Maximum likelihood estimation of spatially and serially correlated panels with random effects

Giovanni Millo[a,1]

[a] *Generali Research and Development, Assicurazioni Generali S.p.A., Via Machiavelli 3 34131 Trieste (Italy)*

Abstract

An estimation framework and a user-friendly software implementation are described for maximum likelihood estimation of panel data models with random effects, a spatially lagged dependent variable and spatially and serially correlated errors. This specification extends static panel data models in the direction of serial error correlation, allowing richer modelling possibilities and more thorough diagnostic assessments. The estimation routines extend the functionalities of the splm package for spatial panel econometrics in the open source R system for statistical computing.

Keywords: Spatial panel, maximum likelihood, serial correlation, R

---

1. Introduction

The econometric literature has recently considered panel regression models with spatially autocorrelated outcomes or disturbances and random or fixed individual effects. After the pioneering works of Anselin (1988) and Case (1991), the more recent methodological contributions by Elhorst (2003) and Baltagi et al. (2003) and the first comprehensive treatments of the subject in Anselin et al. (2008) and Elhorst (2009) have helped the diffusion of spatial panel methods in applied practice, although hindered by the lack of user-friendly software (see Millo and Piras, 2012). Meanwhile, Baltagi et al. (2007b) have extended the spatial panel framework to considering serial correlation in the remainder errors.[2] Lee and Yu (2012), in a recent comprehensive treatment, are the first to analyze a very general specification including spatial lags, spatially and se-

[1] The material in this article originates from the computational part of the author's PhD dissertation at DEAMS, University of Trieste. The views expressed are solely his own and do not necessarily reflect those of his employer. The software described is available in **R** package **splm** since Version 1.1-00.

[2] Notable alternative approaches to space-time dependence are: the dynamic spatial panel framework of Lee and Yu (2010a); and the space-time correlation model of Elhorst (2008), where spatial and serial correlation are modeled jointly.

rially correlated errors together with individual effects.[3] As they observe, "[i]n empirical applications with spatial panel data, it seems that investigators tend to limit their focus on some spatial structures and ignore others, and in addition, no serial correlation is considered" (Lee and Yu, 2012, p. 1370). They also document through Montecarlo simulation the biases due to neglecting serial correlation or some part of the spatial structure, and recommend a general to specific strategy.

This paper describes the implementation in the **R** system for statistical computing (R Development Core Team, 2012) of maximum likelihood estimation of panel models with spatial lags, spatial errors, individual random effects that are/are not spatially autocorrelated and a temporally autocorrelated remainder error term. In economic terms, this specification can accommodate spatial spillovers from the outcome variable at neighbouring locations, spatial diffusion of idiosyncratic shocks, individual heterogeneity and time-persistence of idiosyncratic shocks. In terms of the previous literature, while building on the general estimation theory in Anselin (1988), this framework extends Case (1991) in adding serial error correlation, Baltagi et al. (2007b) in adding a spatially lagged dependent variable. Moreover, unlike Baltagi et al. (2007b) whose primary goal is to derive Lagrange Multiplier diagnostic tests under a restricted specification, we take a general to specific approach aiming primarily at estimation of the full model.[4]

The common specifications (spatial lag and/or spatial error with or without random effects) have already been available for some time in the **R** package **splm** described in Millo and Piras (2012), although implementation details were still unpublished. Here we describe an extended framework allowing for serial correlation of the autoregressive type, along with all the other features. The corresponding software procedures extend the capabilities of the **splm** package in this direction.

One main goal has been to produce a complete set of estimators able to cope with all combinations of the dependence structures considered, while keeping software as easy to read and maintain as possible. To this end, the theoretical estimation framework has been geared towards modularity, so as to have software counterparts to theoretical objects such as spatial filters or covariances, combining them and "plugging them" in the likelihood as prescribed by the reference theory.

In the spirit of the **R** project, we have taken advantage of some peculiar features of the language. In particular abstraction of tasks into functions for easier readability and maintainance; functions as a data type to be passed on as arguments to other functions, in the spirit of correspondence between conceptual and software tools; and lastly, object-orientation, for example, in applying

[3] Their paper appeared as this one was already under revision. We thank an anonymous referee for pointing us at it.

[4] For a review and comparison of the general to specific and specific to general strategies in applied spatial econometrics, see Florax et al. (2003) and Mur and Angulo (2009).

specialized algebraic methods to matrices with a peculiar structure for reliability and performance.

The main contribution of the paper is in combining existing results in the literature – the general estimation framework of Anselin (1988) and analytical derivations in Baltagi et al. (2007b) – into an operational framework for the estimation of a number of models for which, to our knowledge, no algorithms are currently available. Moreover, it describes a user-friendly, open-source implementation in the **R** language which will hopefully open up new modelling possibilities in the field of spatio-temporally correlated panels for a number of applied researchers.

Given the large number of specifications considered, some preliminary words on notation are in order. We consider either cross-sections of N data points (only in Section 4.1) or balanced panels of N data points observed over T time periods. Contrary to standard panel data practice, data are generally meant to be stacked by time, then cross-section (so that the individual index is the "fastest" one) in order to simplify formulas especially as regards spatial filtering representations.[5] In the following, we will denote the composite error term in the standard linear regression model as u; ε will stand for the idiosyncratic error term, as opposed to the random effect μ, so that $u = \mu + \varepsilon$ throughout. Error covariance matrices will be denoted by Ω if unscaled, by Σ if scaled by the innovation's variance σ_e^2, so that $\Omega = \sigma_e^2 \Sigma$. Lastly, as we consider balanced panel datasets with N spatial units observed over T time periods, the dimension of spatial and covariance matrices involved will usually be $NT \times NT$. In some cases, though, we will consider $N \times N$ ($T \times T$) submatrices pertaining to one cross-section (time period), denoting them, e.g., as Σ_N (Σ_T). The spatial weight matrix $W_N = W$ is assumed time invariant, as customary in the literature, and enters spatial panel models as $I_T \otimes W_N$ where $\otimes$ is the Kronecker product, dropping the index N when unambiguous. Software package names are in **bold**, commands and arguments are in `typewriter` font.

Estimation of all models is based on maximum likelihood methods; an assumption of normality is maintained throughout. An assessment of the appropriateness of the procedures presented here under non-normality of errors, in the spirit of Lee and Yu (2012), is left for future work.

The paper is organized as follows: the next section discusses the specification; then we review estimation theory, building on existing approaches to illustrate ours. In the subsequent section we address the practical aspects of estimation, from computational issues to the design of a user-friendly software package. A practical illustration and the conclusions follow.

[5] While important for presentation clarity, this is nevertheless completely transparent for software users, who need only supply suitably indexed `data.frames` or `pdata.frames`.

2. Spatial panel models with error components

In this section we discuss the common specifications of random effects spatial panel models most frequent in the literature. At the end we introduce the general spatial autoregressive model with random effects and both spatially and temporally autoregressive errors. This last is the most general specification we consider here, and also the main subject of the paper.

Spatial panel data models capture spatial interactions across spatial units observed over time. There is an extensive literature both on static as well as dynamic models. Here we consider a general static panel model that includes a spatial lag of the dependent variable and spatial autoregressive disturbances:

$$y = \lambda(I_T \otimes W_N)y + X\beta + u$$

where y is an $NT \times 1$ vector of observations on the dependent variable, X is a $NT \times k$ matrix of observations on the non-stochastic exogenous regressors, I_T an identity matrix of dimension T, W_N is the $N \times N$ spatial weights matrix of known constants whose diagonal elements are set to zero, and λ the corresponding spatial parameter. The disturbance vector is the sum of two terms

$$u = (\iota_T \otimes I_N)\mu + \varepsilon$$

where ι_T is a $T \times 1$ vector of ones, I_N an $N \times N$ identity matrix, μ is a vector of time-invariant individual specific effects (not spatially autocorrelated), and ε a vector of spatially autocorrelated idiosyncratic errors that follow a spatial autoregressive process of the form

$$\varepsilon = \rho(I_T \otimes W_N)\varepsilon + e$$

with ρ as the spatial autoregressive parameter, W_N the spatial weights matrix and $e \sim IID(0, \sigma_e^2)$.[6]

2.1. Spatial panels with (independent) random effects

As in the classical panel data literature, the individual effects can be treated as fixed or random. In a random effects specification, the unobserved individual effects are assumed uncorrelated with the other explanatory variables in the model, and can therefore be safely treated as components of the error term: see, e.g., Assumption RE.1.b in Wooldridge (2002, 10.4). In this case, $\mu \sim IID(0, \sigma_\mu^2)$, and the error term can be rewritten as:

$$\varepsilon = (I_T \otimes B_N^{-1})e$$

where $B_N = (I_N - \rho W_N)$. As a consequence, the composite error term becomes

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B_N^{-1})e$$

[6]The spatial weights matrices in the lag and the error term can differ (see the following). $I_N - \rho W_N$ is assumed non-singular.

and its variance-covariance matrix, if $J_T = \iota_T \iota_T^\top$ is a $T \times T$ matrix of ones, is

$$\Omega_u = \sigma_\mu^2 (J_T \otimes I_N) + \sigma_e^2 [I_T \otimes (B_N^\top B_N)^{-1}]. \tag{1}$$

In deriving several lagrange multiplier (LM) tests, Baltagi et al. (2003) consider a panel data regression model that is a special case of the model presented above in that it does not include a spatial lag of the dependent variable. Elhorst (2003, 2009) defines a taxonomy for spatial panel data models both under the fixed and the random effects assumptions. Following the typical distinction made in cross-sectional models, Elhorst (2003, 2009) defines the fixed as well as the random effects panel data versions of the spatial error and spatial lag models. However, unlike Case (1991), he does not consider a model including both the spatial lag of the dependent variable and a spatially autocorrelated error term. Therefore, the models reviewed in Elhorst (2003, 2009) can also be seen as special cases of this more general specification.

In the following we will use the acronyms SAR (as in Spatial AutoRegressive) to indicate the presence of a spatial lag; SEM (Spatial Error Model) for a spatially autoregressive process in the error[7]. The combined model will be termed SAREM[8]. The suffix RE stands, as usual, for Random Effects.

2.2. Spatially correlated random effects

A different specification for the disturbances was considered in Kapoor et al. (2007). They assume that spatial correlation applies to both the individual effects and the remainder error components. Although the two data generating processes look similar, they do imply different spatial spillover mechanisms governed by a different structure of the implied variance covariance matrix. In this case, commonly referred to as "KKP", the composite disturbance term

$$u = (\iota_T \otimes I_N)\mu + \varepsilon$$

follows a first order spatial autoregressive process of the form:

$$u = \rho(I_T \otimes W_N)u + e.$$

It follows that the variance-covariance matrix of u is:

$$\Omega_u = [I_T \otimes B_N^{-1}]\Omega_\varepsilon [I_T \otimes (B_N{}^\top)^{-1}] \tag{2}$$

where $\Omega_\varepsilon = [\sigma_e^2 I_T + \sigma_\mu^2 J_T] \otimes I_N$ is the typical variance-covariance matrix of a one-way error component model. The variance matrix in (2) is simpler than the one

[7]Although this is by far the most popular specification, the literature has also dealt with different types of spatial diffusion processes in the errors other than the autoregressive one, most notably the spatial moving average (SMA): $\varepsilon = e + \rho(I_T \otimes W_N)e$, see e.g. Fingleton (2008). We do not consider them here.

[8]This is also often labelled as SARAR; we discard this terminology to avoid further confusion with the *serially* autoregressive errors that will be introduced below.

in (1), and therefore its inverse is easier to calculate, as will be discussed below. As Baltagi et al. (2013) observe, the economic meaning of the two models is also different: in the first model only the time-varying components diffuse spatially, in the second spatial spillovers too have a permanent component.[9] We label this latter alternative specification SEM2RE, and its extension to including a spatial lag (see Mutl and Pfaffermayr, 2011) SAREM2RE.

2.3. Serial correlation in idiosyncratic errors

It is possible to generalize the structure of the errors further by introducing serial correlation in the remainder of the error term, together with spatial correlation and random effects. Baltagi et al. (2007b) do so in the context of the Anselin SEMRE, specifying the model errors as the sum of an individual, time-invariant component and an idiosyncratic one which is spatially autocorrelated and has serial correlation in the remainder:

$$\begin{aligned} &\varepsilon = \rho(I_T \otimes W_N)\varepsilon + \nu \\ &\nu_t = \psi \nu_{t-1} + e_t. \end{aligned} \tag{3}$$

The combination of this more general error structure, which we term SEM(2)SRRE because of the addition of Serially autoRegressive errors, with a spatially lagged dependent variable and the estimation of the most general model (SAREM(2)SRRE) will be our main purpose.

3. Alternative specifications and an encompassing one

Before discussing estimation of the specification detailed above, a short discussion of relevant alternative ones is in order. We focus on two different modelling strategies from the ones discussed here. The first, fixed effects methods, although theoretically much different, turns out to be computationally encompassed by the estimation framework discussed here, as discussed below. The second, considering spatial and serial autocorrelation simultaneously in one equation, is an alternative approach to space-time dependence in errors which will not be pursued in this paper but only briefly hinted at for comparison. Lastly, we set our model taxonomy in the context of the comprehensive model of Lee and Yu (2012).

3.1. Random versus fixed effects

The subject of random versus fixed effects is too broad in scope to be summarized here[10]. The modern approach to the issue, tracing back to Mundlak (1978) and summarized, among others, in Wooldridge (2002, 10.2.1), centers on the statistical properties of the individual effects, which despite traditional

[9] Lee and Yu (2012, 2.4) illustrate the difference between this latter specification and SEMRE through the likelihood of the *between* model.

[10] A short introduction with the basic references can be found in Baltagi (2008, 2.3.1)

terminology are always considered as random variates; the crucial distinction becoming whether one can assume them to be uncorrelated with the regressors or not. If uncorrelated, then individual effects can be considered as a component of the error term, and treated in a generalized least squares fashion as will be done in this paper. If not, then the latter strategy leads to inconsistency; the individual effects will have to be estimated or, more frequently, eliminated by first differencing or time-demeaning the data (see Wooldridge, 2002, 10.5). The Hausman (1978) test is the standard device for assessing the hypothesis of no correlation, and hence of using random effects methods. In a spatial setting, Lee and Yu (2012) give an extensive treatment to which the reader is referred here.[11]

From a computational viewpoint, nevertheless, according to the current standard framework of Elhorst (2003), fixed effects estimation of spatial panel models is accomplished as pooled estimation on time-demeaned data. Hence, it is fully encompassed by the methods described in this paper, which can reproduce the results of Elhorst's estimators provided the data are time-demeaned in advance.[12]

While still the standard in applied practice and available software, Elhorst's procedure has been questioned by Anselin et al. (2008) because time-demeaning alters the properties of the joint distribution of errors, introducing serial dependence: see Lee and Yu (2010b, p.257) for a discussion of the issue, and Millo and Piras (2012, p.33) for an evaluation if its practical significance through Montecarlo simulation. To solve the problem, Lee and Yu (2010a, 3.2) suggest a different orthonormal transformation of the data. The possibility of combining this latter transformation with the algorithms discussed in the next sections is out of the scope of the present paper and is left for future work.

3.2. Joint modelling of spatial and serial correlation

In this paper, spatial and serial correlation in the error terms are modelled sequentially, as in Equation 3 above, according to the most common approach in the literature (Baltagi et al., 2007b; Lee and Yu, 2012). Elhorst (2008), in the context of a panel model with neither spatial lags nor individual effects, considers spatial and serial error correlation simultaneously instead of sequentially.[13] In his approach,

$$\begin{aligned} y_t &= X_t\beta + u_t \\ u_t &= \rho W u_t + \psi u_{t-1} + e_t \end{aligned}$$

so that each error depends not only on its current spatial lag and on its own past values, but also on spatially lagged time lags (see Elhorst, 2008, Eq. 5). There-

[11] A spatial Hausman test for models without serial correlation is available in package **splm**, Millo and Piras (see 2012, 7.2).

[12] Infrastructure in package **plm**, in particular `model.matrix` and `pmodel.response` methods for `plm` objects, allow for efficient time-demeaning.

[13] Elhorst (2001) applies the same dependence structure to the regressand instead of the error term.

fore, the combination of the two effects requires stricter stationarity conditions implying a tradeoff between the spatial and the serial correlation coefficients.

This joint specification allows for a richer interaction structure, with more complex time-space influence; in a sense, trading the relative simplicity of sequential "time, then space" dependence for the ability to account for the direct influences of past, neighbouring errors, which are assumed out by the former. By contrast, the sequential specification allows for clearer separation between effects and lends itself more easily to the incorporation of time-invariant individual components, either spatially uncorrelated as in Anselin's SEMRE, or spatially dependent as in KKP's SEM2RE model. Whether the joint specification is more or less appropriate than the sequential one in a given setting is an empirical question. However, the joint approach is not pursued here.

3.3. Comparison with Lee and Yu's general model

The recent paper by Lee and Yu (2012) provides a comprehensive specification in the same line of research of Baltagi et al. (2007b), i.e. modelling correlation in time and space sequentially. Lee and Yu (2012, Eq.1) consider a spatial autoregressive lagged variable; a spatial process for both errors and random effects, with possibly different parameters as in the general random effects model of Baltagi et al. (2013); and moving average (MA) remainder errors in both the idiosyncratic error and the random effect. Translating to our notation:

$$
\begin{aligned}
&y = \lambda(I_T \otimes W_1)y + X\beta + u \\
&u = (\imath_T \otimes \mu) + \varepsilon \\
&\mu = \rho_1 W_2 \mu + (I_N + \delta_1 M_1)\eta \\
&\varepsilon = \rho_2(I_T \otimes W_3)\varepsilon + (I_T \otimes (I_N + \delta_2 M_2))\nu \\
&\nu_t = \psi \nu_{t-1} + e_t
\end{aligned}
$$

where W_1, W_2, W_3, M_1, M_2 are spatial weights matrices, and η, e are i.i.d.. This model is more general than we consider in this paper in that it allows for MA processes in the errors and for different spatial processes generating u and μ. With reference to this model, we will impose $\delta_1 = \delta_2 = 0$ throughout and, alternatively, either $\rho_1 = 0$ (SAREMSRRE) or $\rho_1 = \rho_2$ and $W_2 = W_3$ (SAREM2SRRE), corresponding respectively to the approaches of Baltagi et al. (2007b) (no spatial process in the individual effects) and Kapoor et al. (2007) (same spatial process for both errors and individual effects). Further restrictions will give rise to the various nested specifications, as detailed in the following.

4. Estimation: theoretical framework

In this section we review the theory of maximum likelihood estimation of (static) spatial panel models with random effects.

We will start from models with a spatially lagged dependent variable, spatial error correlation and a general covariance structure for the error, as described by Anselin (1988), without any panel structure.[14] We will proceed introducing random effects and sketching the estimation framework of Elhorst (2003, 2009) for random effects panels with either a SAR or a SEM structure, which is currently the standard in econometric applications. Lastly, we will set out our approach. We will look at the random effects specification for panel models as one particular type of error covariance structure, thus comprising spatial panels in Anselin's general framework.

This approach, with respect to the current solution based on partial time-demeaning, has the advantage of both theoretical and software modularity, leading to a greater flexibility in the choice of the covariance structure: in particular, as is the case of this paper, allowing for the coexistence of random effects and serial correlation together with the spatial effects.[15]

4.1. Spatial models with a general error covariance

Maximum Likelihood estimation with a general error covariance matrix has been outlined in Magnus (1978) (see also Anselin et al., 2008). If the error u is distributed as $N(0, \Omega)$ then the log-likelihood is

$$logL = (C) - \frac{1}{2} ln|\Omega| - \frac{1}{2} u^\top \Omega^{-1} u.$$

Particularizing this likelihood w.r.t. the case at hand and adding a *spatial filter* if needed provides a general framework for ML estimation of the models of interest. Anselin (1988), the classic reference on spatial econometric model estimation by ML, outlines the general procedure for a model with spatial lag, spatial errors and possibly nonspherical residuals as follows. Let us restrict, for the moment, to one cross-section and let our model be

$$\begin{aligned} y &= \lambda W_1 y + X\beta + u \\ u &= \rho W_2 u + \eta \end{aligned} \tag{4}$$

with $\eta \sim N(0, \Omega)$ and, in general, $\Omega \neq \sigma^2 I$. Two special cases of this general model are often found in applied literature: if $\rho = 0$ one has the spatial autoregressive (SAR) model , while if $\lambda = 0$ the spatial (autoregressive) error (SEM) model. Both usually include the hypothesis of spherical errors: $\Omega = \sigma^2 I$. Introducing the now-standard simplifying notation $A = I - \lambda W_1$, $B = I - \rho W_2$ the

[14]In his book Anselin (1988) already considered a SEM panel with random effects, deriving the model likelihood, as a special case.

[15]It must be noted that the algorythm outlined in Elhorst (2003) and based on the combination of partial demeaning with a spectral decomposition of the error covariance matrix does in principle allow for any covariance structure. Yet our implementation turns out much simpler to code and therefore easier to maintain, at the cost perhaps of a slight decrease in efficiency; this last subject remains nevertheless to be investigated, as the different software environments forbid a direct comparison between routines.

model becomes [16]

$$\begin{array}{l} Ay = X\beta + u \\ Bu = \eta. \end{array}$$

If there exists Ω such that $e = \Omega^{-\frac{1}{2}}\eta$ and $e \sim N(0, \sigma_e^2 I)$, and B is invertible, then $u = B^{-1}\Omega^{\frac{1}{2}}e$ and the model (4) can be written as

$$Ay = X\beta + B^{-1}\Omega^{\frac{1}{2}}e$$

or, equivalently,

$$\Omega^{-\frac{1}{2}}B(Ay - X\beta) = e$$

with e a "well-behaved" error.

Still following Anselin, making the estimator operational requires the transformation from the unobservable e to observables. Expressing the likelihood function in terms of y requires calculating the Jacobian of the transformation $J = det(\frac{\partial e}{\partial y}) = |\Omega^{-\frac{1}{2}}BA| = |\Omega^{-\frac{1}{2}}||B||A|$. These determinants are to be added to the log–likelihood, which becomes

$$logL = -\frac{N}{2}ln\pi - \frac{1}{2}ln|\Omega| + ln|B| + ln|A| - \frac{1}{2}e'e$$

where the difference w.r.t. the usual likelihood of the classic linear model is given by the terms of the Jacobian (which is $J = 1$ in that case, see Greene (2003), B.41). The likelihood is thus a function of β, λ, ρ and parameters in Ω.

It will be convenient for our purposes, and without loss of generality, to scale the overall errors' covariance writing it as $B'\Omega B = \sigma_e^2\Sigma$.[17] This likelihood can be concentrated w.r.t. β and the error variance σ_e^2 substituting $e = [\hat{\sigma}_e^2\Sigma]^{-\frac{1}{2}}(Ay - X\hat{\beta})$

$$logL = -\frac{N}{2}ln(\pi\hat{\sigma_e^2}) - \frac{1}{2}ln|\Sigma| + ln|A| - \frac{1}{2\hat{\sigma_e^2}}(Ay - X\hat{\beta})'\Sigma^{-1}(Ay - X\hat{\beta}) \quad (5)$$

and a closed-form GLS solution for $\hat{\beta}$ and $\hat{\sigma_e^2}$ is available for any given set of spatial and other covariance parameters

$$\begin{array}{l} \hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Ay \\ \hat{\sigma_e^2} = \frac{(Ay - X\hat{\beta})'\Sigma^{-1}(Ay - X\hat{\beta})}{N} \end{array} \quad (6)$$

so that a two-step procedure is possible which alternates optimization of the concentrated likelihood and GLS estimation.

[16]The following notation expressing a spatial lag model as $Ay = X\beta + u$ or, equivalently provided A is invertible, $y = A^{-1}(X\beta + u)$ is well known in the literature as "spatal filtering" representation.

[17]Notice that the latter expression is in fact more general, as it does not constrain the heteroskedastic error term η to be spatially lagged, through premultiplication by B, in its entirety. In our case, only the error covariance of the SEM2 specification can be separated into a heteroskedastic error term and a spatial filter, and therefore straightforwardly written as $B'\Omega B$, while the more common SEM specification cannot.

While an iterative solution is generally possible, Anselin (1988) provides a closed-form solution for the spatial lag model and a simplified iterative one for the spatial error model that are used in currently available software solutions, notably in the R package **spdep** (Bivand et al., 2010). Practical implementations of the general solution outlined above incur a number of computational difficulties, so that few computer programs are available that be able to cope with estimation of models more complicated than standard SAR and SEM. In the next paragraph we sketch the currently most popular solution for the estimation of spatial panels; in the following one we will outline our approach.

From here on, we explicitly consider the (balanced) panel structure of the data: N individuals observed over T time periods.

4.2. Spatial random effects panels: the demeaning solution

The standard algorythms for the estimation of SAR and SEM-type spatial panels are due to Elhorst (2003), whose **Matlab** routines are perhaps the most widely used piece of software for the econometric analysis of spatial panel data. As will be clear from the following, his approach is based on a combination of the partial time-demeaning technique familiar from standard panel data (see, eg., Wooldridge, 2002, Ch. 10) with Anselin's Maximum Likelihood framework: data are quasi-time-demeaned in order to eliminate the random effects structure, then standard SAR or SEM estimators are applied to the transformed data so that the first-order conditions in (6) simplify to those of OLS, plus a spatial filter on y in the SAR case. These estimation techniques are well described in Elhorst (2009); we briefly review them here in order to highlight the differences to our approach. We also sketch two straightforward extensions which, although implicit in Elhorst's work, have not been explicitly pursued.

Following Elhorst (2003), to estimate the SARRE model we just need to add spatial filtering on y using $I_T \otimes A = I_T \otimes (I_N - \lambda W)$ and the determinant of the spatial filter matrix, $|I_T \otimes A| = |A|^T$, to the likelihood of the random effects model. As the considerations on transforming variables to get rid of Σ^{-1} and reduce the GLS step to OLS from the standard (non-spatial) RE model still apply[18] [19], an efficient two-step procedure can be based on concentrating the likelihood with respect to β and σ_e^2 as

$$logL = -\frac{NT}{2} ln(2\pi\sigma_e^2) + \frac{N}{2} ln\theta + T ln|A| - \frac{NT}{2} ln(\tilde{e}'\tilde{e}),$$

where θ is the quasi-demeaning parameter and the residuals are those of the demeaned model with a spatial filter on y

$$\tilde{e} = (I_T \otimes A)\tilde{y} - \tilde{X}\beta,$$

[18] See the original paper for the definition of the quasi-demeaning parameter θ as a function of σ_ε^2 and σ_μ^2, with a one-to-one correspondence to the variance ratio ϕ used in this paper.

[19] The validity of Elhorst's procedures rests on a property granting that $\Sigma(I_N \otimes A)y = (I_N \otimes A)\Sigma y$, so that demeaning the spatially lagged data is equivalent to spatially lagging the demeaned data: see Mutl and Pfaffermayr (2011) and Remark A1 in Kapoor et al. (2007).

and maximizing it w.r.t. λ and θ; then iterating until convergence between this maximization and the GLS step, whose first order conditions are

$$\begin{array}{l} \hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}(I_T \otimes A)\tilde{y} \\ \hat{\sigma}_e^2 = \frac{\tilde{e}'\tilde{e}}{NT}. \end{array}$$

The transformation procedure for the SEM model is more complicated, requiring a spectral decomposition of the errors covariance, and is omitted here; it is thoroughly explained, e.g., in Elhorst (2003), pages 19-21. Elhorst's algorithm is also easily adapted to the simpler SEM2 specification, where the random effects are in turn spatially lagged together with the idiosyncratic error term.[20] Again, an efficient two-step procedure can be based on concentrating the likelihood with respect to β and σ_e^2 as

$$logL = -\frac{NT}{2}ln(2\pi\sigma_e^2) + \frac{N}{2}ln\theta + Tln|B| - \frac{NT}{2}ln(\tilde{e}'\tilde{e}),$$

where the residuals from the demeaned model are spatially filtered

$$\tilde{e} = (I_T \otimes B)(\tilde{y} - \tilde{X}\beta),$$

and maximizing it w.r.t. ρ and θ; then iterating between this maximization and the GLS step, whose first order conditions are those of standard OLS on transformed data

$$\begin{array}{l} \hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}\tilde{y} \\ \hat{\sigma}_e^2 = \frac{\tilde{e}'\tilde{e}}{NT}. \end{array}$$

Although not explicitly stated by the author, Elhorst (2003)'s methodology is also easily extended, by combination, to the SAREM(2) specification (for an application see Millo and Pasini, 2010), but it does not lend itself as easily to extensions in the direction of serially correlated errors. As anticipated, in the next section we will abandon the popular demeaning framework, working instead on untransformed data and approaching random effects together with, and not differently from, any other feature of the error covariance.

4.3. General maximum likelihood framework

In this section we discuss our procedure for estimation of the general SAREM-SRRE model with spatial lag and error, random effects and serial correlation in the remainder error term. In economic terms, this specification can accommodate spatial spillovers from the outcome variable at neighbouring locations (SAR), spatial diffusion of idiosyncratic shocks (SEM), individual heterogeneity (RE) and time-persistence of idiosyncratic shocks (SR).

[20] We realized this reading an early working paper version of Elhorst (2003), although to our knowledge the author later chose not to pursue the estimation of this alternative specification in his research.

As anticipated, we build on the framework from Anselin (1988) outlined above, explicitly particularizing and operationalizing it with respect to a number of possible error covariance structures without resorting to variable transformation. As noted above, we start from considering the spatial dependence features together with all the other sources of heteroskedasticity and correlation instead of separating it clearly as done in the original Anselin framework. This has the advantage of keeping some components of the error term (most notably, the random effects) out of the spatial dependence, which can remain a feature of the idiosyncratic error only, in accordance with most applications in the literature (see, e.g.,Anselin et al. (2008); Baltagi et al. (2007b, 2003, 2013); Baltagi and Liu (2008); Baltagi et al. (2007a, 2009); Debarsy and Ertur (2010); Elhorst (2003); Elhorst and Freret (2009); Elhorst (2008, 2009, 2010); Elhorst et al. (2010); Lee and Yu (2010c,a,b, 2009); Mutl (2006); Mutl and Pfaffermayr (2011)) but some clear computational disadvantages, as will be discussed below. Moreover, we consider the alternative specification where the individual effects are lagged together with the idiosyncratic errors, as in Kapoor et al. (2007), which one can straightforwardly express in terms of Anselin's original expression $E(uu^\top) = B^\top \Omega B$, also extending the structure of Ω to include serial correlation. This latter will turn out to be much easier to compute, and the only feasible solution on some large examples.

First we will discuss the combination of a spatial lag with any error covariance structure; then we will review the most significant among the latter; lastly we will give an example of operationalization through the use of analytical expressions for the inverse and determinant of the error covariance matrix Σ.

Optimization will generally be subject to box constraints according to the following rules: the spatial lag and spatial errors coefficients λ and ρ will be bounded between $1/\omega_{min}$ and 1, where ω_{min} is the smallest characteristic root of W (see the standard conditions in Elhorst, 2008, Footnote 1); the serial correlation coefficient will be constrained to the usual stationarity condition $|\psi| < 1$ and the variance ratio of the random effects ϕ to be non-negative.

4.3.1. Spatial lag

Although both the SAR and the SEM specifications are popular in the literature, estimation generally focuses on one effect only, and there are few applications allowing for both of them to be present in the estimated model, one notable exception being the pioneering work of Case (1991). As far as software is concerned, routines for estimating a general SAREM model are available for cross-sectional data in the **R** package **spdep** but not, to our knowledge, in any panel data package. It is nevertheless straightforward, at least as far as expressing the likelihood is concerned, to combine a spatial lag with any error structure, including spatial dependence ones.

The general likelihood for the spatial lag panel model combined with any error covariance structure Σ is a panel version of (5):

$$\begin{aligned} logL = & -\frac{NT}{2} ln(2\pi\sigma_e^2) - \frac{1}{2} ln|\Sigma| + Tln|A| \\ & - \frac{1}{2\sigma_e^2}[(I_T \otimes A)y - X\beta]'\Sigma^{-1}[(I_T \otimes A)y - X\beta]. \end{aligned} \tag{7}$$

The usual iterative procedure a la Oberhofer and Kmenta (1974) can be employed to obtain the maximum likelihood estimates. Starting from an initial value for the spatial lag parameter λ and the error covariance parameters, we obtain estimates for β and σ_e^2 from the first order conditions:

$$\begin{aligned} \hat{\beta} &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(I_T \otimes A)y \\ \hat{\sigma_e^2} &= [(I_T \otimes A)y - X\beta]^\top \Sigma^{-1}[(I_T \otimes A)y - X\beta]/NT. \end{aligned} \tag{8}$$

The likelihood can be concentrated and maximized with respect to the parameters in A and Σ. The estimated values thereof are then used to update the expression for Σ^{-1}. These steps are then repeated until convergence. In other words, for a specific Σ the estimation can be operationalized by a two steps iterative procedure that alternates between GLS (for β and σ_e^2) and concentrated likelihood (for the remaining parameters) until convergence.

This general scheme can be applied to the random effects case, where it provides a simple and effective equivalent to the usual partial time-demeaning procedure, as well as to all the more complicated error covariance specifications discussed in the following.

For example, the spatial autoregressive model with random effects (SAR-RE) can be written as a combination of spatial filtering on the regressand and a random effects structure in the errors:

$$\begin{aligned} (I_T \otimes A)y &= X\beta + u \\ u &= (\imath_T \otimes \mu) + e \end{aligned}$$

hence it can be estimated by "plugging into" the general likelihood (7) the particular scaled error covariance $\Sigma_{RE} = \phi(J_T \otimes I_N) + I_{NT}$ characterized by one parameter: $\phi = \sigma_\mu^2/\sigma_e^2$, the ratio of the variance of the individual effect over that of the idiosyncratic error.

4.3.2. Error structures

In this section we describe the different error structures to be possibly combined with a spatial lag in the way illustrated above.

The Random Effects SEM and SEM2 models. As already discussed, the spatial error, random effects model gives rise to two possible specifications, depending on the interaction between the spatial autoregressive effect and the individual error components: the SEMRE specification first analyzed by Anselin (1988) where only the idiosyncratic error is spatially correlated:

$$\begin{aligned} y &= X\beta + u \\ u &= (\imath_T \otimes \mu) + \varepsilon \\ \varepsilon &= \rho(I_T \otimes W)\varepsilon + e \end{aligned}$$

with the scaled errors' covariance (denoting $\bar{J}_T = J_T/T$ and $E_T = I_T - \bar{J}_T$):

$$\Sigma_{SEMRE} = \bar{J}_T \otimes (T\phi I_N + (B^\top B)^{-1}) + E_T \otimes (B^\top B)^{-1}$$

and that of Kapoor et al. (2007) where the same spatial process applies both to the individual and the idiosyncratic error component:

$$\begin{aligned} y &= X\beta + u \\ u &= (\imath_T \otimes \mu) + \varepsilon \\ u &= \rho(I_T \otimes W)u + e \end{aligned}$$

where the scaled errors' covariance is:

$$\Sigma_{SEM2RE} = (\phi J_T + I_T) \otimes (B^\top B)^{-1}.$$

Serial and spatial correlation in the Random Effects model. Generalizing the structure of the errors further by introducing serial correlation in the remainder of the error term, together with spatial correlation and random effects, we specify the model errors as the sum of an individual, time-invariant component and an idiosyncratic one which is spatially autocorrelated and has serial correlation in the remainder:[21]

$$\begin{aligned} y &= X\beta + u \\ u &= (\imath_T \otimes \mu) + \varepsilon \\ \varepsilon &= \rho(I_T \otimes W)\varepsilon + \nu \\ \nu_t &= \psi \nu_{t-1} + e_t. \end{aligned}$$

To derive the likelihood, Baltagi et al. (2007b) use a Prais-Winsten transformation of the model with random effects and spatial autocorrelation. Following their simplifying notation, define

$$\begin{aligned} V_\psi &= \frac{1}{1-\psi^2} V_1 \\ V_1 &= \begin{bmatrix} 1 & \psi & \psi^2 & \dots & \psi^{T-1} \\ \psi & 1 & \psi & \dots & \psi^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi^{T-1} & \psi^{T-2} & \psi^{T-3} & \dots & 1 \end{bmatrix}; \end{aligned}$$

then the expression for the scaled error covariance matrix Σ can be written as[22]

$$\Sigma_{SEMSRRE} = \phi(J_T \otimes I_N) + V_\psi \otimes (B^\top B)^{-1}.$$

While in principle only an expression of the errors' covariance is needed and its inverse and determinant can be calculated by brute force inside the optimization loop, in practice it is convenient, and even necessary, to rely on simplified analytical expressions to reduce the computational burden and extend the range of feasible sample sizes.

[21]Based on this comprehensive model, Baltagi et al. (2007b) derived a number of conditional and marginal LM tests for the different effects, possibly allowing for the presence of the other ones. In doing so, they also derived the log-likelihoods for all combinations of random effects, spatial and serial correlation and a number of simplified expressions for the inverses and determinants of the covariance matrices involved, which we extensively use in the procedures described here.

[22]See Baltagi et al. (2007b) for details on the derivation of the covariance matrix.

As an example, we report expressions for the inverse and determinant of the error covariance matrix for the most complicated error specification, $\Sigma_{SEMSRRE}$ taken from Baltagi et al. (2007b):

$$\begin{aligned}\Sigma_{SEMSRRE}^{-1} &= V_{\psi}^{-1} \otimes (B^{\top}B) + \tfrac{1}{d^2(1-\psi)^2}(V_{\psi}^{-1}J_T V_{\psi}^{-1}) \\ &\quad \otimes ([d^2(1-\psi)^2\phi I_N + (B^{\top}B)^{-1}]^{-1} - B^{\top}B) \\ |\Sigma_{SEMSRRE}| &= |d^2(1-\psi)^2\phi I_N + (B^{\top}B)^{-1}| \cdot |(B^{\top}B)^{-1}|^{T-1}/(1-\psi^2)^N,\end{aligned}$$

where $\alpha = \sqrt{\frac{1+\psi}{1-\psi}}$ and $d^2 = \alpha^2 + (T-1)$.[23] They can be plugged in the general likelihood (7) to estimate the most general SAREMSRRE model.

Serial and spatial correlation in the KKP model. As an alternative to the SAREMSRRE specification, we consider an extension of the SEM2RE errors a la Kapoor et al. (2007) to serial correlation in the remainder errors.[24] As in the SEM2RE case, the random effects are spatially lagged together with the idiosyncratic ones, while the remainder errors ν in turn are serially correlated:

$$\begin{aligned} y &= X\beta + u \\ u &= (\imath_T \otimes \mu) + \varepsilon \\ u &= \rho(I_T \otimes W)u + \nu \\ \nu_t &= \psi\nu_{t-1} + e_t. \end{aligned}$$

This alternative specification, which to our knowledge is undocumented in the literature but for being one possible restriction of the very general formulation in Lee and Yu (2012), assumes that individual effects follow the same spatial diffusion process as the idiosyncratic errors do. By analogy, we term it SAREM2SRRE. Just as in the SEM2RE case, the error covariance is then again of the $B^{\top}\Omega B$ form (see Footnote 17), which simplifies computations considerably. In fact, the (scaled) error covariance for this model is

$$\Sigma_{SEM2SRRE} = (\phi J_T + V_{\psi}) \otimes (B^{\top}B)^{-1}$$

and, by the properties of Kronecker products, its inverse is

$$\Sigma_{SEM2SRRE}^{-1} = (\phi J_T + V_{\psi})^{-1} \otimes (B^{\top}B)$$

[23] It must be noted that V_{ψ}^{-1} has a simple and self-similar closed-form solution irrespective of dimension. As Baltagi et al. (2007b, App. A.2) note, from the well-known Prais-Winsten transformation $CV_{\psi}C' = I_T$, hence one has $V_{\psi}^{-1} = C'C$. Therefore, it is easily shown that for any T, V_{ψ}^{-1} is bisymmetric and its structure fully described by the $T = 3$ case:

$$V_{\psi}^{-1} = \begin{bmatrix} 1 & -\psi & 0 \\ -\psi & \frac{1-\psi^4}{1-\psi^2} & -\psi \\ 0 & -\psi & 1 \end{bmatrix}$$

to be extended along the main diagonal if $T > 3$ (see also Lee and Yu, 2012, A.2, p.1397). As such, V_{ψ}^{-1} quickly becomes sparse as T grows. Although in practical applications the time dimension is unlikely to be huge, relying on an analytical inverse for V_{ψ} already gives sizeable benefits in speed and numerical stability for moderate T.

[24] A remark from an anonymous reviewer indirectly motivated us to explore this alternative.

so that there is no need for the numerically demanding and unstable inversion of $B^\top B$.[25]

Simpler error structures and a general taxonomy. By restricting to zero the parameters of the most general error covariance structure with spatial and serial correlation plus random effects (SEMSRRE) one can get all simpler cases, e.g serial correlation (SR, if $\lambda = \phi = 0$), simple random effects (RE, if $\lambda = \rho = 0$) or any combination thereof.

In the Table 1 below you can see a taxonomy of the available specifications based on which of the spatial lag and error covariance parameters are non-zero; considering the SEM2SRRE and SEM2RE alternatives, with or without SAR, adds up to 20 different ones. Each of these can be estimated according to the same general principles outlined in the previous sections, and in particular based on the general likelihood in (7) where appropriate expressions for Σ^{-1} and $|\Sigma|$ have been plugged in.

$par \neq 0$		$\psi\rho$	ρ	ψ	(none)
λ	ϕ	SAREM(2)SRRE	SAREM(2)RE	SARSRRE	***SARRE***
(none)	ϕ	SEM(2)SRRE	***SEM(2)RE***	*SRRE*	*RE*
λ	(none)	SAREMSR	SAREM	SARSR	**SAR**
(none)	(none)	SEMSR	**SEM**	*SR*	*OLS*

Table 1: Model taxonomy based on nonzero spatial lag (λ) and error covariance parameters. ϕ is the ratio of the random effect's to the idiosyncratic error's variance, so that $\phi = 0$ means no random effect. Standard spatial models are in **bold**, non-spatial models in *italics*, standard spatial panel models in ***bold italics***.

5. Estimation in practice

In the preceding section we have seen how to translate the most general specification into a computationally manageable likelihood function. We now turn to the practical task of optimizing the latter and obtaining parameter estimates. While in principle it is enough to write the function and feed it to an optimization routine, this becomes a very hard task for the computer if there are many parameters involved. One crucial help is given by the Oberhofer and Kmenta (1974) two-step procedure, which allows to concentrate the likelihood with respect to β and σ_e^2, drastically reducing the parameter space; then, optimization over a maximum of four parameters is a task a modern computer can handle over a fair share of the sample sizes encountered in applied spatial panel data practice. Yet while most of the specifications we discuss can be estimated on a cross-sectional dimension of some thousand data points, some are limited to a few hundredths: in particular, the Anselin-Baltagi style combinations of random

[25]Moreover, Baltagi et al. (2007b) give a closed-form solution for $(\phi J_T + V_\psi)^{-1}$ in terms of V_ψ^{-1}, to which all the simplifying results from the previous footnote apply.

effects and spatial or spatio-temporally correlated errors (SEMRE, SEMSRRE) suffer of numerical problems in the inversion of the cross-product $B^\top B$. As a general rule, space-time correlated errors (SEMRE, SEM2RE, SEMSR, SEMSRRE, SEM2SRRE) are more complicated to estimate, while combinations of spatial lags with timewise correlated errors are relatively fast. Some reasons for this will be explained in the following.

All methods described here allow for W_N and its transformations, the spatial filters A and B, as generic numeric matrices; specialized methods are nevertheless used that consider, and exploit, their sparseness characteristics to improve speed and numerical stability. For the sake of clarity, we start describing the main computational issues without considering the particular nature of A and B. The implementation of these methods will be described at the end of the section as an incremental improvement.

5.1. Computational issues

Combining spatial filtering with GLS, spatial panel models are affected by the two typical computational bottlenecks of both techniques.

The first, which is typical of all spatial econometric literature and has inspired a great deal of research, is the calculation of the determinant of the spatial filters A and B: here, of A and Σ (see Bivand, 2010). The second is the inversion of the Σ matrix and the calculation of the GLS step.

5.1.1. Computing the log determinant

For all the attention reserved to date to efficient computation of the determinant, in this setting it is the GLS step that turns out as the limiting factor, despite there being analytical expressions of the inverse available for each one of the specifications considered here. Performance problems, mainly RAM-related, kick in at sample sizes where, by contrast, brute-force calculation of the determinant is still relatively inexpensive. In fact, all SAR estimators relative to simple structures in Σ would be very fast even if one used a relatively unsophisticated approach to the calculation of $log|A|$ through the standard **R** function based on the LU decomposition (see R Development Core Team, 2012), which does not assume any particular structure for A. The use of sparse matrix methods (see below, Section 5.1.3) begins to give important advantages in this respect when $N > 1000$. From comparisons done on a 3075×3075 real-world proximity matrix of US counties, sophisticated sparse matrix methods dramatically outperform naive, brute-force calculation of the determinant: the naive method does still accomplish its task in a few seconds (to be multiplied by a practical average of about 10 to 30 optimization loops), but sparse matrix methods make estimation of a SAR model faster by an order of magnitude.

Anyway, as we detail in the following, the true challenge lies with the efficiency of the GLS step. For this, together with the sparseness of B, it is crucial to exploit some kind of structure in the (inverse) Σ matrix.

5.1.2. GLS computational strategies

In the two-step optimization procedure outlined above, likelihood parameters enter both A and Σ, implying that all inverses and determinants have to be recomputed at every loop; the optimizer can also be expected to need more loops to converge, the more parameters there are in the concentrated likelihood. Therefore, especially for the richer specifications, it is crucial to have fast and efficient code modules calculating the GLS step inside the likelihood function. To this end, one can either concentrate on minimizing the dimensions of the matrices to invert or on taking advantage of their peculiarities in using specialized software methods.

We therefore classify the various error structures according to the shape of the relevant covariance:

- scalar (OLS)
- block-diagonal by cross-sections (SEM)
- block-diagonal by time periods (SRRE, RE, SR)
- dense (SEMRE, SEM2RE, SEMSR, SEMSRRE, SEM2SRRE)

Basically, any combination of spatial dependence with time persistence, be it AR-type serial correlation or random effects, gives rise to a dense matrix: in these cases we cannot but calculate the GLS step as one; importantly, making use of the analytical expression of the inverse avoids numerical inversion of the full $NT \times NT$ Σ matrix.

In the other cases, excluding the scalar one which is trivial, Σ is block-diagonal either by region or by time; we follow two different routes for the two cases, considering that in the former the block dimension is bound to be bigger in practical applications, as in the typical spatial panel $N >> T$.

By-block GLS. In the by-region block-diagonal case, where $\Sigma = I_T \otimes \Sigma_N$, we exploit the analytical expression of the inverse and the fact that

$$\hat{\beta} = (\sum_{t=1}^{T} X_t^\top \Sigma_N^{-1} X_t)^{-1} (\sum_{t=1}^{T} X_t^\top \Sigma_N^{-1} A y_t) = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} (I_T \otimes A) y$$

in order to reduce the maximum dimension of the objects involved from NT to N. The GLS step becomes a sequence of T matrix products between elements of dimension $N \times K$ and $N \times N$. No numerical inversion is needed any more, and there is no need to store all the zeros in the off-diagonal blocks, with an obvious benefit to RAM usage.

Object-oriented GLS. In the by-time block-diagonal case ($\Sigma = I_N \otimes \Sigma_T$), usually made up of many smaller blocks, we opt instead for an object-oriented solution.[26] After reordering the data internally, we define Σ as a `bdsmatrix` object and use specialized methods available in the **R** package by the same name (Therneau, 2009) to compute $X^\top \Sigma^{-1} X$ and $X^\top \Sigma^{-1}(I_T \otimes A)y$. It may seem inefficient to numerically invert a matrix when we have an analytical expression for the inverse, but the software methods involved are so fast as to beat the by-block solution by far on the sample sizes one is likely to encounter in practice.

Spatial filtering for SAR. Spatial filtering of the data is also done by applying W_N to stacked data by cross-section, precalculating $wy = (I_T \otimes W_N)y$ once and for all at the beginning; thus, no matrix operations are left in this area and updating the spatial filter inside the optimization loop as $(I_T \otimes A)y = y - \lambda wy$ only involves $NT \times 1$ vectors.

The combinations of time-block-diagonal error covariances with spatial lags therefore require multiple reordering of the data to alternatively exploit both structures. Fortunately, simple data reordering is computationally very cheap.[27]

5.1.3. Sparse matrix methods: a transparent object-oriented approach

Sparse matrix methods have always been prominent in spatial econometrics (see LeSage, 1999). In cross-sectional models, calculating the determinant of relatively large spatial filtering matrices has traditionally been the limiting factor. As observed above, in the context of a more complicated panel specification and on a modern computer this is not the primary bottleneck any more: yet numerical efficiency and stability, particularly in the computation and inversion of $B^\top B$ inside the GLS step, already improves in very moderate sized datasets. Computing times, slightly worse for small datasets because of the overhead introduced by data transformation and flow control structures, start to benefit with N in the hundredths and become huge thereafter. Spatial error specifications that are infeasible or borderline if using standard matrix methods become reasonably fast through sparse matrix methods.[28]

The **R** package **spam** (Furrer and Sain, 2010) provides functionality for sparse matrix algebra aimed at end-user transparent interoperability with standard `matrix` objects. This means that once a matrix has been defined as sparse, it can still be treated in a formally identical way as dense ones: specialized meth-

[26] For a more thorough explanation of the object-oriented approach to GLS, see Croissant and Millo (2008, Section 3.3).

[27] Sophisticated methods for spatially lagging a vector instead of simply premultiplying it by W are available in the **R** package `spdep`. Due to the other limiting factors, though, the feasible sample sizes here are such as to make their benefits practically irrelevant.

[28] Testing non-sparse versions of the estimators on the same machine described in Table 2 on artificial data drawn on a 3075×3075 proximity matrix (US counties, see below) and 4 time periods, the SAREM2RE completed the task in about 70 minutes (against 4'12" using sparse methods), the SEM2RE in 39' (2'24"); all other spatial random effects models hit memory limits, while using sparse algebra RAM usage was below 1.3 Gb; by contrast, the SARSRRE converged in 1'40" (against 8"), the SARSR in 49" (2") and the SARRE in 58" (5'): much slower than sparse counterparts, but not limiting in applied settings.

ods will be automatically employed when appropriate, triggered by a matrix's `class` attribute.

Specifically, suitable objects are defined as `spam`. The result of operations on them will in turn be either a `matrix` or a `spam` object which will be treated accordingly by further matrix operations downstream, without any need for code duplication and `if` conditions, retaining the `spam` class as long as sparse, while seamlessly changing into a regular `matrix` as soon as "becoming dense". As an example, with reference to Section 4.3.2, B is defined as sparse; the results of $(B^\top B)$ and $[V_\psi^{-1} \otimes (B^\top B)]$ are still sparse matrices, and are treated as such in the GLS step, while by contrast, $(B^\top B)^{-1}$ is dense, which is the reason behind the huge difference in performance between KKP-style estimators with SEM2SRRE or SEM2RE errors and their Anselin-style counterparts.

As a side note, the $|B|$ and $B^\top B$ operators are coded in an object-oriented fashion on the input side as well, according to the `class` attribute of the object containing the spatial weights: if it is a `matrix` then standard methods are used; if it is a `listw`, the appropriate counterparts are employed. Thus, the estimating functions can indifferently accept a spatial weights matrix W of each type. Spatial lagging for wy (see Section 5.1.2) is written in a flexible fashion too, accepting both matrices and `listw` objects, more for reasons of consistency and ease of maintenance than in order to reap a speed advantage from using the fast `lag.listw` method over this single, and simple, operation.

5.1.4. Computing times

Below we report the timings for all specifications over artificial data in seven different typical sample sizes: the first is a small panel of 49 observations (using the Columbus proximity matrix from package `spdep`) over 7 time periods; the second extends the timespan of the former to 50 time periods; the following ones are artificial "circular" proximity matrices a la Pesaran (see Pesaran and Tosetti, 2011) with 100, 200, 400 elements and 7 or 15 time periods. The last one is the proximity matrix of all 3075 adjacent counties of the USA (from example data in package **spam**). The data are generated according to the SAREMSRRE specification with two regressors and an intercept.

Despite the computational burden, estimation is reasonably fast even on the bigger examples, very fast for any error structure with only time dependence, even if coupled with a SAR term. By contrast, spatial error dependence specifications prove to be harder to estimate, the SEM2 predictably less so than the SEM. The fails are all due to singular matrix errors in the calculation of $(B^\top B)^{-1}$, while total RAM usage remains in the region of 1 Gb (gross of about 300 Mb for the operating system).

5.1.5. Reliability

Multi-parameter optimization of a relatively complex likelihood can have reliability issues, like all numerical procedures. The main issues here concentrate on the SEMSRRE and SEMRE error cases, with or without spatial lag, with occasional fails due to the $B^\top B$ matrix becoming computationally singular. The

	49x7	49x50	100x7	200x7	200x15	400x15	3075x4
sar+semsrre	13.07	92.67	27.86	59.07	124.36	-	-
sar+sem2srre	8.41	65.55	8.70	18.59	22.32	135.42	252.21
sar+semre	7.95	82.14	15.07	41.00	-	-	-
sar+sem2re	5.27	49.71	5.17	11.23	23.64	81.18	254.91
sar+semsr	4.34	57.03	5.05	10.62	23.30	65.80	243.64
sar+srre	1.32	5.16	1.65	2.54	1.94	2.91	8.17
sar+sem	1.70	19.29	2.08	4.77	8.60	38.34	114.51
sar+re	0.94	3.01	0.90	1.04	3.01	2.25	5.14
sar+sr	0.76	3.60	0.86	1.61	1.26	3.47	2.04
sar+ols	0.27	0.41	0.28	0.28	0.27	1.43	0.72
semsrre	8.82	92.78	14.94	36.68	103.85	-	-
sem2srre	4.94	49.64	4.37	8.39	20.62	75.55	253.09
semre	5.26	67.44	11.35	31.56	-	-	-
sem2re	3.06	36.54	2.68	5.88	13.36	50.78	144.18
semsr	2.06	25.51	2.88	4.44	10.55	46.44	168.19
srre	0.14	0.95	0.13	0.28	0.31	0.95	0.57
sem	0.92	8.59	1.00	2.22	5.01	15.56	53.71
re	0.09	0.57	0.08	0.18	0.34	0.52	0.38
sr	0.08	0.61	0.09	0.17	0.36	0.46	0.39
ols	0.01	0.03	0.02	0.03	0.05	0.07	0.08

Table 2: Computing times (seconds) for seven different sample sizes (NxT): 49x7, 49x50, 100x7, 200x7, 200x15, 400x15, 3075x4. R 2.15 on Linux Mint 9, Intel Core i7 720QM with 4 Gb RAM. Missing values are fails due to singular matrix errors.

frequency of such problems increases with bigger examples, where the degree of sparseness of W typically grows. For the reasons given above, the alternative SEM2SRRE, SEM2RE specifications are free from this problem and no failures have been recorded during extensive trials.

Statistical inference is based on standard errors obtained from the GLS step for β, from the numerical Hessian evaluated at the ML parameter values using finite differences as far as the error components are concerned. As such, given the optimal values of the error components and spatial lag parameter, GLS-based standard errors are based on a closed-form solution. By contrast, while the standard error estimates based on the numerical Hessian are generally accurate[29], negative estimates of parameter variances may happen, especially when estimation ends up in corner solutions or when the estimated model is overspecified with respect to the true data generating process (DGP), i.e., when one or more of ϕ, ψ, ρ and λ are zero. The by far most frequent case happens with $\hat{\sigma}_\phi$ when the 'true' value is zero. This behaviour, although undesirable, is generally confined to unambiguous cases where the parameter in question is

[29] For an evaluation of the accuracy of this method, see Section 8 in Millo and Piras (2012).

ostensibly not significant.

As is always the case with numerical optimization of complex problems, it can be useful to try different settings for the optimizer, which one can easily do (see Section 5.2.1), or even different optimization routines (see Section 5.2.2): in particular, although marginally slower, the `''BFGS''` method has proven more resilient than `''nlminb''` to the problem of negative covariances.

5.2. Software design

The software mentioned in this paper extends the **splm** add-on package for spatial panel econometrics (see Millo and Piras, 2012) inside the **R** project for statistical computing (R Development Core Team, 2012). In the following paragraphs we shortly discuss the user interface and then single out the main guiding principles behind the organization of the underlying computing functions.

5.2.1. User interface

Estimators in **splm** are organized according to estimation method into two main functions, `spgm` for generalized moments and `spml` for maximum likelihood. Each combination of individual effects (fixed, random, none), spatial error and spatial lag can be estimated by either method (see, again, Millo and Piras, 2012).

The estimators assuming no serial correlation are therefore already available at user level through the general wrapper function `spml`, specifying the `model` argument as `''random''` or `''none''`, the `spatial.error` argument as `''b''` for (Anselin/) Baltagi type (SEM) spatial dependence, `''kkp''` for Kapoor, Kelejian and Prucha-type dependence (SEM2), or `''none''` for no spatial error; and `lag` as `TRUE` or `FALSE` for adding a spatial lag or not (see Millo and Piras, 2012, Section 5.1).

The new estimators allowing for serial correlation, not being documented yet in the literature and excluding fixed effects, are for the time being not available through `spml`. In fact, when estimating random effects or pooling models, the `spml` function in turn calls a second-level wrapper, `spreml`, now also available at user level[30], whose syntax is illustrated below in Section 6 and which provides a more flexible yet still user-friendly way of calling the different estimator functions, specifying the `error` argument (covariance structure) as in Table 3.

$par \neq 0$	$\psi\rho$	ρ	ψ	(none)
ϕ	`semsrre`	`semre`	`srre`	`re`
(none)	`semsr`	`sem`	`sr`	`ols`

Table 3: Admissible values for the `errors` argument to `spreml()` and corresponding nonzero parameters in the errors covariance. `sem2re`, `sem2srre` can be used in place of `semre`, `semsrre` to obtain the specifications a la KKP instead of those a la Anselin/Baltagi.

[30]Since version 1.0-03.

The other specification-related argument to `spreml` is boolean: `lag` as `TRUE` or `FALSE`, meaning whether to combine the chosen error structure with a spatial lag or not. Moreover, two different spatial matrices can be supplied for lag and error, defaulting to being the same if just one is supplied.

Concerning the optimization step, initial values for the covariance and spatial lag parameters can be set through the `initval` argument and chosen from `''zeros''` (setting all to zero) and `''estimate''`; when `initval` is set to `estimate` the initial values are retrieved from the estimation of nested specifications. As an example, when estimating `semsrre`, the initial value for the serial correlation parameter is taken to be the estimated ψ from a panel regression with serially correlated errors. Analogously, the initial value of ρ is the estimated spatial autocorrelation coefficient from the `sem` model; and, finally, an initial value for ϕ is obtained estimating a random effect model. Another possibility is to feed `initval` an arbitrary numeric vector of initial values.

Any admissible parameter can be passed on to the optimizer function by means of the special "`...`" (*dots*) argument[31]; moreover, the boolean `quiet` argument allows tracing each optimization loop, both as a status indicator and as a diagnostic tool.

At the output end, estimation produces a regular `splm` object. `splm` objects loosely inherit from `plm` objects (Croissant and Millo, 2008) and therefore behave like most model objects in **R**, exposing, inter alia, `coef` and `vcov` elements to be used in diagnostics and model summary methods. `splm` objects, though, extend the `plm` class adding special equivalents of "coef" and "vcov" elements for the SAR and the error parameters (respectively `arcoef`, `vcov.arcoef`, `errcomp` and `vcov.errcomp`) which the specialized `summary.splm` method uses to display estimates and diagnostics. These elements are `NULL` if not present in the specification, thus preserving compatibility.[32]

Consistently with the rest of the **R** system, all model objects have a `logLik` element, recording the value of the log-likelihood at the optimum, complying with the generic `logLik()` extractor function in base R and allowing for likelihood ratio tests on covariance parameters.

As customary in the **R** environment, the `summary` method prints a short description of the model, the most recent call, a summary of the residuals and the table of estimated coefficients. The **splm** specific part of the output (between the summary of the residuals and the table of the estimated coefficients) reports on the estimated spatial lag and error components along with standard errors from the numerical hessian.

[31]The special "..." argument is used in **R** as a placekeeper to allow passing on an arbitrary number of optional parameters to a lower-level function (see R Development Core Team, 2012).

[32]See Millo and Piras (2012, Par. 7.3) for an illustration of interoperability between `splm` objects and generic testing functions from other R packages, and Zeileis (2006) for a general discussion of object-orientation in econometric software.

5.2.2. Computing engine

The estimator functions are separate but constructed modularly, with a common main structure and separate code modules for the likelihoods, the GLS steps and the SAR spatial filters. This helps keeping maintenance costs at a minimum despite the proliferation of different estimators. Another guiding principle, in the spirit of Open-Source software, has been readability of the code. To this end, we have as far as possible constructed software counterparts to theoretical objects such as spatial filters or covariances, combining them and "plugging them in" as prescribed by the reference theory.

In the spirit of the **R** project, we have employed some peculiar features of the language. In particular abstraction of tasks (like, e.g., computing the determinants of spatial filters) into functions makes code easier to produce and maintain and future enhancements easier. As functions are a data type in R, and can therefore be passed on as arguments to other functions, the choice of which method (and corresponding software module) to use can be made at runtime; moreover, new methods can be passed on to the existing structure without any need to reprogram it.

The optimizer function can be chosen specifying a `method` argument in the call to `spreml`. The `method` defaults to `''nlminb''`, based on the PORT routines (Gay, 1990) and available in package **nlme** (Pinheiro et al., 2012), but the user is also allowed to choose between all the constrained optimizers supported by package **maxLik** (Toomet et al., 2012). While the latter are still deemed experimental in the package documentation, `method=''BFGS''` has proven only marginally slower than `''nlminb''` and well-behaved as regarding the issues of corner solutions and negative variance estimates discussed above.

6. Examples

In this section we illustrate estimation by means of examples. As Lee and Yu (2012, p.1394) recommend, in each case we run various specifications, and especially the comprehensive ones, on the data in order to assess the properties of more specific ones in a general-to-specific fashion.

We reconsider three examples from the applied literature, with particular focus on the innovative features of the estimators described in the paper: serial correlation and the coexistence of spatial lags and errors. In the first case, the extended estimators discussed in this paper basically validate previous results, the additional features identifying effects which either prove redundant (the spatial lag) or significant but of limited magnitude (random effects and serial correlation). The second starts from a model with time persistence in errors (both individual effects and serial correlation) and no evidence of spatial effects when testing for *either* spatial lag *or* error, where by contrast spatial dependence features appear only when tested jointly inside an encompassing SAREM model. In the last example, a classic from the panel data literature, serial correlation, once included in the model, shows up with alarming magnitude, questioning the stationarity of the data; at the same time, the estimate of ϕ becomes unstable.

The illustrations therefore demonstrate the importance of the general model both for estimation and as a diagnostic tool.

6.1. Indonesian rice farming

Due both to the substantial relevance of the research question and to the availability of detailed data collected through governmental programs, the analysis of production efficiency in Indonesian rice farming has been a recurring subject of applied papers since Erwidodo (1990) and later work by Lee and Schmidt (1993); Trewin et al. (1995); Horrace and Schmidt (1996); Druska and Horrace (2004); Feng and Horrace (2012). Moreover, the related subject of market demand for rice in Indonesia provided the empirical application for the original paper on SAREM panel models, Case (1991).

The rice farming example, focusing on many small farms which can reasonably be seen as randomly drawn from a bigger population, provides a good example of the usefulness of random effects methods in the econometric analysis of spatial panels. 171 rice farms in Indonesia are observed over six growing seasons, three wet and three dry, between 1975 and 1983. The farms are located in six different villages of the Chimanuk River basin in West Java. The production frontier equation relates rice output to the following inputs: seed, urea, phosphate (tsp), labour hours (lab) and land (size), all but phosphate in logs. Dummy variables account for the use of high yield varieties of seed (high), or for a mix of seed varieties (mixed) and for the use of pesticides. Dummy variables are also added for the six villages and for the season being a wet one. Following Druska and Horrace (2004), the proximity matrix is constructed considering all the farms of the same village as neighbours.[33] [34]

As Druska and Horrace (2004) summarize, "[o]f the six villages included in the sample, two are on the north coast of the island in an area with average altitudes of 10-15 meters above sea level. Another three villages are in an area (600-1100 meters above sea level) in the central part of West Java. The last village is in the center of the island with an average altitude of 375 meters. The infrastructure in the Cimanuk River Basin is fairly heterogeneous. Some of the villages (in both high and lowland areas) lack reliable transportation systems, and local roads are almost impassable in the wet (rainy) season. Other villages, located in close proximity to province capital cities, are highly accessible along paved, all-weather roads." As such, one can expect both village-level heterogeneity and spatial correlation between farms belonging to the same village. Spatial dependence is easier to justify for the error terms, due to spillovers across neighbouring farms in idiosyncratic factors and climate conditions; more

[33] Data and weights matrix are available in the **splm** package as, respectively, `RiceFarms` and `riceww`.

[34] The full model formula statement is `ricefm` ← `log(goutput)` ∼ `log(seed) + log(urea) + phosphate + log(totlabor) + log(size) + I(pesticide>0) + I(varieties=="high") + I(varieties=="mixed") + as.factor(region) + I(as.numeric(time) %in% c(1,3,5))`; hence the syntax for estimating, e.g., the SAREMSRRE model is `mod` ← `spreml(ricefm, data=RiceFarms, w=riceww, errors=''semsrre'', lag=T)`.

difficult to find reasons for the inclusion of a spatial lag of the dependent variable, as it seems unrealistic for the outcome in one farm to influence those of neighbours.

The estimation results of all specifications are reported in Table 6.1[35] [36]. When considering spatial error dependence, the effect of phosphate (tsp) roughly halves; the coefficient on pesticide use in turn becomes negative and non significant; the magnitude of the influence from using high yield or mixed seeds is reduced and the positive effect for the Ciwangi region disappears, showing its nature as an artefact of neglected spatial correlation.

As for error covariance parameters, there is some evidence of individual effects, but their variance is less than one fifth of that of the idiosyncratic errors; the serial correlation coefficient $\hat{\psi}$ is also significant and positive, but rather small. While the spatial lag coefficient $\hat{\lambda}$ is not significant, provided spatial error correlation has been controlled for, the latter, at $\hat{\rho} = 0.65 - 0.74$, is consistent with the results of Druska and Horrace (2004) (based on a generalized moments estimator) and supports their choice of specification. It is noteworthy how a spatial lag specification, which as observed has a weaker theoretical basis, nevertheless yields a highly significant coefficient of about 0.4 if the SEM term is omitted.

6.2. Italian non-life insurance

Millo and Carmeci (2011) analyze the determinants of per-capita equilibrium consumption of non-life insurance in all 103 Italian provinces over five years, 1998 to 2002, based on socioeconomic characteristics of territory: per-capita income (rgdp) and wealth as proxied by bank deposits (bank), real lending rates (rirs), territorial density of population (den) and of the distribution network (agen); demographic characteristics as average family size (fam) and schooling (school) and the prevailing level of trust; lastly, on the share of agriculture on value added and on the level of inefficiency of civil justice (inef).[37]

As some of the regressors are time-invariant while others have little variability over time, fixed effects are not an option; hence they make the case for controlling heterogeneity through five macroregional variables (NW, NE, Centre, South and Islands, meaning Sicily and Sardinia) plus provincial random

[35] The estimates of the βs are not directly comparable between SAR (upper half of the table) and non-SAR specifications (lower half) since the marginal effects of the models extended to include a spatially lagged dependent variable are not equal to the coefficient estimates of those variables (on this issue, see LeSage and Pace, 2009, p. 74)

[36] A function `allmodels.R` replicating the estimates in the table on user-supplied data has been added to the supplementary materials for convenience of the reader. The arguments are, as customary, model formula, data and spatial weights matrix; the output is a matrix of estimated parameters. E.g., Table 6.1 is reproduced as `print(allmodels(ricefm, data=RiceFarms, w=riceww))`. We thank an anonymous referee for the suggestion.

[37] Data and weights matrix are available in the **splm** package as, respectively, `Insurance` and `itaww`. The model formula is `log(ppcd) ~ log(rgdp) + log(bank) + log(den) + rirs + log(agen) + school + vaagr + log(fam) + log(inef) + log(trust) + d99 + d00 + d01 + d02 + NorthWest + NorthEast + South + Islands`.

	1	seed	urea	tsp	lab	size	pest	high	mix	lambda	phi	psi	rho
sar+semsrre	4.03	0.12	0.13	0.61	0.23	0.51	-0.01	0.12	0.10	0.17	0.17	0.09	0.67
sar+sem2srre	4.00	0.12	0.13	0.61	0.23	0.51	-0.01	0.12	0.10	0.18	0.16	0.09	0.66
sar+semre	4.03	0.12	0.13	0.65	0.23	0.51	-0.01	0.12	0.10	0.17	0.20		0.67
sar+sem2re	3.98	0.12	0.13	0.65	0.23	0.51	-0.01	0.12	0.10	0.18	0.19		0.67
sar+semsr	4.04	0.12	0.13	0.58	0.23	0.51	-0.01	0.13	0.10	0.17		0.20	0.65
sar+srre	2.59	0.12	0.12	1.01	0.21	0.50	0.02	0.12	0.10	0.40	0.11	0.11	
sar+sem	4.01	0.12	0.14	0.65	0.22	0.51	-0.01	0.13	0.09	0.17			0.64
sar+re	2.55	0.12	0.13	1.07	0.20	0.50	0.03	0.11	0.09	0.40	0.15		
sar+sr	2.66	0.12	0.13	0.96	0.21	0.50	0.01	0.13	0.10	0.38		0.18	
sar+ols	2.65	0.12	0.14	1.02	0.20	0.51	0.02	0.12	0.08	0.39			
semsrre	5.23	0.12	0.13	0.58	0.23	0.50	-0.01	0.12	0.10		0.17	0.09	0.73
sem2srre	5.23	0.12	0.13	0.58	0.23	0.50	-0.01	0.12	0.10		0.16	0.09	0.73
semre	5.23	0.12	0.13	0.62	0.23	0.50	-0.01	0.12	0.10		0.20		0.74
sem2re	5.23	0.12	0.13	0.62	0.23	0.50	-0.01	0.12	0.10		0.19		0.74
semsr	5.22	0.12	0.13	0.55	0.23	0.51	-0.01	0.13	0.10			0.20	0.71
srre	5.01	0.14	0.15	1.33	0.22	0.47	0.03	0.18	0.16		0.00	0.16	
sem	5.21	0.12	0.14	0.62	0.22	0.51	-0.01	0.13	0.09				0.71
re	5.02	0.14	0.15	1.47	0.21	0.46	0.05	0.18	0.16		0.06		
sr	5.01	0.14	0.15	1.33	0.22	0.47	0.03	0.18	0.16			0.16	
ols	5.04	0.14	0.15	1.39	0.21	0.47	0.04	0.18	0.15				

Table 4: Parameter estimates for all specifications on the rice farming model. Regressors are logged, except for tsp and the dummy variables pest, high and mix. tsp is divided by 1000 for readability.

effects, in the same spirit as the previous example. Unlike the other examples, serial correlation, which is a known feature of insurance data, is included from the beginning, while spatial effects are checked for both by means of the Baltagi et al. (2007b) LM test (C.1) and through the Pesaran (2004) CD test, and ultimately excluded. One would therefore expect the extension of their specification to the full SAREMSRRE model to be redundant.

The results in Table 6.2 confirm time persistence as the most important feature to be accounted for. With respect to a purely spatial model, the estimated effect of GDP halves when considering any type of time persistence. For bank deposits this positive bias is even higher, and it is highest for interest rates. The coefficient of trust on the contrary is seriously underestimated if omitting time persistence.

The estimates of the random effects' variance $\hat{\phi}$ and the serial correlation coefficient $\hat{\psi}$ are very stable across all specifications. If considered jointly, $\hat{\phi} = 10.2 - 10.9$ and $\hat{\psi} = 0.52 - 0.53$; else, if taken in isolation, one predictably picks up the effect of the other. Moreover, the precision of estimates (not shown in the table) is good and the algorithm successfully distinguishes between the contribution of the time-invariant component μ and that of serial correlation in ε to error persistence[38].

Interestingly, from the point of view of spatial effects, while neither a SAR nor a SEM effect prove significant if taken in isolation, estimating a SAREM specification identifies a positive SAR and a negative SEM term, both of moderate magnitude, a behaviour consistent with the ex-ante considerations in Millo and Carmeci (2011, 5.1). These two effects ostensibly compensate each other in reduced specifications[39]; taking them into consideration does not change the qualitative conclusions of the study exception made for the loss of significance of one variable (the density of the distribution network), but suggests further insights into the spatial dimension of the phenomenon.

6.3. The productivity puzzle

Munnell (1990) investigated the productivity of public capital (roads, water facilities, other infrastructure) in 48 US States observed over 17 years[40]. Her model is a Cobb-Douglas production function relating the gross social product (gsp) of a given state to the input of public capital (pcap), private capital (pc) and labour (emp); state unemployment rate (unemp) is meant to capture business cycle effects. For further details, see also Example 3 in Baltagi (2008).

In Table 6.3 we present parameter estimates for all possible specifications.[41]

[38] See the discussion of this issue in the next example.

[39] The Baltagi et al. (2007b) LM (C.1) test employed in the original paper, asymptotically equivalent to the Wald test on the SEM term in a spatial model, unsurprisingly accepted the null of no spatial effect.

[40] Data are available in the **Ecdat** package (Croissant, 2010). The spatial weights matrix based on binary contiguity for the US states is included in **splm** as `usaww`.

[41] Table 6.3 is reproduced as `print(allmodels(log(gsp) ~ log(pcap) + log(pc) + log(emp) + I(unemp/100), data=Produc, w=usaww))`.

	1	rgdp	bank	den	rirs	agen	school	vaagr	fam	inef	trust	lambda	phi	psi	rho
sar+semsrre	-1.58	0.30	0.13	0.09	-1.16	0.14	0.16	-0.66	-0.09	-0.19	1.13	0.25	10.92	0.53	-0.35
sar+sem2srre	-1.94	0.32	0.13	0.06	-1.06	0.14	0.12	-0.63	-0.09	-0.14	1.03	0.34	10.77	0.53	-0.43
sar+semre	-0.84	0.28	0.10	0.08	-1.49	0.13	-0.07	-0.53	-0.16	-0.20	1.22	0.23	13.70		-0.35
sar+sem2re	-1.32	0.30	0.10	0.06	-1.40	0.14	-0.05	-0.52	-0.16	-0.15	1.10	0.33	13.26		-0.42
sar+semsr	-2.09	0.32	0.14	0.06	-0.95	0.14	0.18	-0.77	-0.05	-0.15	0.95	0.36		0.95	-0.44
sar+srre	-1.14	0.31	0.17	0.08	-1.41	0.17	0.11	-0.79	-0.13	-0.18	1.48	0.04	10.16	0.53	
sar+sem	-4.72	0.63	0.41	0.05	-6.82	0.23	-0.33	-0.26	-0.15	-0.12	0.38	0.06			0.10
sar+re	-0.14	0.27	0.12	0.08	-1.89	0.16	-0.14	-0.77	-0.21	-0.19	1.58	0.02	13.18		
sar+sr	-1.53	0.32	0.20	0.08	-1.21	0.17	0.19	-0.83	-0.09	-0.18	1.39	0.05		0.94	
sar+ols	-5.24	0.67	0.40	0.04	-6.14	0.21	-0.32	-0.25	-0.16	-0.11	0.37	0.09			
semsrre	-1.04	0.31	0.16	0.08	-1.38	0.17	0.13	-0.77	-0.12	-0.18	1.58		10.72	0.52	-0.08
sem2srre	-1.05	0.31	0.17	0.07	-1.40	0.17	0.11	-0.79	-0.13	-0.18	1.59		10.57	0.53	-0.03
semre	-0.15	0.28	0.11	0.08	-1.78	0.16	-0.13	-0.69	-0.19	-0.19	1.65		13.68		-0.10
sem2re	-0.14	0.28	0.12	0.07	-1.84	0.16	-0.14	-0.74	-0.20	-0.19	1.66		13.58		-0.04
semsr	-1.40	0.32	0.20	0.08	-1.21	0.17	0.19	-0.84	-0.09	-0.18	1.50			0.94	-0.01
srre	-1.02	0.31	0.17	0.08	-1.41	0.17	0.11	-0.79	-0.13	-0.18	1.57		10.46	0.53	
sem	-4.21	0.60	0.43	0.05	-7.35	0.24	-0.38	-0.29	-0.14	-0.13	0.44				0.18
re	-0.07	0.27	0.12	0.07	-1.88	0.16	-0.14	-0.76	-0.21	-0.19	1.63		13.41		
sr	-1.39	0.32	0.20	0.08	-1.21	0.17	0.19	-0.84	-0.09	-0.18	1.49			0.94	
ols	-5.06	0.69	0.41	0.04	-6.23	0.22	-0.40	-0.30	-0.18	-0.10	0.48				

Table 5: Parameter estimates for all specifications on the Italian insurance model. Regressors are logged, except for rirs, school and vaagr, which are divided by 100 to make coefficients easier to read.

Estimating the full model with two possible sources of spatial correlation gives us an implicit Wald test for SAR vs. SEM, which in this case favours the latter. From the results we gather also that errors persistence in time is so high as to cast a doubt on the appropriateness of a static specification on levels, or at least warrant some unit root testing. The very high dispersion of the estimate of the random effects variance is also a consequence of the high persistence of the remainder error: for a discussion of related computational problems see Calzolari and Magazzini (2012).[42]

Munnell's model is well known for yielding spurious significance to public capital if not accounting for individual effects (Baltagi and Pinnoi, 1995); from the results it is apparent that controlling for serial error correlation also reduces the relative parameter's estimate. Notice also, by contrast, the relative stability of the SEM parameter ρ; how the SAR parameter λ picks up the spatial correlation when ρ is left out of the specification (exception made for the case of spherical errors); and especially the dramatic reduction in the parameter of private capital (pc) in any specification accounting for serial correlation in the remainder errors. Lastly, $\phi = 0$ in the SRRE model is an example of optimization ending up in a corner solution, as discussed in Calzolari and Magazzini (2012); interestingly, the problem is avoided when also controlling for spatial correlation.

Summing up: the example higlights the usefulness of our new estimation procedure in different ways. One of the main economic results of estimation, the magnitude of private capital productivity, is drastically changed by extending the specification to serially correlated errors. From the point of view of spatial correlation, the evidence in favour of SEM versus SAR indicates that it seems to be due to a diffusion process in the idiosyncratic shocks affecting the economy, rather than to the outcomes of each state influencing each other. Lastly, from a diagnostic viewpoint: the estimate for ψ being so close to one casts the shadow of nonstationarity on the specification as a whole and warrants unit root testing. These features would hardly have emerged without considering serial correlation in a comprehensive model.

7. Conclusions

This paper describes the theoretical approach and the practical implementation in the R software of maximum likelihood estimation of panel models incorporating: random effects and spatial dependence in the error terms; a spatially lagged dependent variable; and possibly also a serial dependence structure in the remainder of the error term.

[42] This is a structural aspect of the problem, unrelated to the precision of the numerical Hessian-based t-test appearing in the summary. In fact, as regards this last aspect, retrieving the likelihoods from the SAREMSRRE and SAREMSR models and performing an asymptotically equivalent likelihood ratio test we get the test statistic $LR = 2(2023.046 - 2022.924) \sim \chi^2_1$ which corresponds to a p-value of 0.6217, very similar to the 0.6216 we get from the t-test.

model	1	pcap	pc	emp	unm	λ	ϕ	ψ	ρ
sar+semsrre	2.96	0.04	0.07	0.91	-0.25	0.01	8.20	0.989	0.62
sar+semre	2.37	0.04	0.24	0.74	-0.35	0.00	7.53		0.54
sar+sem2re	2.29	0.05	0.24	0.74	-0.37	0.00	6.68		0.52
sar+semsr	2.91	0.04	0.07	0.91	-0.25	0.01		0.991	0.61
sar+srre	1.24	0.08	0.02	0.74	-0.27	0.30	0.92	0.997	
sar+sem	1.33	0.14	0.37	0.56	-0.90	0.01			0.52
sar+re	1.66	0.01	0.23	0.67	-0.58	0.16	21.32		
sar+sr	1.24	0.08	0.02	0.74	-0.27	0.30		0.997	
sar+ols	1.67	0.15	0.31	0.60	-0.66	-0.00			
semsrre	3.05	0.04	0.07	0.91	-0.25		9.08	0.988	0.63
semre	2.39	0.04	0.24	0.74	-0.34		7.50		0.54
sem2re	2.32	0.04	0.25	0.74	-0.36		6.62		0.53
semsr	3.04	0.04	0.07	0.91	-0.25			0.991	0.62
srre	2.74	0.10	0.07	0.88	-0.53		0.00	0.987	
sem	1.41	0.14	0.37	0.56	-0.86				0.52
re	2.14	0.00	0.31	0.73	-0.61		5.00		
sr	2.74	0.10	0.07	0.88	-0.53			0.987	
ols	1.64	0.16	0.31	0.59	-0.67				

Table 6: Parameter estimates for all specifications on the Munnell model. Regressors are logged, except for unemp. The latter is divided by 100 for readability. The SRRE estimator hit a corner solution at 0 for ϕ.

We have started by sketching a taxonomy of spatial panel models, beginning with the two basic random effects (RE) specifications used in the literature: the spatial autoregressive (SAR) RE model containing a spatially lagged dependent variable and a group-specific, time-invariant component in the error term, and the spatial error (SEM) RE model, with both a group-specific component and a spatial dependence structure in the error term. We have discussed the combination of SAR and SEM features as in Case (1991) and the extension of the SEM specification by Baltagi et al. (2007b) to an encompassing model allowing for random effects, serial and spatial correlation in the error term.

Building on the combination of Anselin (1988)'s general estimation framework with the likelihood functions for the full model and the models restricted at the different levels derived by Baltagi et al. (2007b) leads to an estimation procedure for the full model and any zero-restriction of it, giving rise to a total of 20 different specifications. We have described our implementation of the extended procedure in the R language and discussed it critically, with particular attention to the issues of performance and reliability, also providing applied examples taken from the empirical panel literature.

The techniques described in this paper extend the functionalities of the **splm** package inside the **R** project for statistical computing, thus opening up the possibility of analyzing spatio-temporal (static) panels with a rich structure, estimating the coefficients and conducting specification searches from general to specific through Wald or likelihood ratio tests as an alternative to Lagrange multiplier testing.

Thanks to the efficiency of R, especially as regards the availability of user-friendly infrastructure for sparse matrices, and the power of modern computers, these estimators can be employed on problems of relatively high dimension, spanning many fields of the applied literature. Yet for the richest specifications in the Anselin-Baltagi variant there are limits, starting from N in the hundredths, which exclude many interesting case studies. Therefore, future developments concern mostly the aspects of performance and stability, concentrating on the numerical optimization step. The occasional problem of negative estimates for the parameters' variance, although confined to borderline cases, would find a solution through analytical derivation of the parameters' covariance matrix for the full SAREM(2)SRRE models, which is another interesting task for future work. Further directions for development also concern the interoperability of the present framework with the orthonormal transformation of Lee and Yu (2012) in order to provide functionality for their Hausman test for serially correlated models. Lastly, it would be interesting to assess the behaviour of the proposed estimators under nonnormality.

Acknowledgments

I am heavily indebted to Gianfranco Piras for fruitful collaboration in the development of the **splm** package which originated the extensions discussed here, for the essential performance tip of including the GLS step into the optimization routine and for some code snippets in the sparse matrix functionalities. In turn, Roger Bivand inspired the work of us both and of the spatial econometric R community at large. On the broad subject of spatial panels I also acknowledge countless useful comments and suggestions from Elisa Tosetti, Gaetano Carmeci and Giacomo Pasini. Yves Croissant deserves the merit for originally devising the framework for panel data econometrics with R and my gratitude for years of collaboration; in this instance, he also contributed the Rice Farming example data. The quest for modularity and especially for correspondence between statistical "objects" and software tools is inspired by the work of, and discussions with, Achim Zeileis. The paper has benefited greatly from the observations of the Associate Editor and two anonymous referees. Lastly, I thank for useful comments and suggestions participants to the following conferences: European Regional Science Association, Liverpool 2008; Applied Statistics Conference, Bled 2008; useR 2008, Dortmund; useR 2009, Rennes; Italian Congress of Econometrics and Empirical Economics, Ancona 2009; Spatial Econometric Association, Toulouse 2011; Computational and Financial Econometrics, London 2011; and to internal seminars at DEAMS (University of Trieste) and the University of Innsbruck.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publisher, Dordrecht.

Anselin, L., Le Gallo, J., Jayet, H., 2008. Spatial panel econometrics. In: Matyas, L., Sevestre, P. (Eds.), The econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice (3rd Edition). Springer-Verlag, Berlin Heidelberg, pp. 624 – 660.

Baltagi, B., 2008. Econometric Analysis of Panel Data, 4th edition. Wiley, New York.

Baltagi, B., Egger, P., Pfaffermayr, M., 2007a. Estimating models of complex fdi: are there third-country effects? Journal of Econometrics 140 (1), 260–281.

Baltagi, B., Egger, P., Pfaffermayr, M., 2009. A generalized spatial panel data model with random effects. CPR- Working paper series.

Baltagi, B., Egger, P., Pfaffermayr, M., 2013. A generalized spatial panel data model with random effects. Econometric Reviews 32, 650–685.

Baltagi, B., Liu, L., 2008. Testing for random effects and spatial lag dependence in panel data models. Statistics and Probability Letters 78, 3304–3306.

Baltagi, B., Pinnoi, N., 1995. Public capital stock and state productivity growth: Further evidence from an error components model. Empirical Economics 20, 351–359.

Baltagi, B., Song, S., Jung, B., Koh, W., 2007b. Testing for serial correlation, spatial autocorrelation and random effects using panel data. Journal of Econometrics 140 (1), 5–51.

Baltagi, B., Song, S., Koh, W., 2003. Testing panel data regression models with spatial error correlation. Journal of Econometrics 117 (1), 123–150.

Bivand, R., with contributions by Micah Altman, Anselin, L., Assuncao, R., Berke, O., Bernat, A., Blanchet, G., Blankmeyer, E., Carvalho, M., Christensen, B., Chun, Y., Dormann, C., Dray, S., Halbersma, R., Krainski, E., Lewin-Koh, N., Li, H., Ma, J., Millo, G., Mueller, W., Ono, H., Peres-Neto, P., Piras, G., Reder, M., Tiefelsdorf, M., , Yu., D., 2010. spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-27.

Bivand, R. S., 2010. Computing the jacobian in spatial models: an applied survey. NHH Discussion paper series 20/2010, 1–30.

Calzolari, G., Magazzini, L., 2012. Autocorrelation and masked heterogeneity in panel data models estimated by maximum likelihood. Empirical Economics 43 (1), 145–152.

Case, A., 1991. Spatial patterns in household demand. Econometrica 59 (4), 953–965.

Croissant, Y., 2010. Ecdat: Data sets for econometrics. R package version 0.1-6. URL http://CRAN.R-project.org/package=Ecdat

Croissant, Y., Millo, G., 2008. Panel Data Econometrics in R: The plm Package. Journal of Statistical Software 27 (2), 1–43. URL http://www.jstatsoft.org/v27/i02/

Debarsy, N., Ertur, C., 2010. Testing for spatial autocorrelation in a fixed effects panel data model. Regional Science and Urban Economics 40, 453–470.

Druska, V., Horrace, W. C., 2004. Generalized moments estimation for spatial panel data: Indonesian rice farming. American Journal of Agricultural Economics 86 (1), 185–198.

Elhorst, J., 2001. Dynamic models in space and time. Geographical Analysis 33, 119–140.

Elhorst, J., 2003. Specification and estimation of spatial panel data models. International Regional Sciences Review 26 (3), 244–268.

Elhorst, J., 2008. Serial and spatial error correlation. Economics Letters 100 (3), 422–424.

Elhorst, J., 2009. Spatial panel data models. In: Fischer, M. M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis. Springer, Berlin, Heidelberg, New York.

Elhorst, J., 2010. Dynamic panels with endogenous interactions effects when T is small. Regional Science and Urban Economics 40 (5), 272–282.

Elhorst, J., Freret, S., 2009. Yardstick competition among local governments: French evidence using a two -regimes spatial panel data model. Journal of Regional Science 49, 931–951.

Elhorst, J., Piras, G., Arbia, G., 2010. Growth and convergence in a multiregional model with space-time dynamics. Geographical Analysis 42 (3), 338–355.

Erwidodo, 1990. Panel data analysis on farm-level efficiency, input demand and output supply of rice farming in west java, indonesia. Ph.D. thesis, Michigan State University. Department of Agricultural Economics.

Feng, Q., Horrace, W. C., 2012. Alternative technical efficiency measures: skew, bias and scale. Journal of Applied Econometrics 27 (2), 253–268.

Fingleton, B., 2008. A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. Spatial Economic Analysis 3 (1), 27–44.

Florax, R., Folmer, H., Rey, S., 2003. Specification searches in spatial econometrics: the relevance of hendry's methodology. Regional Science and Urban Economics 3 (5), 557–579.

Furrer, R., Sain, S. R., 2010. spam: A sparse matrix R package with emphasis on mcmc methods for gaussian markov random fields. Journal of Statistical Software 36 (10), 1–25.
URL http://www.jstatsoft.org/v36/i10/

Gay, D., October 1990. Usage summary for selected optimization routines. Computing Science Technical Report 153, AT&T Bell Laboratories, Murray Hill, NJ 07974.

Greene, W., 2003. Econometric Analysis, 5th Edition. Prentice Hall.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–71.

Horrace, W. C., Schmidt, P., 1996. Confidence statements for efficiency estimates from stochastic frontier models. Journal of Productivity Analysis 7 (2-3), 257–282.

Kapoor, M., Kelejian, H., Prucha, I., 2007. Panel data model with spatially correlated error components. Journal of Econometrics 140 (1), 97–130.

Lee, L., Yu, J., 2009. A unified transformation approach to the estimation of spatial dynamic panel data models: stability, spatial cointegration and explosive roots. Ohio state University, manuscript.

Lee, L., Yu, J., 2010a. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154 (2), 165–185.

Lee, L., Yu, J., 2010b. Some recent developments in spatial panel data models. Regional Science and Urban Economics 40 (5), 255–271.

Lee, L., Yu, J., 2010c. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory 26, 564–597.

Lee, L., Yu, J., November 2012. Spatial panels: random components versus fixed effects. International Economic Review 53 (4), 1369–1412.

Lee, Y. H., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical efficiency. The measurement of productive efficiency: Techniques and applications, 237–255.

LeSage, J., 1999. The theory and practice of spatial econometrics. Tech. rep., University of Toledo (US).

LeSage, J., Pace, K., 2009. Introduction to Spatial Econometrics. CRC Press, Boca Raton, FL.

Magnus, J., 1978. Maximum likelihood estimation of the gls model with unknown parameters in the disturbance covariance matrix. Journal of Econometrics 7, 281– 312.

Millo, G., Carmeci, G., 2011. Non-life insurance consumption in italy: a subregional panel data analysis. Journal of Geographical Systems 13 (3), 273–298.

Millo, G., Pasini, G., 2010. Does social capital reduce moral hazard? a network model for non-life insurance demand. Fiscal Studies 31 (3), 341–372.

Millo, G., Piras, G., 2012. splm: Spatial panel data models in R. Journal of Statistical Software 47 (1), 1–43.
URL `http://www.jstatsoft.org/v47/i01/`

Mundlak, Y., 1978. On the pooling of time series and cross-section data. Econometrica 46, 69–85.

Munnell, A., January/February 1990. Why has productivity growth declined? productivity and public investment. New England Economic Review, 3–22.

Mur, J., Angulo, A., March 2009. Model selection strategies in a spatial setting: Some additional results. Regional Science and Urban Economics 39, 200–213.

Mutl, J., 2006. Dynamic panel data models with spatially autocorrelated disturbances, phD Thesis, University of Maryland, College Park.

Mutl, J., Pfaffermayr, M., 2011. The hausman test in a cliff and ord panel model. Econometrics Journal 14 (1), 48–76.

Oberhofer, W., Kmenta, J., 1974. A general procedure for obtaining maximum likelihood estimates in generalized regression models. Econometrica 42 (3), 579–590.

Pesaran, H., Tosetti, E., 2011. Large panels with common factors and spatial correlations. Journal of Econometrics 161 (2), 182–202.

Pesaran, M. H., 2004. General diagnostic tests for cross section dependence in panels. Working Paper 1229, CESifo.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Development Core Team, 2012. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-103.

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-12-7.
URL `http://www.R-project.org/`

Therneau, T., 2009. bdsmatrix: Routines for Block Diagonal Symmetric matrices. R package version 1.0.
URL `http://CRAN.R-project.org/package=bdsmatrix`

Toomet, O., Henningsen, A., with contributions from Spencer Graves, Croissant, Y., 2012. maxLik: Maximum Likelihood Estimation. R package version 1.1-2. URL `http://CRAN.R-project.org/package=maxLik`

Trewin, R., Weiguo, L., Erwidodo, S., Bahri, S., 1995. Analysis of the technical efficiency over time of west javanese rice farms. Australian Journal of Agricultural and Resource Economics 39 (2), 143–163.

Wooldridge, J., 2002. Econometric analysis of cross-section and panel data. MIT Press.

Zeileis, A., 2006. Object-oriented computation of sandwich estimators. Journal of Statistical Software 16 (9), 1–16.
URL `http://www.jstatsoft.org/v16/i09/`