbf{\Phi}_0 + \mathbf{\Phi}_1\mathbf{W}$ is 0.99.

For both models, we estimated the vector of parameters $\beta = (\phi_{01}, \phi_{11}, \phi_{02}, \phi_{12}, \phi_{03}, \phi_{13})'$ by the method of least squares. Table 1 shows the theoretical and estimated values of the parameters, for increasing number of observations T. It can be seen that, for both models, $\widehat{\beta}_T$ approaches β as T increases. To compare the rate of convergence, the empirical mean squared errors of $\widehat{\beta}_T$, calculated as the average of 1000 Monte Carlo repetitions of $\|\widehat{\beta}_T - \beta\|^2$, are also shown. It can be seen that for the model on the border of the stationarity region (Model 2), the convergence of $\widehat{\beta}_T$ to β is not worse.

Next, we investigate how well the limiting normal distribution of $\sqrt{T}(\widehat{\beta}_T - \beta)$, as described in Theorem 3.3, matches the finite sample distribution. For this we perform

Table 1: Parameter values and least squares estimates.

	β	T = 50	100	500	1000	10000
Model 1	$\phi_{01} = 0.3$	0.2789	0.2916	0.2971	0.3000	0.2999
	$\phi_{11} = 0.4$	0.4039	0.3932	0.3991	0.4008	0.3995
	$\phi_{02} = 0.1$	0.0937	0.0976	0.0998	0.1003	0.1003
	$\phi_{12} = 0.3$	0.2994	0.2968	0.2998	0.2994	0.2999
	$\phi_{03} = 0.1$	0.0949	0.1001	0.0970	0.1001	0.1000
	$\phi_{13} = 0.3$	0.2902	0.3006	0.3002	0.2993	0.3000
	MSE	0.1519	0.0748	0.0149	0.0070	0.0007
Model 2	$\phi_{01} = 0.99$	0.9574	0.9738	0.9864	0.9881	0.9898
	$\phi_{11} = 0.1$	0.1014	0.1002	0.1001	0.1014	0.1005
	$\phi_{02} = 0.1$	0.0819	0.0896	0.0965	0.0988	0.1001
	$\phi_{12} = 0.03$	0.0252	0.0209	0.0297	0.0291	0.0299
	$\phi_{03} = 0.1$	0.0844	0.0897	0.0977	0.0984	0.1000
	$\phi_{13} = 0.03$	0.0176	0.0222	0.0294	0.0285	0.0299
	MSE	0.1061	0.0524	0.0089	0.0043	0.0004

a Monte Carlo simulation with 1000 replications of a GSTAR(1_1) model with the above parameters, and for each replication obtain a realization of the random vector $\sqrt{T}(\widehat{\beta}_T - \beta)$. Since we cannot visualize the entire joint distribution of the six-dimensional vector $\sqrt{T}(\widehat{\beta}_T - \beta)$, we look at the marginal distributions and the joint distribution of just two components. Figure 1 shows kernel density estimates (bandwidths are determined by the normal reference method) of 1000 realizations of the first component $\sqrt{T}(\widehat{\phi}_{01} - \phi_{01})$, together with the asymptotic normal density, for increasing values of T (top row), and contour lines of the two-dimensional kernel density estimates for two parameters from different locations $\sqrt{T}(\widehat{\phi}_{01} - \phi_{01})$ and $\sqrt{T}(\widehat{\phi}_{02} - \phi_{02})$, together with the bivariate normal contour lines (bottom row). The normal approximation is very good even for $T = 50$. The above simulations demonstrate that the consistency and asymptotic normality of the least squares estimator in GSTAR models is clearly observed for as little as 50 observations, and these asymptotic properties also hold for parameter values near the border of the stationarity region.

4.2 Monthly tea production in Western Java

Finally, we apply a GSTAR model to a multivariate time series of the monthly tea production in the western part of Java. The series consists of 96 observations of monthly

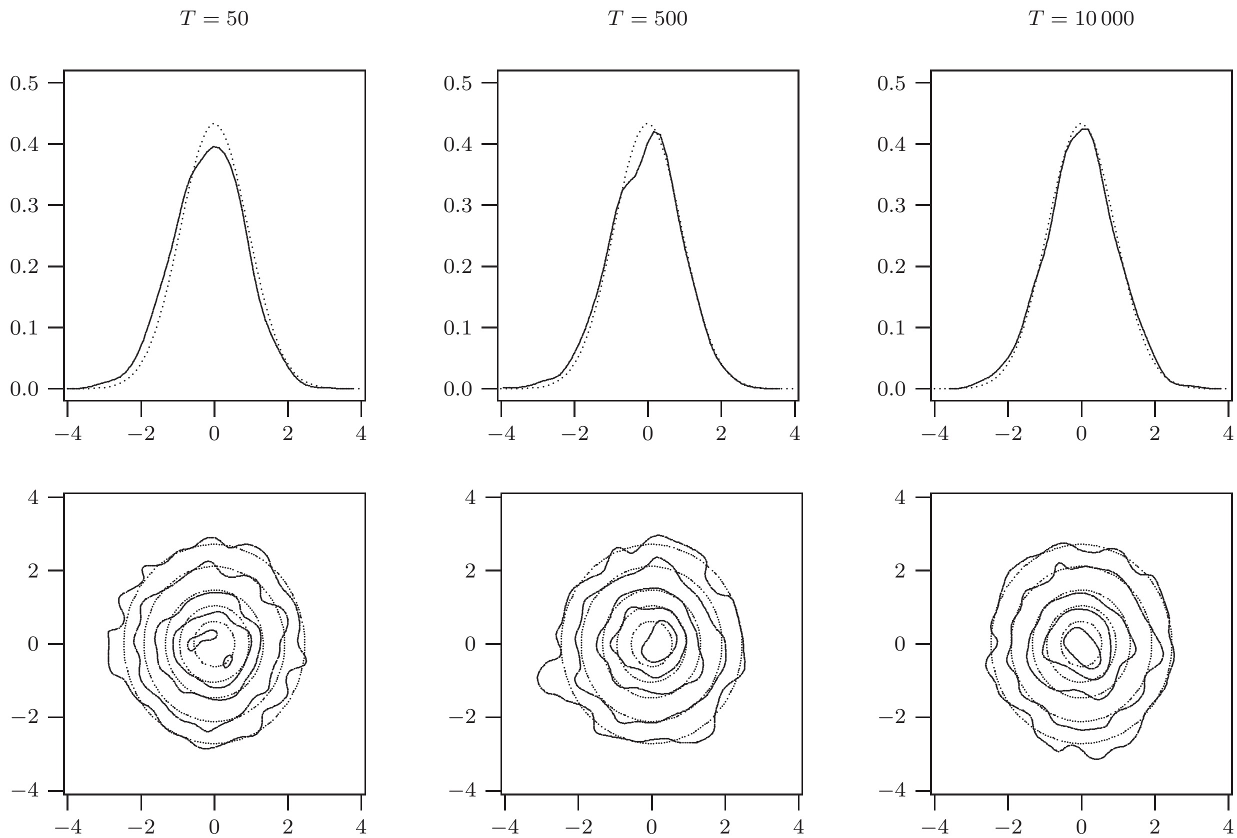

Figure 1: Kernel density estimates and contour lines of two-dimensional kernel density estimates of 1000 realizations versus the normal density and bivariate normal contour lines.

tea production collected at 24 cities from January 1992 to December 1999. Thus, there are $N = 24$ sites, each with $T = 96$ observations. Figure 2 shows the locations of the sites. If this 24-variate time series is modelled by a vector autoregressive model, at least $24^2 = 576$ parameters must be estimated, which is clearly not feasible, given the number of observations. Moreover, we know the exact locations of the sites, and hence the sites inter-distances, which can be used to quantify the dependencies between the sites. There is no apparent trend or seasonality in the series, so we only subtracted the global mean vector from the time series.

The sites are not on a regular grid, hence it is not immediately obvious how to define neighbors of spatial lag higher than one. The temporal order can be chosen by examining the spatial autocorrelation and partial autocorrelation functions, defined for ordinary STAR models in PFEIFER and DEUTSCH (1980a). The partial autocorrelation function should be used here with caution, since its definition applies only to STAR models and not

Figure 2: The sites of the tea production in western part of Java.

directly to GSTAR models. This examination already suggests temporal order 2 or 3; we considered temporal and spatial lags up to order three.

Two types of weight matrices are chosen. The first type is based on inverse distances. Given a vector of cutoff distances $0 = d_0 < d_1 < \cdots < d_\lambda$, the weight matrix of spatial lag k is based on inverse weights $1/(d_{ij}+1)$ for sites i and j whose Euclidean distance d_{ij} is between d_{k-1} and d_k, and weight zero otherwise. In order to meet the condition $\sum_j w_{ij}^{(k)} = 1$ for $k = 1, 2, 3$, we took $d_1 = \infty$ (all sites are first order neighbor), $(d_1, d_2) = (3, \infty)$, and $(d_1, d_2, d_3) = (3, 4.5, \infty)$. The second type is based on binary weights in a similar fashion: for a fixed site i, all sites j with a distance between d_{k-1} and d_k are considered as kth order neighbors of site i, and all the corresponding weights are equal. We took $d_1 = \infty$, $(d_1, d_2) = (1.5, \infty)$ and $(d_1, d_2, d_3) = (1.5, 3, \infty)$. All weight matrices are normalized in such a way that for all locations the weights add up to one. Higher order weight matrices may also be derived from a first lag matrix according to the approach in ANSELIN and SMIRNOV (1996). However, this results in almost the same weight matrices as the binary ones.

We compared the models by the Akaike information criterion $\log(\det(\widehat{\Sigma}_T)) + 2k/n$ for VAR models (see MCQUARRIE and TSAI (1998)), where $k = p+\lambda_1+\cdots+\lambda_p+N(N+1)/2$ is the number of parameters and $n = T + 1 - p$ is the effective series length. Table 2 lists the values of the AIC for all considered models. The column 'cut-off' indicates whether for the largest kth spatial lag all remaining sites are kth order neighbors or only a part of them are. For instance, the 1-all GSTAR(2_{11}) model with inverse distance weights corresponds to the choice $d_1 = \infty$ as cut-off value, whereas the other GSTAR(2_{11}) model corresponds to $(d_1, d_2) = (3, \infty)$.

All GSTAR models provide a significantly better fit than their STAR counterpart, whose AIC values are all above 490. We left out the AIC values for the other GSTAR models of temporal order $p = 3$ or higher, since they do not provide better fits. Overall, the two GSTAR(2_{11}) models have the smallest AIC values, with the 1-all model that has

Table 2: The AIC for GSTAR($p_{\lambda_1,\ldots,\lambda_p}$).

Temporal order	Spatial order	cut-off	Weight matrix: Distances	Weight matrix: Binary
$p = 1$	λ_1			
	1	1-all	488.16	488.12
	1		488.27	488.74
	2	2-all	488.46	488.96
	2		488.56	489.46
	3	3-all	488.94	489.29
$p = 2$	(λ_1, λ_2)			
	$(1, 1)$	1-all	487.48	487.19
	$(1, 1)$		487.91	488.36
	$(1, 2)$	2-all	488.68	488.88
	$(1, 2)$		488.83	488.85
	$(2, 1)$	2-all	488.20	488.53
	$(2, 1)$		488.26	488.86
	$(2, 2)$	2-all	488.71	488.88
	$(2, 2)$		488.80	489.39
	$(1, 3)$	3-all	489.30	489.39
	$(2, 3)$	3-all	489.19	489.80
	$(3, 1)$	3-all	488.67	488.80
	$(3, 2)$	3-all	489.11	489.34
	$(3, 3)$	3-all	489.57	489.65
$p = 3$	$(\lambda_1, \lambda_2, \lambda_3)$			
	$(1, 1, 1)$	1-all	487.98	487.75
	$(1, 1, 1)$		488.48	489.24

all sites as first order neighbors providing the best fit. In general, the k-all models provide better fits than their counterparts with finite cutoff values. The type of weight matrix does not seem to affect the overall picture, although the models with inverse distance weights have the best overall performance. The fit could be possibly improved further, using a weight matrix that incorporates characteristics relevant for tea production, such as average precipitation at sites or moistness of soil.

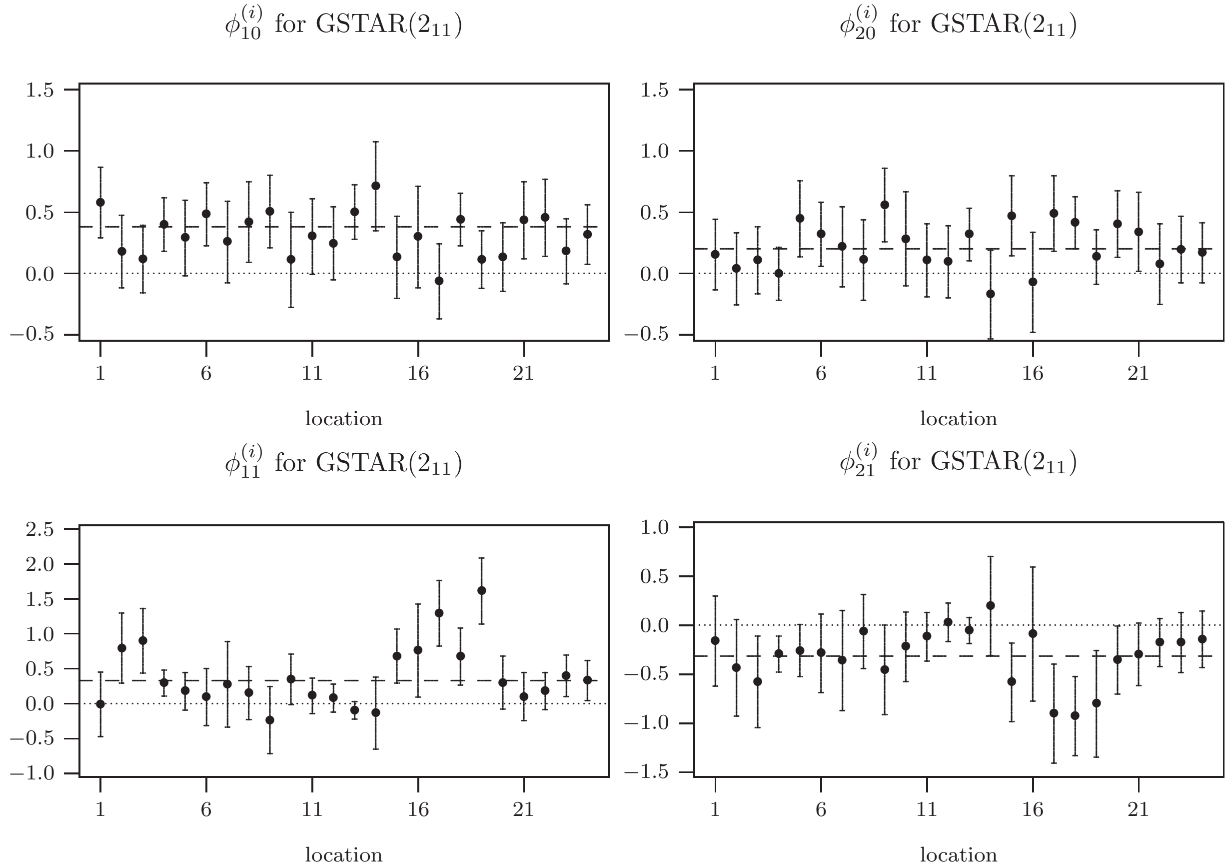

Figure 3: Estimated temporal and spatial parameters with 95% confidence intervals for the GSTAR(2_{11}) model with inverse distance weights and all sites as first order neighbors.

Figure 3 shows the estimates of the temporal and spatial regressive parameters $\phi_{s0}^{(i)}$ and $\phi_{s1}^{(i)}$ for all sites together with their 95% confidence intervals in the 1-all GSTAR(2_{11}) model with inverse distance weights. The dotted line indicates zero and the dashed line the estimated parameters $\widehat{\phi}_{10} = 0.38$, $\widehat{\phi}_{20} = 0.20$, $\widehat{\phi}_{11} = 0.33$, and $\widehat{\phi}_{21} = -0.32$ of the corresponding STAR model (recall that in a STAR model these parameters are the same for all locations). Note that many parameters $\phi_{sk}^{(i)}$ are significantly different from zero. More importantly, there are significant differences in parameter values at different sites. For all pairs of sites we tested the hypotheses $\phi_{sk}^{(i)} = \phi_{sk}^{(j)}$. For many pairs these hypotheses are rejected at 5% level. For example, for the sites 14 and 17 the hypothesis $\phi_{10}^{(i)} = \phi_{10}^{(j)}$ is rejected with p-value 0.0022, and for the sites 9 and 14 the hypothesis $\phi_{20}^{(i)} = \phi_{20}^{(j)}$ is rejected with p-value 0.0027.

In all, the empirical application shows that GSTAR models outperform STAR models due to their greater flexibility, and hence, should be preferred. Furthermore, the appropriate choice of the temporal order is more important for the model fit than the choice of spatial order. Finally, the exact choice of the weight matrix does not seem to influence the results very much.

5 Appendix: proofs

Proof of Theorem 3.1: The theorem follows from a general result in ANDERSON and KUNITOMO (1992) for autoregressive models. From their Lemma 2 on page 229 we conclude that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}(t-1)\mathbf{Z}(t-1)' \to \mathbf{\Gamma} \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}(t-1)\mathbf{e}(t)' \to \mathbf{0}, \tag{25}$$

in probability. Together with (10) and (11) this implies that

$$\frac{1}{T}\mathbf{X}'\mathbf{X} \to \mathbf{M}(\mathbf{I}\otimes\mathbf{\Gamma})\mathbf{M}' = \text{diag}(\mathbf{M}_1\mathbf{\Gamma}\mathbf{M}_1', \ldots, \mathbf{M}_N\mathbf{\Gamma}\mathbf{M}_N) \tag{26}$$

and $\mathbf{X}'\mathbf{u}/T \to \mathbf{0}$ in probability. Since $\mathbf{\Gamma}$ is nonsingular, we conclude that $\mathbf{M}(\mathbf{I}\otimes\mathbf{\Gamma})\mathbf{M}'$ is nonsingular, which means that $\mathbf{X}'\mathbf{X}/T$ is nonsingular with probability tending to one. The theorem now follows from the identity $\mathbf{X}'\mathbf{X}(\widehat{\beta}_T - \beta) = \mathbf{X}'\mathbf{u}$. ∎

The proof of Lemma 3.1 requires the following result.

Lemma 5.1 *Under conditions (C2),(C3*), and (C4*),*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{e}(t)\mathbf{e}(t+\Delta)' \to \begin{cases} \mathbf{\Sigma} & \text{if } \Delta = 0, \\ 0 & \text{if } \Delta = 1, 2, \ldots, \end{cases}$$

with probability one.

Proof of Lemma 5.1: For the case $\Delta = 0$ consider $M_T = \sum_{t=1}^{T} X_t$, with $X_t = e_r(t)e_s(t) - \sigma_{t,rs}$, for $r, s = 1, 2, \ldots, N$ fixed and $\sigma_{t,rs}$ being the rsth element of $\mathbf{\Sigma}_t$. Then by conditions (C2) and (C3*), M_T is a martingale with respect to $\{\mathcal{F}_t\}$. Since $m \geq 1$, it follows from Jensen's inequality:

$$E\left[|X_t|^m \mid \mathcal{F}_{t-1}\right] \leq 2^{m-1}\left(E\left[(\mathbf{e}(t)'\mathbf{e}(t))^m \mid \mathcal{F}_{t-1}\right] + \sigma_{t,rs}^m\right) \leq 2^m E\left[(\mathbf{e}(t)'\mathbf{e}(t))^m \mid \mathcal{F}_{t-1}\right].$$

Condition (C4*) implies that $\sum_{t=1}^{\infty} t^{-m}E[|X_t|^m \mid \mathcal{F}_{t-1}] < \infty$ with probability one, so that we conclude from Theorem 2.18 in HALL and HEYDE (1980) that $\sum_{t=1}^{T} t^{-1}X_t$ converges with probability one. By Kronecker's lemma (see, e.g., HALL and HEYDE, 1980, Section 2.6) this implies that $T^{-1}\sum_{t=1}^{T}(e_r(t)e_s(t) - \sigma_{t,rs}) = T^{-1}\sum_{t=1}^{T} X_t$ converges to zero with probability one, so that the case $\Delta = 0$ follows from condition (C3*).

For $\Delta = 1, 2, \ldots$ fixed, consider $M_T' = \sum_{t=1}^{T-\Delta} X_t'$, with $X_t' = e_r(t)e_s(t+\Delta)$. This is again a martingale with respect to $\{\mathcal{F}_t\}$. By the Cauchy-Schwarz inequality:

$$\left(\sum_{t=1}^{\infty} E\left[t^{-m}|X_t'|^m \;\middle|\; \mathcal{F}_{t-1}\right]\right)^2 \leq \sum_{t=1}^{\infty} E\left[t^{-m}\|\mathbf{e}(t)\|^{2m}\middle|\, \mathcal{F}_{t-1}\right] \cdot \sum_{t=1}^{\infty} E\left[t^{-m}\|\mathbf{e}(t+\Delta)\|^{2m}\middle|\, \mathcal{F}_{t-1}\right].$$

The first term is finite with probability one by condition (C4*). For the second term we write

$$\sum_{t=1}^{\infty} E\Big[E\Big[t^{-m}\|\mathbf{e}(t+\Delta)\|^{2m}\Big|\,\mathcal{F}_{t+\Delta-1}\Big]\Big|\,\mathcal{F}_{t-1}\Big] = E\Big[\sum_{t=1}^{\infty} E\Big[t^{-m}\|\mathbf{e}(t+\Delta)\|^{2m}\Big|\,\mathcal{F}_{t+\Delta-1}\Big]\Big|\,\mathcal{F}_{t-1}\Big]$$
$$\leq (1+\Delta)^m E\left[\sum_{t=1}^{\infty}(t+\Delta)^{-m} E\Big[\|\mathbf{e}(t+\Delta)\|^{2m}\Big|\,\mathcal{F}_{t+\Delta-1}\Big]\Big|\,\mathcal{F}_{t-1}\right].$$

Again by condition (C4*),

$$\sum_{t=1}^{\infty}(t+\Delta)^{-m} E\Big[\|\mathbf{e}(t+\Delta)\|^{2p}\Big|\,\mathcal{F}_{t+\Delta-1}\Big] \leq \sum_{t=1}^{\infty} t^{-m} E\Big[\|\mathbf{e}(t)\|^{2m}\Big|\,\mathcal{F}_{t-1}\Big] < \infty \quad \text{a.s.},$$

so we conclude that $\sum_{t=1}^{\infty} t^{-m} E\big[|X_t'|^m \mid \mathcal{F}_{t-1}\big] < \infty$ with probability one. By a similar reasoning as above it follows that $T^{-1}\sum_{t=1}^{T-\Delta} e_r(t)e_s(t+\Delta) \to 0$ a.s.. ■

Proof of Lemma 3.1: First consider $\mathbf{Y}(t) = \mathbf{Z}(t) - \mathbf{A}^t\mathbf{Z}(0) = \sum_{j=0}^{t-1}\mathbf{A}^j\mathbf{e}(t-j)$. Then for each $i, j = 1, 2, \ldots, N$ fixed:

$$\frac{1}{T}\sum_{t=1}^{T} Y_i(t)Y_j(t) = \sum_{r=1}^{N}\sum_{s=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{t-1}[\mathbf{A}^\tau]_{ir}\, e_r(t-\tau)\right)\left(\sum_{\theta=0}^{t-1}\left[\mathbf{A}^\theta\right]_{js} e_s(t-\theta)\right) \qquad (27)$$

For $i, j, r, s = 1, 2, \ldots, N$ fixed, we apply Lemma 2 in CRISTOPEIT and HELMES (1980) with $a_\tau = [\mathbf{A}^\tau]_{ir}$, $b_\theta = \left[\mathbf{A}^\theta\right]_{js}$, $\xi_t = e_r(t,\omega)$, $\zeta_t = e_s(t,\omega)$, and $h = k = 0$. We have that $\sum_{\tau=0}^{\infty}|a_\tau| \leq \sum_{\tau=0}^{\infty}\|\mathbf{A}^\tau\| \leq \sqrt{N}\sum_{\tau=0}^{\infty}\|\mathbf{A}^\tau\|_2$, where $\|\mathbf{A}\|_2 = \sup\|\mathbf{A}\mathbf{x}\|/\|\mathbf{x}\|$ with $\|\cdot\|$ the Euclidean norm. According to Lemma 5 in (Anderson and Kunitomo(1992)), $\|\mathbf{A}^\tau\|_2 \leq q\tau^N\lambda^\tau$, where λ denotes the largest absolute eigenvalue of $\mathbf{A}$ and $q > 0$ is a suitable constant independent of τ. From Lemma 5.1 it follows that for all ω in a set of probability one, $T^{-1}\sum_{t=1}^{T}\xi_t^2 \to \sigma_{rr}$, $T^{-1}\sum_{t=1}^{T}\zeta_t^2 \to \sigma_{ss}$, and $T^{-1}\sum_{t=1}^{T-\Delta}\xi_t\zeta_{t-\Delta}$ and $T^{-1}\sum_{t=1}^{T-\Delta}\xi_{t+\Delta}\zeta_t$ converge to σ_{rs} for $\Delta = 0$, and to zero otherwise. Lemma 2 in CRISTOPEIT and HELMES (1980) now implies that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{t-1}[\mathbf{A}^\tau]_{ir}\, e_r(t-\tau)\right)\left(\sum_{\theta=0}^{t-1}\left[\mathbf{A}^\theta\right]_{js} e_s(t-\theta)\right) \to \sum_{\tau=0}^{\infty}[\mathbf{A}^\tau]_{ir}\,[\mathbf{A}^\tau]_{js}\,\sigma_{rs},$$

with probability one. Together with (27) we find that

$$\frac{1}{T}\sum_{t=1}^{T} Y_i(t)Y_j(t) \to \sum_{\tau=0}^{\infty}\sum_{r=1}^{N}\sum_{s=1}^{N}[\mathbf{A}^\tau]_{ir}\,[\mathbf{A}^\tau]_{js}\,\sigma_{rs} = \sum_{\tau=0}^{\infty}\left[\mathbf{A}^\tau\mathbf{\Sigma}(\mathbf{A}^\tau)'\right]_{ij},$$

with probability one. Now, write

$$\mathbf{Y}(t)\mathbf{Y}(t)' = \mathbf{Z}(t)\mathbf{Z}(t)' - \mathbf{A}^t\mathbf{Z}(0)\mathbf{Z}(t)' - \mathbf{Z}(t)\mathbf{Z}(0)'(\mathbf{A}^t)' + \mathbf{A}^t\mathbf{Z}(0)\mathbf{Z}(0)'(\mathbf{A}^t)'.$$

Then, to prove the lemma we are left with showing that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\mathbf{Z}(t)' \to \mathbf{0} \tag{28}$$

with probability one, and

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\mathbf{Z}(0)'(\mathbf{A}^t)' \to \mathbf{0} \tag{29}$$

with probability one. As before, we have from condition (C1):

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\mathbf{Z}(0)'(\mathbf{A}^t)'\right\| \le \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{Z}(0)\|^2\|\mathbf{A}^t\|_2^2 \le \|\mathbf{Z}(0)\|^2\cdot\frac{1}{T}\sum_{t=1}^{T}(qt^N\lambda^t)^2 \to 0,$$

with probability one, where λ denotes the largest absolute eigenvalue of $\mathbf{A}$. To prove (28), write

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\mathbf{Z}(t)' = \frac{1}{T}\sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\mathbf{Z}(0)'(\mathbf{A}^t)' + \frac{1}{T}\sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\sum_{\theta=0}^{t-1}(\mathbf{A}^\theta\mathbf{e}(t-\theta))'.$$

According to (29) the first term tends to zero. Next consider the ijth element of the second term for $i,j=1,2,\dots,N$ fixed:

$$\sum_{r=1}^{N}\sum_{s=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left((\mathbf{A}^t)_{ir}Z_r(0)\right)\left(\sum_{\theta=0}^{t-1}(\mathbf{A}^\theta)_{js}e_s(t-\theta)\right).$$

This can be treated again by applying Lemma 2 in CRISTOPEIT and HELMES (1980) to $a_\tau = 1_{\{\tau=0\}}$, $b_\theta = (\mathbf{A}^\theta)_{js}$, $\xi_t = (\mathbf{A}^t)_{ir}Z_r(0,\omega)$, and $\zeta_t = e_s(t,\omega)$. Clearly, $\sum_{\tau=0}^{\infty}|a_\tau| < \infty$ and as before $\sum_{\theta=0}^{\infty}|b_\theta| < \infty$. From (29) and Lemma 5.1 it follows that for all ω in a set of probability one, $T^{-1}\sum_{t=1}^{T}\xi_t^2 \to 0$ and $T^{-1}\sum_{t=1}^{T}\zeta_t^2 \to \sigma_{ss}$. Condition (C1) implies that the process $\mathbf{Z}(t)$ is causal, which means that $\mathbf{Z}(0)$ only depends on $\mathbf{e}(s)$, $s\le 0$. This implies that for any $\Delta \ge 1$, the process $M_T = \sum_{t=1}^{T-\Delta}\xi_t\zeta_{t-\Delta}$ is a martingale. As in the proof of Lemma 5.1 we conclude that for all ω in a set of probability one, $T^{-1}\sum_{t=1}^{T-\Delta}\xi_t\zeta_{t-\Delta} \to 0$, and similarly $T^{-1}\sum_{t=1}^{T-\Delta}\xi_{t+\Delta}\zeta_t \to 0$. Lemma 2 in CRISTOPEIT and HELMES (1980) now implies that

$$\frac{1}{T}\sum_{t=1}^{T}(\mathbf{A}^t)_{ir}Z_r(0)\sum_{\theta=0}^{t-1}(\mathbf{A}^\theta)_{js}e_s(t-\theta) \to 0 \quad \text{a.s..}$$

This proves (28), and hence $T^{-1}\sum_{t=1}^{T}\mathbf{Z}(t)\mathbf{Z}(t)' \to \mathbf{\Gamma}$ with probability one. To prove the second statement, first note that $M_T' = \sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\mathbf{e}(t)'$ is a martingale. As in the proof of Lemma 5.1 it follows that $T^{-1}\sum_{t=1}^{T}\mathbf{A}^t\mathbf{Z}(0)\mathbf{e}(t)' \to \mathbf{0}$ with probability one. Hence, it suffices to consider for $i,j=1,2,\dots,N$ fixed:

$$\frac{1}{T}\sum_{t=1}^{T}Y_i(t-1)e_j(t) = \sum_{r=1}^{N}\sum_{s=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{t-2}[\mathbf{A}^\tau]_{ir}\,e_r(t-1-\tau)\right)e_j(t).$$

Again apply Lemma 2 in CRISTOPEIT and HELMES (1980) with $a_\tau = [\mathbf{A}^\tau]_{ir}$, $b_\theta = 1_{\{\theta=0\}}$, $\xi_t = e_r(t,\omega)$, $\zeta_t = e_j(t,\omega)$, $h = 1$, and $k = 0$. As before it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{t-2}[\mathbf{A}^\tau]_{ir}\, e_r(t-1-\tau)\right) e_j(t) \to 0$$

with probability one, which proves the lemma. ■

Proof of Theorem 3.3: From Theorem 4 in ANDERSON and KUNITOMO (1992) it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\text{vec}\left(\mathbf{Z}(t-1)\mathbf{e}(t)'\right)$$

converges to a multivariate random vector with mean zero and covariance matrix $\boldsymbol{\Sigma}\otimes\boldsymbol{\Gamma}$. From the identity $\mathbf{X}'\mathbf{X}(\widehat{\beta}_T - \beta) = \mathbf{X}'\mathbf{u}$ together with property (26) for $\mathbf{X}'\mathbf{X}$ and representation (11) for $\mathbf{X}'\mathbf{u}$, it follows that $\sqrt{T}(\widehat{\beta}_T - \beta)$ converges in distribution to a $2N$-variate normal distribution with zero mean and covariance matrix $(\mathbf{M}(\mathbf{I}\otimes\boldsymbol{\Gamma})\mathbf{M}')^{-1}\mathbf{M}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Gamma}\right)\mathbf{M}'((\mathbf{M}(\mathbf{I}\otimes\boldsymbol{\Gamma})\mathbf{M}')^{-1}$. ■

Proof of Theorem 3.4: Under the conditions of Theorem 3.1, the convergence in probability of $\widehat{\boldsymbol{\Gamma}}_T$ has already been established in (25). For $\widehat{\boldsymbol{\Sigma}}_T$ we have the following the identity

$$\begin{aligned}
\frac{T-2}{T}\widehat{\boldsymbol{\Sigma}}_T &= \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{Z}(t)-\widehat{\mathbf{A}}\mathbf{Z}(t-1)\right)\left(\mathbf{Z}(t)-\widehat{\mathbf{A}}\mathbf{Z}(t-1)\right)' \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}(t)\mathbf{e}(t)' + (\mathbf{A}-\widehat{\mathbf{A}})\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}(t-1)\mathbf{Z}(t-1)'(\mathbf{A}-\widehat{\mathbf{A}})' \\
&\quad +(\mathbf{A}-\widehat{\mathbf{A}})\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}(t-1)\mathbf{e}(t)' + \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}(t)\mathbf{Z}(t-1)'(\mathbf{A}-\widehat{\mathbf{A}})'
\end{aligned}$$

where $\mathbf{A} = \boldsymbol{\Psi}_{10} + \boldsymbol{\Psi}_{11}\mathbf{W}$ and $\widehat{\mathbf{A}} = \widehat{\boldsymbol{\Psi}}_{10} + \widehat{\boldsymbol{\Psi}}_{11}\mathbf{W}$ is the *restricted* least squares estimator. According to Theorem 2 in ANDERSON and KUNITOMO (1992), the first term on the right hand side converges to $\boldsymbol{\Sigma}$ in probability. From Theorem 3.1, and (25) it follows that the three other terms on the right hand side converge to zero in probability. Under the conditions of Theorem 3.2, the almost sure convergence of $\widehat{\boldsymbol{\Gamma}}_T$ has already been established in Lemma 3.1. The almost convergence of $\widehat{\boldsymbol{\Sigma}}_T$ follows from the above identity, Lemma 5.1, Theorem 3.2, and Lemma 3.1. ■

Proof of Theorem 3.5: The STAR(1_1) model coincides with model (5) with all $\phi_{0i} = \phi_0$ and all $\phi_{1i} = \phi_1$. This means that in the linear model $\mathbf{Y} = \widetilde{\mathbf{X}}\phi + \mathbf{u}$ for STAR(1_1), $\mathbf{Y}$ and $\mathbf{u}$

are as before, $\phi = (\phi_0, \phi_1)$, and $\widetilde{\mathbf{X}}$ is the $NT \times 2$ matrix obtained by stacking the matrices $\mathbf{X}_i$ as defined in (6):

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} = \mathbf{X} \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix},$$

with $\mathbf{X} = \text{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_N)$ is the matrix in the proof of Theorem 3.1. It follows that the least squares estimator $\widehat{\phi}_T$ satisfies $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\widehat{\phi}_T - \phi) = \widetilde{\mathbf{X}}'\mathbf{u}$, where

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{I} & \cdots & \mathbf{I} \end{bmatrix} \mathbf{X}'\mathbf{X} \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} \quad \text{and} \quad \widetilde{\mathbf{X}}'\mathbf{u} = \begin{bmatrix} \mathbf{I} & \cdots & \mathbf{I} \end{bmatrix} \mathbf{X}'\mathbf{u}.$$

From (26) it follows that under conditions (C1)-(C4),

$$\frac{1}{T}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \to \begin{bmatrix} \mathbf{I} & \cdots & \mathbf{I} \end{bmatrix} \begin{pmatrix} \mathbf{M}_1\boldsymbol{\Gamma}\mathbf{M}_1' & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{M}_N\boldsymbol{\Gamma}\mathbf{M}_N \end{pmatrix} \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} = \sum_{i=1}^{N} \mathbf{D}_{ii}, \tag{30}$$

where $\mathbf{D}_{ii}$ is defined in (15), and since $\mathbf{X}'\mathbf{u}/T \to \mathbf{0}$ in probability (see the proof of Theorem 3.1) also $\widetilde{\mathbf{X}}'\mathbf{u}/T \to \mathbf{0}$ in probability. As in the proof of Theorem 3.2, under conditions (C1), (C2), (C3*), and (C4*), the convergence is with probability one. Because $\boldsymbol{\Gamma}$ is nonsingular, also $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/T$ is nonsingular eventually, so that part (a) follows from that identity $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\widehat{\phi}_T - \phi) = \widetilde{\mathbf{X}}'\mathbf{u}$. For part (b), first note that from the proof of Theorem 3.3 we know that $\mathbf{X}'\mathbf{u}/\sqrt{T}$ is asymptotically normal with mean zero and covariance matrix $\mathbf{M}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Gamma})\mathbf{M}'$. This means that $\widetilde{\mathbf{X}}'\mathbf{u}/\sqrt{T}$ is asymptotically normal with mean zero and covariance matrix

$$\begin{bmatrix} \mathbf{I} & \cdots & \mathbf{I} \end{bmatrix} \begin{pmatrix} \sigma_{11}\mathbf{D}_{11} & \cdots & \sigma_{1N}\mathbf{D}_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1}\mathbf{D}_{N1} & \cdots & \sigma_{NN}\mathbf{D}_{NN} \end{pmatrix} \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} = \sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{ij}\mathbf{D}_{ij}.$$

Together with (30), part (b) follows from the identity $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\widehat{\phi}_T - \phi) = \widetilde{\mathbf{X}}'\mathbf{u}$. ∎

Proof of Theorem 3.6: First consider the linear model corresponding to the GSTAR($p_{\lambda_1,\ldots,\lambda_p}$) model as described in Section 3.3. If we write $\mathbf{N}_i = \text{diag}(\mathbf{N}_{i1}, \ldots, \mathbf{N}_{ip})$, where for $i = 1, 2, \ldots, N$ and $s = 1, 2, \ldots, p$

$$\mathbf{N}_{is} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ w_{i1}^{(1)} & \cdots & w_{i,i-1}^{(1)} & 0 & w_{i,i+1}^{(1)} & \cdots & w_{iN}^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{i1}^{(\lambda_s)} & \cdots & w_{i,i-1}^{(\lambda_s)} & 0 & w_{i,i+1}^{(\lambda_s)} & \cdots & w_{iN}^{(\lambda_s)} \end{pmatrix}, \tag{31}$$

then similar to (8) we can write $\mathbf{X}_i' = \mathbf{N}_i \begin{pmatrix} \mathbf{Z}^{(p)}(p-1) & \mathbf{Z}^{(p)}(p) & \cdots & \mathbf{Z}^{(p)}(T-1) \end{pmatrix}$. This means that, if $\mathbf{N} = \mathrm{diag}(\mathbf{N}_1, \mathbf{N}_2, \ldots, \mathbf{N}_N)$, then similar to (10) and (11),

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \mathbf{N}\left(\mathbf{I} \otimes \sum_{t=p}^{T} \mathbf{Z}^{(p)}(t-1)\mathbf{Z}^{(p)}(t-1)'\right)\mathbf{N}', \\ bX'\mathbf{u} &= \mathbf{N}\left(\sum_{t=p}^{T} \mathrm{vec}\left(\mathbf{Z}^{(p)}(t-1)\mathbf{e}(t)'\right)\right). \end{aligned}$$

Condition (C1*) is equivalent with condition (C1) for $\mathbf{B}$ in model (21). Conditions (C2)-(C4) on $\mathbf{e}(t)$ with $\mathbf{\Sigma}$ imply the same conditions for $\mathbf{e}^{(p)}(t)$ with $\mathbf{\Sigma}^{(p)}$ as defined in (22). We conclude that conditions (C1)-(C4) hold in VAR(1) model (21). Finally, since $\mathbf{\Sigma}$ is nonsingular, it follows from Lemma 3 in ANDERSON and KUNITOMO (1992) that $\mathbf{\Gamma}^{(p)}$ is nonsingular. From here on the proof of $\hat{\beta}_T \to \beta$ in probability is completely similar to that of Theorem 3.1, where instead of (26), we now have

$$\frac{1}{T}\mathbf{X}'\mathbf{X} \to \mathbf{N}(\mathbf{I} \otimes \mathbf{\Gamma}^{(p)})\mathbf{N}' = \mathrm{diag}(\mathbf{N}_1\mathbf{\Gamma}^{(p)}\mathbf{N}_1', \ldots, \mathbf{N}_N\mathbf{\Gamma}^{(p)}\mathbf{N}_N).$$

Conditions (C3*)-(C4*) on $\mathbf{e}(t)$ with $\mathbf{\Sigma}$ imply the same conditions for $\mathbf{e}^{(p)}(t)$ with $\mathbf{\Sigma}^{(p)}$. This means that the proof of $\hat{\beta}_T \to \beta$ with probability one is the same as that of Theorem 3.2. This proves part (i). For part (ii), note that condition (C5) on $\mathbf{e}(t)$ implies the same condition for $\mathbf{e}^{(p)}(t)$. The matrix $\mathbf{N}$ plays the same role for the extended model (21), as the matrix $\mathbf{M}$ plays in the GSTAR(1_1) model. Hence, from here on the proof is the same as that of Theorem 3.3, with

$$\frac{1}{\sqrt{T}}\sum_{t=p}^{T} \mathrm{vec}\left(\mathbf{Z}^{(p)}(t-1)\mathbf{e}(t)'\right)$$

converging to a multivariate random vector with mean zero and covariance matrix $\mathbf{\Sigma} \otimes \mathbf{\Gamma}^{(p)}$.
■

Proof of Theorem 3.7: The proof is completely similar to that of Theorem 3.5, with matrices $\mathbf{N}_i$, $\mathbf{N}$, and $\mathbf{\Gamma}^{(p)}$ instead of $\mathbf{M}_i$, $\mathbf{M}$, and $\mathbf{\Gamma}$. ■

References

ANDERSON, T.W. and N. KUNITOMO (1992). Asymptotic distributions of regression and autoregression coefficients with martingale difference disturbances. *Journal of Multivariate Analysis*, **40**, 221-243.

ANDERSON, T.W. and J.B. TAYLOR (1979). Strong consistency of least squares estimates in dynamic models. *Annals of Statistics*, **7**, 484-489.

ANSELIN, L. and O. SMIRNOV (1996). Efficient algorithms for constructing proper higher order spatial lag operators. *Journal of Regional Science*, **36**, 67-89.

BENNET, R.J. (1979). *Spatial Time series, Analysis-Forecasting-Control*. San Fransisco: Holden-Day, Inc.

BESAG, J.S. (1974). Spatial interaction and the statistical analysis of lattice system. *Journal of the Royal Statistical Society, Series B*, **36**, p.197-242.

BOROVKOVA, S.A., H.P. LOPUHAÄ, and B. NURANI (2002). Generalized STAR models with experimental weights, *Proceedings of the 17th International Workshop on Statistical Modelling* 2002, Chania.

BOX, G.E.P. and G.M. JENKINS (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day Inc.

BROCKWELL, P.J. and R.A. DAVIS (1996). *Introduction to Time Series and Forecasting*. New York: Springer-Verlag, Inc.

CLIFF, A.D. and J. ORD (1973). *Spatial Autocorrelation*. London: Pioneer.

CLIFF, A.D. and J. ORD (1975a). Model building and the analysis of spatial pattern in human geography. *Journal of the Royal Statistical Society, Series B*, **37**, 297-348.

CLIFF, A.D. and J. ORD (1975b). Space-time modelling with an application to regional forecasting. *Transactions of the Institute of British Geographers*, **66**, 119-128.

CLIFF, A.D. and J. ORD (1981). *Spatial Processes, Models and Applications*, New York: Methuen.

CRISTOPEIT, N. and K. HELMES (1980). Strong consistency of the least-squares estimators in linear regression models. *Annals of Statistics*, **8**(4), 778-788.

DI GIACINTO, V. (2006). A generalized space-time ARMA model with an application to regional unemployement analysis in Italy. *International Regional Science Review*, **29**, 159-198.

DUFLO, M., R. SENOUSSI and A. TOUATI (1991). Propriétes asymptotique presque sûre de l'estimateur carrés d'un modèle autorégressif vectoriel. *Annales d'Institute Henri Poincaré*, **27**, 1-25.

EPPERSON, B.K. (1993). Spatial and space-time correlations in systems of subpopulations with genetic drift and migration. *Genetics*, **133**, 711-727.

GARRIDO, R. (2000). Spatial interaction between the truck flows through the Mexico-Texas border. *Transportation Research, Part A*, **34**, 23-33.

GIACOMINI, R. and C.W.J. GRANGER, (2004). Aggregation of space-time processes. *Journal of Econometrics*, **118**, 7-26.

HALL, P. and C.C. HEYDE (1980). *Martingale Limit Theory and Its Applications*. Academic Press.

HANNAN, E.J. (1970). *Multiple Time Series*. New York: John Wiley & Sons.

HERNANDEZ-MURILLO, R.and , M.T. OWYANG (2004). The information content of regional employment data for forecasting aggregate conditions. *The Federal Reserve Bank of St.Lous Working Paper 2004-005B*.

KAMARIANAKIS, Y. and P.P. PRASTACOS (2005). Space-time modelling of traffic flow. *Computers and Geosciences*, **31**, 119-133.

KYRIAKIDIS, P.C.and A.G. JOURNEL (1999). Geostatistical space-time models: a review. *Mathematical Geology*, **31**(6), 651-683.

LAI, T.L. and C.Z. WEI (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamical systems. *Annals of Statistics*, **10**(1), 154-166.

LIU, H. and D.E. BROWN (1998). Spatial-temporal event prediction: A new model. *Proc. IEEE Conf. on Systems, Man and Cybernetics*, October 1998, San Diego.

MARTIN, R. and J. OEPPEN (1975). The identification of regional forecasting models using space-time correlation functions. *Transactions of the Institute of British Geographers*, **66**, 95-118.

MCQUARRIE, A.D.R. and C.L. TSAI (1998). *Regression and Time Series Model Selection*, Word Scientific.

PFEIFER, P.E. and S.J. DEUTSCH (1980a). A Three Stage Iterative Procedure for Space-Time Modeling. *Technometrics*, **22**(1), 35-47.

PFEIFER, P.E. and S.J. DEUTSCH (1980b). Identification and Interpretation of First Order Space-Time ARMA Models. *Technometrics*, **22**, 397-408.

PFEIFER, P.E. and S.J. DEUTSCH (1980c). Stationary and Invertibility Regions for Low Order STARMA Models. *Commun. Stat. & Simulation and Computation*, **9**(5), 551-562.

RAO, C.R. (1968). *Linear Statistical Inference and Its Applications.* Wiley.

SARTORIS, A. (2005). A STARMA model for homicides in the city of Sao Paulo. *Proceedings of the Spatial Economics Workshop*, Kiel Institute for World Economics, April 8-9, 2005.

STOFFER, D.S. (1986). Estimation and Identification of Space-Time ARMAX Models in the Presence of Missing Data. *Journal of the American Statistical Association*, **81** (395), 762-772.

TERZI, S. (1995) Maximum likelihood estimation of a generalized STAR($p; 1_p$) model. *Statistical Methods and Applications*, **3**, 377-393.

WHITE, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press.