THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 10-47

Wolfgang Polasek
Institute for Advanced Studies, Vienna, Austria
University of Porto, Porto, Portugal
The Rimini Centre for Economic Analysis (RCEA), Rimini, Italy

Richard Sellner
Institute for Advanced Studies, Vienna, Austria

Carlos Llano
Universidad Autónoma de Madrid, Facultad de Ciencias Económicas y Empresariales, Departamento de Análisis Económico, Madrid, Spain

CHOW-LIN METHODS IN SPATIAL MIXED MODELS

Chow-Lin methods in spatial mixed models

Wolfgang Polasek[a], Richard Sellner[b], Carlos Llano[c]

[a]*Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna, Austria, and University of Porto, Rua Campo Alegre, Portugal*
[b]*Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna, Austria*
[c]*Universidad Autónoma de Madrid, Facultad de Ciencias Económicas y Empresariales, Departamento de Análisis Económico, 28049 Madrid*

Abstract

Missing data in dynamic panel models occur quite often since detailed recording of the dependent variable is often not possible at all observation points in time and space. In this paper we develop classical and Bayesian methods to complete missing data in panel models. The Chow-Lin (1971) method is a classical method for completing dependent disaggregated data and is successfully applied in economics to disaggregate aggregated time series. We will extend the space-time panel model in a new way to include cross-sectional and spatially correlated data. The missing disaggregated data will be obtained either by point prediction or by a numerical (posterior) predictive density. Furthermore, we point out that the approach can be extended to more complex models, like flow data or systems of panel data. The panel Chow-Lin approach will be demonstrated with examples involving regional growth for Spanish regions.

Keywords: Space-time interpolation, Spatial panel econometrics, MCMC, Spatial Chow-Lin, missing regional data, Spanish provinces, MCMC, NUTS: nomenclature of territorial units for statistics.
JEL classification: C11, C15, C52, E17, R12 .

1. Introduction

Regional data restrictions are one of the most common and unwanted limitations for applied regional scientists. In many fields linked to economics, geography and environmental science data scarcity arise when analyzing phenomena at a fine spatial scale such as provinces, counties or districts. Similarly, it is difficult to find rich databases covering simultaneously different spatial levels, which impede the right evaluation of any phenomena affected by the spatial scale under consideration (population density, agglomeration of economic activity, segregation and integration processes, etc).

Email address: polasek@ihs.ac.at (Wolfgang Polasek)

Data restrictions at a disaggregated or lower level spatial scale could cause problems for several reasons. First, the larger the spatial scope of a survey the larger the number of observations required to reach a reasonable level of accuracy at such levels. This reason leads to a trade-off between the chosen spatial scale and the cost of collecting the data. Apart, data scarcity at disaggregated spatial scales is linked to problems of statistical and legal data protection of the surveyed agents. Due to this issue, it is common to encounter data restrictions in the smallest and less developed regions (provinces or equivalent areas) of a country, where there is a limited number of observations and the sample could almost coincide with the target population. Another potential source of a data-scale discontinuity is the variation of geographical boundaries, as it has been observed in the case of some Eastern European countries, which have undergone a profound transformation in their internal organization after gaining independence since 1989.
Consequently, past and present census data related to the most important economic and demographic variables (as agreed by the UN) is not available at a disaggregated spatial scale. Additionally, it is common that some of the best statistical sources for analyzing demographic and social issues -the census data- are available according to UN recommendations every 10 years. Thus many census data find often limited connections with other data sources that might be published on e.g. yearly basis on a different spatial scale (i.e. educational outcome at the county level in the US). Moreover, it could also be the case that the right 'spatial scale' for analyzing a specific economic phenomenon requires the use of data at a particular level of aggregation.

In relation to these complementary causes of data scarcity at certain disaggregated spatial levels, there is a prolific literature dealing with the sensibility analysis of the spatial scale on different types of analysis. Moreover, several authors have dealt with this issues suggesting quantitative techniques able to solve the problem. What it is of much interest for the approach presented in this paper is that although some of these works were developed in alternative fields, they take into account the concept of spatial autocorrelation and spatial heterogeneity.

According to Jelinski and Wu (1996) one of the most comprehensive treatment of the sensitivity of analytical results to the definition of the spatial units is found under the geographical concept of 'the Modifiable Areal Unit Problem' (MAUP) (Openshaw and Taylor (1979) Openshaw and Taylor (1981); Openshaw (1984); Fotheringham and Wong (1991); Amrhein and Wong (1996); Sui (2000)). According to this literature, the MAUP arises from the fact that areal units are usually arbitrarily determined, in the sense that they can be aggregated to form units of different sizes or spatial arrangements. As it is posed by Wong (2003), the impact of the MAUP is significant partly because the correlation of variables will change when data gathered at different scale levels are used. As he explains, in general, data are spatially 'smoothed' when they are aggregated to adjacent values, and thus less variation is preserved at the aggregated level (see Fotheringham and Wong, 1991). But if data have a strong positive spatial autocorrelation, the aggregation process will not remove much information as compared to negatively spatially autocorrelated data. Several research efforts have tried to correct or adjust for correlation among variables (e.g. Holt et al., 1996) in order to obtain relatively consistent statistical results across scale levels, but most of these procedures are either too computationally intensive or impractical. To this regard, focusing on the regression framework, Fotheringham et al. (2001) suggested that the spatially weighted regression could be a potential solution to the scale effect, while King (1995) opted for an error-bound method. Moreover, Wong (2001) proposed a spatial correlation approach for analyzing count variables, which can yield results relatively consistent across scale levels.

Regarding this literature, it is important to note that the main discussion is about how to conciliate the results obtained when regressing the same model (with the same variables) at different spatial scales -when data is available-, due to the heterogeneity of space and the presence of different source of spatial autocorrelation at each unit level. Linked to this topic, but with singular connotations, other authors focus on developing methods for estimating unavailable data at a certain spatial scale taking into account observed relations of the same variables in the past (but in the same spatial units) or at different levels of spatial aggregation (but in the same period of time).

For example, among the first group, Baltagi and Li (2006) considered the problem of prediction in a panel data regression model with spatial autocorrelation in the context of a simple demand equation for liquor. Their model was based on a panel of 43 states over the period 1965-1994, and took into account the spatial autocorrelation due to neighboring states and the individual heterogeneity across states. Then based on the model, they compared the performance of several predictors of the states' demand for liquor for 1 year and 5 years ahead, using OLS, fixed effects ignoring spatial correlation, fixed effects with spatial correlation, random-effects GLS estimator ignoring spatial correlation and random-effects estimator accounting for the spatial correlation. In this article, they found that for forecasts 2-5 years ahead, estimators that take into account the heterogeneity across the states yield the best forecasts.

In relation to the other group, Pavía et al. (2008) use geo-statistical procedures to build a spatial model of voting patterns, testing the model in three elections in Spain. They apply kriging (a spatial model) and co-kriging (in a spatiotemporal model version) to improve the accuracy of election night forecasts. The estimates use polling stations as basic locations in the context of election night forecasting. The idea is to forecast the unavailable polling stations (NAPS) from the current polling stations' incoming results and then, by aggregation, predict the final outcomes. Three alternative sets of predictions - spatial, spatiotemporal, and temporal forecasts - are obtained in each election for eight different moments in the election night. Spatial forecasts are based on kriging and provide estimates for each NAPS from the vote distribution observed in the stations in its vicinity. Spatiotemporal forecasts use co-kriging to find NAPS estimates based on both the spatial distribution of the vote and the relationship that exists at each station between votes of consecutive elections. Finally, temporal forecasts are made to provide comparison with the spatial strategies. Then they compare the results with actual outcomes and also to predictions made using models that use only historical data from polling stations in previous elections. According to their results, the use of spatial information strongly improves the accuracy of the prediction.

Finally, we find an additional example trying to overcome the data scarcity problem in lower spatial units for demographic variables. Wu and Murray (2005) discussed how population information is typically available for analysis in aggregate socioeconomic reporting zones, such as census blocks in the United States and enumeration districts in the United Kingdom. However, such data masks underlying individual population distributions and may be incompatible with other information sources (e.g. school districts, transportation analysis zones, metropolitan statistical areas, etc.). Moreover, as we do, they link this data scarcity issue with the modifiable areal unit problem (MAUP) described above. Then, they use impervious surface fraction derived from Thematic Mapper (TM) imagery to derive the underlying population of an urban region. A co-kriging method was developed to interpolate population density by modeling the spatial correlation and cross-correlation of population and impervious surface fraction.

From an alternative perspective, Polasek et al. (2009) have recently developed new methods for estimating unavailable data using spatial interpolation methods based on the Chow-Lin method (see Chow and Lin, 1971), the workhorse of the interpolation techniques successfully applied to complete data in time series problems. In that paper, the authors developed a procedure able to estimate omitted data at a low spatial level using available data at the aggregated one. The method was developed for cross section type of data. Based on this first attempt, this paper extend the spatial interpolation method to the case of panel data. We think of a regional data set that is completely observed at an aggregate level (like NUTS-2) and has to be broken down into smaller regional units (e.g. NUTS-3) conditional on observed disaggregated indicators. By means of this new approach, we will be able to tackle with some of the problems described above regarding the presence of 'holes' in spatial and temporal dimensions. In addition, we will be able to show that the results obtained with the panel data approach show better results compared to the ones obtained with the method designed for cross sections (see Polasek et al., 2009). Our new method is developed using spatial econometrics techniques, both for classic and Bayesian models, where the later has to be estimated by MCMC. To evaluate the new method, we estimate the panel model using the GDP of the Spanish regions at two different spatial scales for the period 2000-2004. With that model we forecast the GDP for the 52 Spanish provinces (at NUTS-3 level), based only on the information for the 18 Spanish regions (i.e. NUTS-2 GDP as dependent variable), and the high frequency socio-economic indicators at the NUTS-3 level. Then, to compare the results obtained with the actual series available at the NUTS-3 level, we computed forecast criteria. We point out that a significant spatial lag parameter leads to an improvement (through the so called gain term) in the spatial Chow-Lin prediction of the disaggregated data. The Bayesian MCMC method yield the best result among the models in the forecast experiment.

1.1. The Chow-Lin method: outline and assumptions

The Chow-Lin method can be considered as a prediction method for subunits that are unobserved at this disaggregated spatial or time scale. The process can be viewed as inter-diction, because it forecast on a scale 'between' the observed scale or fine-casting, because it makes forecasts at a finer scale. While interpolation refers to mathematical (deterministic) models for data completion at a disaggregate level, we use the word inter-prediction for a statistical (stochastic) model to predict missing disaggregate observation based on the observed aggregated data.

The general framework for all Chow-Lin 'inter-prediction' methods (whether in space or time or both) can be summarized as follows:

1. Establish a disaggregate model that will be used for the inter-diction.
2. Derive the aggregate model and the reduced form.
3. Estimate the disaggregated parameters with the observed aggregated data.
4. Complete the data by forecasting with the disaggregated model (inter-prediction).

For a successful application of the method we need the following assumptions:

Assumption 1. *Structural similarity: The aggregated model for* $\mathbf{y}_c$ *and the disaggregated model for* $\mathbf{y}$ *are structurally similar. This implies that variable relationships that are observed on an aggregated level are following the same empirical law as on a disaggregated level: the regression parameters in both models are the same.*

Assumption 2. *Error similarity: The spatially correlated errors have a similar error structure on an aggregated level and on a disaggregated level: The spatial correlations are not significantly different.*

Assumption 3. *Reliable indicators: The indicators to make the formats on a disaggregated level have sufficiently large predictive power: The* R^2 *(or the* F *test) is significantly different from zero.*

The paper is organized as follows. Section 2 outlines the Bayesian model of the spatial Chow-Lin (CL) method for cross-sectional data and reviews the results of Polasek et al. (2009), along with the error covariance matrix needed for the improved prediction of the missing values, which leads to the so-called spatial gain terms for predictions. In section 3 we extend the approach to a spatial panel model assuming a seemingly unrelated type of covariance structure and Chow-Lin method for panel data. The procedures are given in sections 4 and 5. The Bayesian Chow-Lin model for completing panel data is outlined in section 6. In the next section (section 7), we then apply the spatial panel Chow-Lin method to Spanish NUTS-2 and NUTS-3 data. As we observe all data on the disaggregated level, we will evaluate the quality of the spatial Chow-Lin method by comparing the predicted values for the NUTS-3 GDP to their observed values and calculate the usual forecast accuracy criteria. A final section concludes.

2. Review: Bayesian Chow-Lin method for completing spatial cross-sectional data

First, we review the cross-sectional spatial autoregressive (SAR) model with missing y observations as in Polasek et al. (2009). In the spatial Chow-Lin model we are interested in a cross-sectional vector $y_d : n \times 1$ that should have been observed at a certain point in time t, but could actually not be observed and is completely or partially missing. Instead we can observe the shorter, aggregated vector $y_a = Cy_d : N \times 1$ where C is a $N \times n$ aggregation matrix consisting of 0's and 1's, indicating which cells have to be aggregated together. First, we consider a disaggregated spatial regression model (indicated by the subscript d), because it is the model where we can make the Chow-Lin forecasts for the missing y_d observations on the left hand side of the SAR model:

$$y_d = \rho_d W y_d + X_d \beta_d + \varepsilon_d, \quad \varepsilon_d \sim N[0, \sigma^2 I_n]. \tag{1}$$

The *reduced form (RF)* is obtained by defining the spread matrix $R = I_n - \rho_d W$ for an appropriately chosen weight matrix W and the spatial correlation coefficient ρ_d:

$$y_d = R^{-1} X_d \beta_d + R^{-1} \varepsilon_d, \quad R^{-1} \varepsilon_d. \sim \mathcal{N}[0, \sigma^2 (R'R)^{-1}]. \tag{2}$$

The prior distribution for the parameters of the disaggregated model $\theta_d = (\beta_d, \sigma^{-2}, \rho_d)$ is proportional to

$$p(\boldsymbol{\beta}_d, \sigma^{-2}, \rho_d) \propto p(\beta_d) \cdot p(\sigma^{-2}) = \mathcal{N}[\beta_d \mid \beta_*, H_*] \cdot \Gamma[\sigma^{-2} \mid s_*^2, n_*],$$

since we assume an uniform prior for the spatial correlation $\rho_d \sim \mathcal{U}[-1, 1]$.

The C-aggregation of the *reduced form* model is obtained by multiplying with the $N \times n$ matrix C

$$Cy_d = CR^{-1}X_d\beta_d + CR^{-1}\varepsilon_d, \quad CR^{-1}\varepsilon_d \sim \mathcal{N}[0, \sigma^2 C(R'R)^{-1}C']. \tag{3}$$

We will write shorter for the covariance matrix:

$$\sigma^2\Omega(\rho_d) = \sigma^2 C(R'R)^{-1}C'.$$

The prior distribution for $\theta_d = [\beta_d, \rho_d, \sigma^2]$ of this reduced form model is given by

$$p(\theta_d) = \mathcal{N}[\beta \mid \beta_*, H_*] \cdot \Gamma[\sigma^{-2} \mid s_*^2, n_*]$$

since the prior for ρ_d is assumed to flat: $p(\rho_d) \propto 1$. $\Gamma[a, b] = \Gamma[ab/2, b/2]$ stands briefly for the gamma distribution of the residual precision σ^{-2}.

The joint distribution of the disaggregated model is with $\mathcal{D} = \{y_a, X_d\}$

$$p(\theta_d, \mathcal{D}) = \mathcal{N}[y_a \mid CR^{-1}X_d\beta_d, \sigma^2\Omega(\rho_d)]p(\theta_d). \tag{4}$$

2.1. MCMC for the Chow-Lin SAR model

For the MCMC procedure we need from the joint distribution in (4) three full conditional distributions (fcd's) which are briefly denoted by $p(\rho_d \mid \theta^c), p(\beta_d \mid \theta^c)$, and $p(\sigma^2 \mid \theta^c)$, where the disaggregated parameters are collected by the vector $\theta_d = (\rho_d, \beta_d, \sigma^2)$. Furthermore, θ denotes all the parameter of the model and θ^c the complementary parameter set that we need for the fcd's.

The MCMC procedure for the SAR model consists of 3 blocks of sampling, as is shown in the next theorem:

Theorem 1 (MCMC in the SAR model).
We consider the SAR model in (1) with joint distribution in (4). Then the MCMC estimation procedure is given by

1. *Draw* β *from* $\mathcal{N}\left[\beta \mid b_{**}, H_{**}\right]$
2. *Draw* ρ_i *by a Metropolis step:* $\rho_{new} = \rho_{old} + \mathcal{N}[0, \tau^2]$
3. *Draw* σ^{-2} *from* $\Gamma[\sigma^{-2} \mid s^2_{**}, n_{**}]$
4. *Repeat until convergence.*

Proof 1. *The proof can be found in Polasek et al. (2009) or in the appendix.*

2.2. Completing data by inter-prediction

We obtain the posterior predictive distribution in the following way, by integrating over the conditional predictive distribution for an unknown observation y_p with the posterior distribution $p(\beta, \rho, \sigma^{-2} \mid y_a)$ with the data $\mathcal{D} = (\mathbf{y}_a, X_d)$:

$$p(y_p \mid \mathcal{D}) = \int \int \int p(y_p \mid \beta, \sigma^{-2}) p(\beta, \rho, \sigma^{-2} \mid y_a) d\beta \; d\rho \; d\sigma^{-2},$$

where the posterior normal-gamma density $p(\beta_d, \rho_d, \sigma^{-2} \mid \mathbf{y}_a)$ is computed by a numerical procedure yielding the MCMC sample Θ_{MCMC} of size J of the θ_d parameters:

$$\Theta_{MCMC} = \left\{(\beta_j, \rho_j, \sigma_j^2), \quad j = 1, ..., J\right\}.$$

From this output we find a predictive sample of the unknown disaggregated vector y_d by drawing from the reduced form (which depends on the matrix W and on the known regressors X_d). This is the Chow-Lin formula for MCMC samples:

$$y_d^{(j)} \sim \mathcal{N}[R_j^{-1} X_d \beta_d^j + g_j, \sigma_j^2[(R_j' R_j)^{-1} - G_j]], \quad j = 1, ..., J, \tag{5}$$

using the spread matrix $R_j = I_n - \rho_j W$ for each j. g is the gain vector and G is the gain matrix for the mean and variance matrix, respectively, which are defined by

$$G_j = (R_j' R_j)^{-1} C' [C (R_j' R_j)^{-1} C']^{-1} C (R_j' R_j)^{-1}], \tag{6}$$

$$g_j = (R_j' R_j)^{-1} C' [C (R_j' R_j)^{-1} C']^{-1} (y_a - \hat{y}_{a,j})], \tag{7}$$

where we use the aggregated residuals $\hat{e}_a = y_a - \hat{y}_a$ and the current predictions $\hat{y}_{a,j} = R^{-1}_{a,j} X_a \beta_j$.

3. Completing data in spatial space-time panel (STP) models

We adopt the following notation: Let $Y_a : T \times N$ be the aggregate panel matrix for T aggregated time points and N aggregate cross-sectional units and $Y_d : m \times n$ be the disaggregate panel matrix. We assume that the aggregation has to be done in both dimensions, time and space:

$$Y_a = C_1 Y_d C_2'.$$

The time aggregation matrix is $C_1 : T \times m$ and the space aggregation matrix is $C_2 : N \times n$. Because aggregation in time always needs equal time periods of a basic period S called 'season', we have the integer equality for the dimensions

$$ST = m$$

since m is a T-multiple of the basic season S. Then the C_1 matrix can be written as a Kronecker product:

$$C_1 = I_T \otimes 1_m' = I_m \otimes 1_S,$$

where $1_m = (1, ..., 1)' : 1 \times m$ is a vector of m ones and 1_S is a S vector of ones . The spatial aggregation matrix faces irregularities and can be written as block diagonal matrix:

$$C_2 = diag(1_{n_1}, ..., 1_{n_N}) : N \times n \quad \text{with} \quad \sum_{i=1}^{N} n_i = n,$$

where the $n_i's$ are the lengths of the aggregates and $1_{n_i} : n_i \times 1$ is a column vector of ones and indexes the areas where n_i units are aggregated. For the space-time Chow-Lin procedure we have to vectorize the aggregation equation:

$$y_a = (C_2 \otimes C_1) vecY_d = C y_d$$

with the joint aggregation matrix $C = C_2 \otimes C_1$ and the vectorized data matrices $vecY_a = y_a$ and $vecY_d = y_d$. For a model with K regressors the indicator model we need K disaggregated panel matrices

$$X_k^d : n \times n, \quad \text{for} \quad k = 1, ..., K$$

as a 'panel indicators'[1] . A single vectorized panel indicator is just a $mn \times 1$ vector $vecX^d = x_d$ of regressors. Note that indicator matrices $\{X_k^d, k = 1, ..., K\}$ has to have the same dimension as Y_d. The disaggregated model for the missing y_d variable is a linear regression model involving all vectorized panel matrices. This leads to the following Chow-Lin model:

Definition 1 (The non-spatial Chow-Lin panel model). *For the K vectorized indicator matrices $\{X_k^d, k = 1, ..., K\}$ and the dependent panel variable Y_d we define the non-spatial Chow-Lin panel regression model:*

$$\begin{aligned} y_d &= X_d\beta_d + \varepsilon_d, \quad \varepsilon_d \sim \mathcal{N}[0, \Omega \otimes \sigma^2 V] \\ (mn \times 1) &= (mn \times K)(K \times 1) + (mn \times 1) \end{aligned} \tag{8}$$

where Ω is a $n \times n$ and V is a $m \times m$ correlation matrix for the time dimension. A simpler assumption is the homoskedastic case $\varepsilon_d \sim \mathcal{N}[0, \sigma^2 I_{nm}]$.

Note: The V matrix could be a time series covariance matrix assuming autoregressive errors for the error terms ε_{it} for all N cross sections in the panel $i = 1, ..., N$:

$$\varepsilon_{it} = \phi_i \varepsilon_{i,t-1} + u_{it}, \quad u_i \sim \mathcal{N}[0, \sigma_i^2 I_T], \quad i = 1, ..., N. \tag{9}$$

with the error vector $u_i = (u_{i1}, ..., u_{iT})'$ in each cross section i. Clearly a simplifying assumption is the homoskedastic case $\sigma_i^2 = \sigma^2$ and the 'homo-dynamic' case $\phi_i = \phi$.

The correlation matrix of the error term is then

$$V = \frac{\sigma_i^2}{1 - \phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & ... & ... & \phi^{T-1} \\ \phi & 1 & \phi & \phi^2 & ... & \phi^{T-2} \\ & & & ... & 1 & \phi \\ \cdots & & & & & \\ \phi^{T-1} & \phi^{T-2} & ... & & \phi & 1 \end{pmatrix} \tag{10}$$

and the inverse of this covariance matrix V is tri-diagonal

$$V^{-1} = \frac{1}{1 - \phi^2} \begin{pmatrix} 1 & -\phi & 0 & ... & ... & 0 \\ -\phi & 1 + \phi^2 & -\phi & 0 & ... & 0 \\ 0 & ... & -\phi & 1 + \phi^2 & -\phi & 0 \\ \cdots & & & & & \\ 0 & 0 & ... & 0 & -\phi & 1 \end{pmatrix}. \tag{11}$$

[1]The indices d and a are used as a sub- or a superscript, respectively, for convenience of notation.

3.1. Spatial space-time panel (STP) models

After having defined the ordinary space-time panel (STP) model in (1) we can turn now to the spatial autoregressive (SAR) extension if this model.

Definition 2 (The SAR space-time panel (STP) model). *The spatial extension of the space-time panel model (8) has the following disaggregated form*

$$\begin{aligned} y_d &= \rho_d (W \otimes I_n) y_d + X_d \beta_d + \varepsilon_d, \quad \varepsilon_d \sim \mathcal{N}[0, \Omega \otimes \sigma^2 V] && (12) \\ && (13) \\ (mn \times 1) &= (mn \times mn)(mn \times 1) + (mn \times K)(K \times 1) + (mn \times 1). \end{aligned}$$

We can write this model also as

$$y_d = \rho_d \tilde{y}_d + X_d \beta_d + \varepsilon_d,$$

since the spatial lag vector $\tilde{y}_d$ has the special form

$$\tilde{y}_d = vec(Y_d W') = (W \otimes I_m) y_d.$$

The spatial filter form of the SAR-STP model is given by

$$\tilde{R} y_d = X_d \beta_d + \varepsilon_d, \quad \varepsilon_d \sim \mathcal{N}[0, \Omega \otimes \sigma^2 V]. \tag{14}$$

with

$$\tilde{R} = (I_{nm} - W \otimes I_m) \rho_d = (I_n - \rho_d W) \otimes I_m = R \otimes I_m$$

and the spread matrix $R = I_n - \rho_d W$ as before.
The reduced form of the spatial filter model (14) is

$$y_d = \tilde{R}^{-1} X_d \beta_d + \tilde{R}^{-1} \varepsilon_d, \quad \tilde{R}^{-1} \varepsilon_d \sim \mathcal{N}[0, \Omega_\rho \otimes \sigma^2 V]. \tag{15}$$

which has a similar covariance structure since Ω_ρ is given by

$$\Omega_\rho = (R' \Omega^{-1} R)^{-1} : (n \times n). \tag{16}$$

For the estimation we need the aggregated reduced form of the spatial panel model

$$C y_d = C \tilde{R}^{-1} X_d \beta_d + C \tilde{R}^{-1} \varepsilon_d, \quad C \tilde{R}^{-1} \varepsilon_d \sim \mathcal{N}[0, C(\Omega_{\rho_d} \otimes \sigma^2 V) C']. \tag{17}$$

Since only the aggregated data are completely observed we have to derive the aggregated model from the disaggregated model. In compact notation the ARF of the spatial panel STP model (17) is:

$$y_a = X_{a\rho}\beta_d + \varepsilon_{a\rho}, \quad \varepsilon_{a\rho} \sim \mathcal{N}[0, C_2\Omega_\rho C_2' \otimes \sigma^2 C_2 V C_1']. \tag{18}$$

with $y_a = Cy_d$, $X_{a\rho} = C\tilde{R}^{-1}X_d$, and $\varepsilon_{a\rho} = C\tilde{R}^{-1}\varepsilon_d$. The covariance matrix of the ARF panel model is

$$V_{a\rho} = C_2\Omega_\rho C_2' \otimes C_1 V C_1' = \Omega_a \otimes V_a \tag{19}$$

with $\Omega_a = C_2\Omega_\rho C_2'$ and $V_a = C_1 V C_1'$ with Ω_ρ in (16) and V given in (10).

4. Least squares estimation for spatial space-time panel (STP) models

This augmented reduced form of the STP model (18) is the starting point for classical and Bayesian estimation procedures. The first method we discuss is simple least squares: Assume that there are K panel indicators $X_1^d, \ldots, X_K^d$ available at the disaggregated level, where the first one is a matrix of ones $X_1^d = 1_m \otimes 1_n'$, then we define the aggregated regressor matrix X_a consisting of vectorized panel data matrices:

$$X_a = (vecX_1, \ldots, vecX_K) : (TN \times K) \quad \text{and} \quad CX_d = X_a. \tag{20}$$

Again, the *aggregated model* is obtained by multiplying with the aggregation matrix C as in (18) but now the aggregation matrix is given by:

$$\begin{aligned} X_a &= (vecC_1 1_m 1_n' C_2', vec(C_1 X_2 C_2'), ..., vec(C_1 X_K C_2')) = \\ &= (vecX_{a_1}, vecX_{a_2}, ..., vecX_{a_K}) : mn \times K. \end{aligned}$$

The relationship between the disaggregated and the aggregated indicator matrix is: $X_k : (m \times n) \to X_k^a : (T \times N)$. The transposed matrix X_a' is given by

$$X_a' = \begin{pmatrix} vec' X_{a_1} \\ .. \\ vec' X_{a_K} \end{pmatrix} : K \times TN. \tag{21}$$

This leads to the following estimation procedure 1 (EP1): A 2-step estimate is given like in the simple spatial Chow-Lin model (see Polasek et al., 2009). First, we consider the usual GLS estimate $\beta_d^{GLS}:(K\times 1)$:

$$\beta_d^{GLS}=(X_a'V_{a\rho}^{-1}X_a)^{-1}(X_a'V_{a\rho}^{-1}y_a),$$

and then we get the "weighted" GLS estimate using the spatial lag $\tilde{y}_a=vecY_aW'$ as dependent variable:

$$\beta_{WLS}=(X_a'V_{a\rho}^{-1}X_a)^{-1}(X_a'V_{a\rho}^{-1})(R\otimes I_m)y_a.$$

Theorem 2 (GLS estimation in the space-time panel (STP)model).
The GLS and WLS estimates in the spatial STP model (12) can be found by the $K\times 1$ vector estimates

$$\beta_{GLS}=M_{\tilde{X}\tilde{X}}^{-1}M_{\tilde{X}\tilde{Y}} \tag{22}$$

$$\beta_{WLS}=M_{\tilde{X}\tilde{X}}^{-1}M_{\tilde{X}Y\tilde{W}'} \tag{23}$$

where $M_{\tilde{X}\tilde{X}}$ and $M_{\tilde{X}\tilde{Y}}$ are cross-moments matrices of the aggregated and transformed observations $\tilde{X}=(\Omega_a\otimes V_a)^{-1}X_a$, and $Y\tilde{W}'=(\Omega_a\otimes V_a)^{-1/2}Y_aW'$.

Proof 2. *The moment matrices can be simplified for computations:*

$$M_{\tilde{X}\tilde{X}}=\left(\begin{pmatrix} vec'X_{a_1}\\ \dots\\ vec'X_{a_K}\end{pmatrix}(\Omega_a\otimes V_a)^{-1}X_a\right)=$$

$$\begin{pmatrix} trX_{a1}'\Omega_a^{-1}X_{a1}V_a^{-1} & \dots & trX_{aK}'\Omega_a^{-1}X_{a1}V_a^{-1}\\ \dots & \dots & \\ trX_{aK}'\Omega_a^{-1}X_{a1}V_a^{-1} & \dots & trX_{aK}'\Omega_a^{-1}X_{aK}V_a^{-1}\end{pmatrix}$$

using the formula $tr\ ABCD=vec'D'(C'\otimes A)vecB$ and $vec'D=(vecD)'$ denotes the row vectorization. In the same way we find for the $(K\times K)$ cross-moment matrix

$$M_{\tilde{X}\tilde{Y}}=\left(X_a'(\Omega_a\otimes V_a)^{-1}Y_a\right)=\begin{pmatrix} trX_{a1}'\Omega_a^{-1}Y_aV_a^{-1}\\ \dots\\ trX_{aK}'\Omega_a^{-1}Y_aV_a^{-1}\end{pmatrix}$$

and

$$M_{\tilde{X}\tilde{Y}W'}=\left(X_a'(\Omega_a\otimes V_a)^{-1}Y_aW'\right)=\begin{pmatrix} trX_{a1}'\Omega_a^{-1}Y_aW'V_a^{-1}\\ \dots\\ trX_{aK}'\Omega_a^{-1}Y_aW'V_a^{-1}\end{pmatrix}$$

with the covariance matrices Ω_a and V_a given in (19). The minimum of the spatial ρ is found by minimizing the error sum of squares (ESS) over a grid of ρ_d values.

Based on these estimates we can propose the second stepwise estimation procedure (EP2): *The 3-step estimator*: In the first step we calculate the simplified GLS and WLS estimates of theorem 2 by using identity matrices for Ω_a and V_a by setting the spatial and the time correlation to zero: $\rho_d = 0$ and $\phi = 0$, in the covariance matrices (19). Then, in a second step, we get from the $T \times N$ residual matrix $\hat{E}_a$ (of the simplified GLS estimation with $\hat{e}_a = y_a - X_a\beta_{GLS}$) an estimate for $\hat{\Omega} = \hat{E}_a'\hat{E}_a/N$ and also an average ϕ autocorrelation coefficients from the N time series in the panel data set ($\hat{\phi} = \sum_{i=1}^N \hat{\phi}_i/N$), and as a third step we can estimate the spatial ρ_d again using the new ϕ estimate.

4.1. Chow-Lin GLS point prediction

The forecasting of the disaggregated observations has to be done by the general Goldberger (1962) formula (the subscript d is suppressed)

$$\hat{y} = X\beta_{GLS} + G\hat{e},$$

where the $G\hat{e}$ is an improvement of the estimated error term $\hat{e} = (y - X\beta_{GLS})$ using the 'Goldberger gain' matrix

$$G = V_{a\rho}^{-1}\tilde{C}'(\tilde{C}V_{a\rho}^{-1}\tilde{C}')^{-1}. \tag{24}$$

Next, the gain matrix can be partitioned using (19)

$$\begin{aligned} CV_{a\rho}^{-1}C' &= (C_2 \otimes C_1)\Omega_a^{-1} \otimes V_a^{-1}(C_2' \otimes C_1') = \\ &= C_2\Omega_a^{-1}C_2' \otimes C_1V_a^{-1}C_1'. \end{aligned} \tag{25}$$

Therefore the Goldberger gain matrix G in (24) can be simplified to $G = (G_2 \otimes G_1)$ with

$$G_2 = \Omega_a^{-1}C_1'(C_1\Omega_a^{-1}C_1')^{-1} \text{ and } \quad G_1 = V_a^{-1}C_2'(C_2V_a^{-1}C_2')^{-1}. \tag{26}$$

The gain-in-mean of the Chow-Lin prediction is

$$g = G\hat{e} = (G_2 \otimes G_1)vec\hat{E} = vecG_1\hat{E}G_2' \tag{27}$$

The Chow-Lin forecasts can be calculated as

$$\hat{y} = X\beta_{GLS} + vec\ G_1 \hat{E} G_2' \tag{28}$$

or with $vecB_{GLS} = \beta_{GLS}$ and $vecY_0 = X\beta_{GLS}$

$$\hat{Y} = Y_0 + G_1 \hat{E} G_1', \tag{29}$$

where we have used the vectorisation relation $\hat{e} = vec\hat{E}$.

The gain-in-variance or the improvement term for the variance (stemming from the conditional normal density formula) is

$$V_G = (G_2 C_1 \Omega_a^{-1} \otimes G_1 C_2 V_a^{-1})$$

5. MCMC in spatial space-time panel (STP) models

The Bayesian panel STP model estimation follows the same line as in the cross-section model and can be summarized by the MCMC procedure similar as in theorem (1). We consider the STP model with the prior distribution

$$p(\theta) = \prod_{i=1}^{n} \mathcal{N}[\beta_i^d \mid b_i^*, H_*] \mathrm{W}[\Omega \mid \Omega_*, \nu_*] \mathcal{U}_{-1,1}(\rho) \mathcal{U}_{-1,1}(\phi) \tag{30}$$

where $\mathcal{U}_{[-1,1]}(\rho_d)$ and $\mathcal{U}_{-1,1}(\phi)$ stands for a uniform distribution for the two correlation coefficients. Note that simpler formulas can be obtained if we assume a 'large' or '0-diffuse' prior for the betas:

$$p(\beta_d) = \mathcal{N}[\beta_d \mid 0, H_* = g\mathbf{I}_K] \tag{31}$$

centered at mean 0 and with the scalar g being large (e.g. $g = 10^3$ or 10^6). Then the posterior distribution of the parameter θ can be simulated using MCMC.

Theorem 3 (MCMC in the space-time panel (STP) model).
The MCMC procedure for the spatial STP model in (12) and the prior (30)

1. *Draw* β *from* $\mathcal{N}\left[\beta \mid b_{**}, H_{**}\right]$
2. *Draw* ρ *by a Metropolis step:* $\rho_{new} = \rho_{old} + \mathcal{N}[0, \tau_1^2]$
3. *Draw* ϕ *by a Metropolis step:* $\phi_{new} = \phi_{old} + \mathcal{N}[0, \tau_2^2]$
4. *Draw* σ^{-2} *from* $\Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}]$
5. *Repeat until convergence.*

Proof 3. *a) The fcd for the* β *regression coefficients is*

$$\begin{aligned} p(\beta \mid \mathcal{D}, \theta^c) &= \mathcal{N}[\beta \mid b_*, H_*] \cdot \mathcal{N}[Cy \mid CR^{-1}X\beta, \sigma^2 V_{a\rho}] \\ &= \mathcal{N}\left[\beta \mid b_{**}, H_{**}\right] \end{aligned}$$

with $\mathcal{D} = (y_a, X_d)$ *being the data base*[2] *for the STP model and the parameters*

$$\begin{aligned} H_{**}^{-1} &= H_*^{-1} + \sigma^{-2} X_d^{\top} R'^{-1} C' V_{a\rho}^{-1} C R^{-1} X_d, \\ b_{**} &= H_{**}[H_*^{-1} b_* + \sigma^{-2} X_d' \mathbf{R}'^{-1} C' V_{a\rho}^{-1} C y_d] \end{aligned}$$

using the covariance matrices in (19). These formulas can be written as

$$\begin{aligned} H_{**}^{-1} &= H_*^{-1} + \sigma^{-2} \mathbf{M}_{\tilde{X}_d R \tilde{X}_d}, \\ b_{**} &= H_{**}[H_*^{-1} b_* + \sigma^{-2} M_{\tilde{X}_d R \tilde{Y}_d}] \end{aligned}$$

where the moment matrices are given as before, but now the contain an additional spatial transformations by the inverse spread matrix R:

$$C(R^{-1} \otimes I_m) vec X_k^d = (C_2 \otimes C_1) vec X_k^d R'^{-1} = vec\ C_1 X_k^d R'^{-1} C_2' \text{ for } \quad k = 1, ..., K.$$

The regressor matrix now looks like

$$X_{a\rho} = (vec X_{a\rho,1}, vec X_{a\rho,2}, ..., vec X_{a\rho,K})$$

and the $K \times K$ *moment matrices are*

$$M_{\tilde{X} R \tilde{X}} = \left(\begin{pmatrix} vec' X_{a\rho,1} \\ ... \\ vec' X_{a\rho,K} \end{pmatrix} (\Omega_a \otimes V_a)^{-1} X_{a\rho} \right) =$$
$$\begin{pmatrix} tr\ X_{a\rho,1}' \Omega_a^{-1} X_{a\rho,1} V_a^{-1} & ... & tr\ X_{a\rho,K}' \Omega_a^{-1} X_{a\rho,1} V_a^{-1} \\ ... & ... & \\ tr\ X_{a\rho,K}' \Omega_a^{-1} X_{a1} V_a^{-1} & ... & tr\ X_{aK}' \Omega_a^{-1} X_{a\rho,K} V_a^{-1} \end{pmatrix}$$

and for the cross-product moments

$$M_{\tilde{X} R \tilde{Y}} = \left(X_{a\rho}' (\Omega_a \otimes V_a)^{-1} Y_a \right) = \begin{pmatrix} tr X_{a\rho,1}' \Omega_a^{-1} Y_a V_a^{-1} \\ ... \\ tr X_{a\rho,K}' \Omega_a^{-1} Y_a V_a^{-1} \end{pmatrix}.$$

[2] Note that W and C are strictly speaking also part of the data base

These matrices have usually small dimensions ($K \times K$), since the number of indicators is limited and can be easily built up by a loop in a computer program.

Note: If we assume a "0-diffuse" prior as in (31), then the posterior moments have simpler expressons

$$\begin{aligned}\mathbf{H}_{**}^{-1} &= g^{-1}I_K + \sigma^{-2}\mathbf{M}_{\tilde{X}R\tilde{X}},\\ \mathbf{b}_{**} &= \sigma^{-2}\mathbf{H}_{**}\mathbf{M}_{\tilde{X}R\tilde{Y}}.\end{aligned}$$

b) The fcd for the residual inverse variance is

$$p(\sigma^{-2} \mid \mathcal{D}, \theta^c) = \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}] \tag{32}$$

with $n_{**} = n_* + n$ *and*

$$s_{**}^2 n_{**} = s_*^2 n_* + ESS_\rho$$

where the error sum of squares ESS_ρ *is - using (19) - given by*

$$ESS_\rho = (y_a - CR^{-1}X_d\beta_d)'\mathbf{V}_{a\rho}^{-1}(y_a - CR^{-1}X_d\beta_d). \tag{33}$$

Using matrix notation we find

$$ESS_\rho = vec'\ E_a(\Omega_a \otimes V_a)^{-1}vecE_a = tr\ V_a^{-1}E_a\Omega_a^{-1}E_a' \tag{34}$$

with the residual matrix $E_a = (e_1, ..., e_N) : T \times N$ *defined by* $vecE_a = y_a - CR^{-1}X_d\beta_d$ *or* $e_i = y_{a,i} - X_{a\rho,i}\beta_{d,i}, i = 1, ..., N$ *with* $X_{a\rho} = CR^{-1}X_d$.

c) The fcd for the spatial rho

For the ρ_d *we use a Metropolis step*

$$\rho_{new} = \rho_{old} + \mathcal{N}[0, \tau^2]$$

with the acceptance ratio

$$\alpha = min\left[1, \frac{p(\rho_{new})}{p(\rho_{old})}\right],$$

where $p(\rho_d)$ *is the (kernel of) the full conditional for* ρ_d*, in our case the kernel is just stemming from the likelihood function:*

$$\begin{aligned}p(\rho_d) &= |\Omega_\rho|^{-\frac{T}{2}} exp\left(-\frac{1}{2\sigma^2}ESS_\rho\right)\\ &= |R\Omega^{-1}R|^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}ESS_\rho\right\}.\end{aligned}$$

From (19) we find $|\Omega_a \otimes V_a|^{-\frac{1}{2}} \propto |\Omega_a|^{-\frac{T}{2}} \propto |\Omega_\rho|^{-\frac{T}{2}}$ *and the error sum of squares* ESS_ρ *given in (33) contains* ρ.

d) The fcd for the correlation parameter ϕ

For the ϕ *we use a Metropolis step:*

$$\phi_{new} = \phi_{old} + \mathcal{N}[0, \tau_2^2]$$

with the acceptance ratio

$$\alpha = min\left[1, \frac{p(\phi_{new})}{p(\phi_{old})}\right];$$

where $p(\phi)$ *consists of the (kernel of) the full conditional distribution for* ϕ*, in our case the kernel is just stemming from the likelihood function:*

$$p(\phi) = |\mathrm{V}|^{-\frac{N}{2}} exp\left(-\frac{1}{2\sigma^2}ESS_\phi\right).$$

From (19) we find

$$|\Omega_a \otimes V_a|^{-\frac{1}{2}} \propto |V_a|^{-\frac{N}{2}} \propto |\mathbf{V}|^{-\frac{N}{2}}$$

and $ESS_\phi = ESS_\rho$ *given in (33) contains* ϕ.

e) The fcd for the inverse SUR covariance matrix Ω^{-1}

The SUR precision matrix

$$p(\Omega^{-1} \mid \mathcal{D}, \theta^c) = \mathrm{W}_m[\Omega^{-1} \mid \Omega_{**}, \nu_{**}] \tag{35}$$

follows a (m-dim.) Wishart distribution with $\nu_{**} = \nu_* + 2$ *and*

$$\Omega_{**} = \Omega_* + CU_\rho U_\rho' C' \tag{36}$$

where U_ρ *is the residual matrix, constructed from the vectorized residuals* $vecU_\rho = y_a - R^{-1}X_a\beta_a$. *There is no closed form expression possible, since the inverse spread matrix* R^{-1} *destroys the Kronecker product structure of the multivariate equation.*

5.1. Completing data by prediction: inter-prediction

We obtain the posterior predictive distribution in the following way, by integrating over the conditional predictive distribution with the posterior distribution:

$$p(y_p \mid \mathcal{D}) = \int \int \int p(y_p \mid \beta, \sigma^{-2}) p(\beta, \rho, \sigma^{-2} \mid \mathcal{D}) d\beta \; d\rho \; d\sigma^{-2}$$

where the posterior normal-gamma density $p(\beta_d, \rho_d, \sigma^{-2} \mid \mathcal{D})$ is given numerically by a MCMC sample, i.e. a posterior sample of the θ parameters of the STP model:

$$\Theta_{MCMC} = \{(\beta_j, \rho_j, \phi_j, \sigma_j^2, \Omega_j), \quad j = 1, ..., J\}.$$

From this output we find a predictive sample of the unknown vector y_d by drawing from the reduced form in (15), which depends on the matrix W and on the known regressors X_d:

$$\{y^{(j)} \sim \mathcal{N}[R_j^{-1}X\beta_j + g_j, (R_j'\Omega_j^{-1}R_j)^{-1} \otimes \sigma_j^2 V_j - G_j]\} \tag{37}$$

where g_j is given by (27), G_j is given as in (6) and the spread $R_j = I_n - \rho_j W$ computed for all MCMC draws $j = 1, ..., J$ in Θ.

6. The Bayesian Chow-Lin model for completing panel data

We consider a panel spatial autoregressive model as in (12)

$$y_d = \rho_d W y_d + X_d \beta_d + \varepsilon_d, \quad \varepsilon_d \sim \mathcal{N}[0, \Omega \otimes \sigma^2 I_n]$$

with the residuals $\varepsilon_d = vec\ E_d$ from the stacked residual matrix $E_d : m \times n$. The prior information for the parameters $\theta = (\rho, \beta, \sigma^2, \Omega)$ is blockwise independent

$$\begin{aligned} p(\theta) &= p(\rho)p(\beta)p(\sigma^2)p(\Omega) \quad with & (38) \\ p(\rho) &= U(-1,1); \quad p(\beta) = \mathcal{N}[\beta_*, H_*]; \\ p(\sigma^{-2}) &= \Gamma[s_*^2, n_*], \quad p(\Omega^{-1}) = \mathrm{W}[\Omega_*, \nu_*], & (39) \end{aligned}$$

where U is a uniform, W a Wishart and Γ a Gamma-2 distribution.

Consider the SAR panel Chow-Lin model (in short SAR-PCL) and let us denote the 3 conditional distributions by $p(\rho \mid \theta^c), p(\beta \mid \theta^c)$, and $p(\sigma^{-2} \mid \theta^c)$ where θ^c denotes the complementary parameters for the f.c.d.'s, respectively.

The MCMC procedure for the panel Chow-Lin model (SAR-PCL) consists of 4 blocks of sampling, as given in the next theorem:

Theorem 4 (MCMC in the SAR-PCL model).
The MCMC estimation for the SAR-PCL model (12) with prior (38) involves the following iterations:

Step 1. *Draw* β *from* $\mathcal{N}[\beta \mid \lfloor_{**}, \mathcal{H}_{**}]$
Step 2. *Draw* ρ *by a Metropolis step:* $\rho_{new} = \rho_{old} + \mathcal{N}[0, \tau^2]$
Step 3. *Draw* σ^{-2} *from* $\Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}]$
Step 4. *Draw* Ω^{-1} *from* $\mathrm{W}[\Omega^{-1} \mid \Omega_{**}, \nu_{**}]$
Step 5. *Repeat until convergence.*

Proof 4 (Proof of Theorem 4). *The first three fcd's are the same as in Theorem (1). We now show that the fcd for the* Ω^{-1} *is derived in the following way. Recall that the reduced form of the panel SAR model is given by*

$$y \sim \mathcal{N}[R^{-1}X\beta, \Omega \otimes \sigma^2 (R'R)^{-1}]. \tag{40}$$

This leads to the likelihood function

$$p(\Omega^{-1} \mid y) = |\Omega \otimes \sigma^2 (R'R)^{-1}|^{-1/2} exp \left\{ -\frac{1}{2\sigma^2} e_d' (\Omega \otimes \sigma^2 (R'R)^{-1})^{-1} e_d \right\}, \tag{41}$$

with $e_a = y_a - R^{-1}X_a\beta_d = vec(E_a)$ *the vectorization of the residual matrix* $E_a : T \times N$*. This can be written in compact form*

$$p(\Omega^{-1} \mid y) = |\Omega|^{-n/2}\sigma^{-nT}|R| \; exp\left\{-\frac{1}{2\sigma^2}tr \; E_a\Omega E_a'(R'R)^{-1}\right\}. \tag{42}$$

Now this expression has to be combined with the kernel of the prior distribution

$$p(\Omega^{-1} \mid y) \propto |\Omega|^{-\nu_*/2} \; exp\left\{-\frac{1}{2\sigma^2}tr \; \Omega_*\Omega\right\} = \mathrm{W}[\Omega_{**}, \nu_{**}]. \tag{43}$$

and yields a Wishart distribution with $\nu_{**} = \nu_* + n$ *and* $\Omega_{**} = \Omega_* + E'(R'R)^{-1}E$*.*

6.1. Completing data by Chow-Lin prediction: Inter-prediction

We obtain the posterior predictive distribution in the same way as before: Using the above MCMC procedure we obtain a posterior sample of the θ parameters: $\Theta_{MCMC} = \{(\beta_j, \rho_j, \sigma_j^2, \Omega_j), \quad j = 1, ..., J\}$. Again, from this MCMC output we find a predictive sample $y^{(j)}$ by drawing from the reduced form (which depends on the matrix W and on the known disaggregated indicators (regressors) X_d):

$$\{y^{(j)} \sim \mathcal{N}[R_j^{-1}X\beta_j + g_j, \Omega_j \otimes \sigma_j^2[(R_j'R_j)^{-1} - G_j]], \quad j = 1, ..., J\}$$

computed as $R_j = (I_n - \rho_j W)$ for each j. g_j is the 'gain-in-mean vector' for the mean and G_j is the 'gain-in-variance' matrix, respectively, which are defined by

$$G_j = (R_j'R_j)^{-1}C'[C(R_j'R_j)^{-1}C']^{-1}C(R_j'R_j)^{-1}], \tag{44}$$

$$g_j = (R_j'R_j)^{-1}C'[C(R_j'R_j)^{-1}C']^{-1}(y_a - \hat{y}_{a,j})], \tag{45}$$

where we use the aggregated residuals $\hat{e}_a = y_a - \hat{y}_a$ and the current predictions $\hat{y}_{a,j} = R_{a,j}^{-1}X_a\beta_j$ for each j.

7. Application of the spatial Chow-Lin to Spanish regions

In this section, the performance of the classical and Bayesian Chow-Lin method is evaluated using actual data for the Spanish GDP at NUTS-2 and NUTS-3 level for the period 2000-2004[3]. Spain has 18 regions (NUTS-2) and 52 provinces (NUTS-3). The associated C matrix is constructed from the hierarchical structure of the NUTS-3 regions embedded in NUTS-2 regions. Note that, in contrast to the temporal Chow-Lin method where each aggregated period (year) has the same number of disaggregated seasons (4 quarters,

[3]All data and the aggregation matrix C for Spanish provinces are available from the authors upon request.

12 months etc.), in the spatial framework the number of provinces (NUTS-3) varies for each region (NUTS-2). In Spain, the number of provinces by regions range between one and nine, and seven regions are single unit regions, having just one province. In such provinces no aggregation is possible (and this fact has to be taken into account in the evaluation procedure of Chow-Lin methods). This heterogeneity in terms of size and administrative structure makes Spanish regions a real challenge and a good testing ground for spatial Chow-Lin methods.

7.1. The Spanish sub-national data

The regressors used for the aggregate model are described in Table 1. Note that the indicators should be available at the NUTS-2 and NUTS-3 level. Usually, due to the data limitation problems described above, the number and quality of indicators available at this spatial level is lower than for the NUTS-2 level. However, in the Spanish case it is possible to obtain some reliable indicator variables that are able to proxy the GDP by the demand and supply side. All regressors enter in log levels to explain GDP (for NUTS-2) for the year 2004 (or the years 2000-2004 in the panel case). The NUTS-2 GDP series were calculated by aggregating NUTS-3 GDP. Therefore, it is possible to compare the Chow-Lin predicted values with the actual data available. As a spatial weight matrix W we use the row normalized matrix for the inverse distances between the NUTS-3 provinces.

The first aim is to find an appropriate aggregated SAR model, using different indicator variables, which should be correlated with the 'GDP', both at the regional and provincial level. Table 2 shows the results obtained for the two best models[4] , using the SAR program of LeSage (1997). In these two models the spatial term ρ is positive and significant. Based on the best cross-sectional models in Polasek et al. (2009), our first model consists of three the variables 'Employment', international 'Exports' and 'Imports' that are able to explain by a $R^2 = 99.96\%$ of the spatial distribution of the 'GDP' in the cross-sectional models.

The second model estimated for the spatial panel data set includes an agglomeration dummy variable that takes the value 1 for Madrid and Barcelona (Mad_Bar), and 0 otherwise. Although the R^2 is slightly lower than for the cross-section models (see Polasek et al., 2009), the level of significance for all the variables increases as well as the importance of the spatial effect, whose positive coefficients vary from 0.12 to 0.14. The 'Mad_Bar' variable shows negative coefficients with acceptable significance levels, pointing out to higher levels of concentration in employment and international trade in Madrid and Barcelona than in terms of 'GDP'. Probably this result is connected with differences in productivity (GDP/employment ratios by regions) and

[4]Due to space limitations, we omit the results for variables like 'capital-stock', 'number of trucks' and 'number of banks', which did not improve the results.

the higher concentration of traders and headquarters in these two regions, which tends to overvalue their amount of imports and exports.

7.2. Evaluation of the spatial Chow-Lin method

The evaluation of the spatial Chow-Lin (CL) follows the evaluation methods for predictions in statistical models. This follows from the fact that unknown y's have to be predicted while the predictors are fully observed. In the Spanish case we are in the fortunate position of knowing the disaggregated y values, so we can compute the prediction accuracy. This is done for the classical and Bayesian prediction as well as for the method with and without the *Gain* (see equation 24) term. After that we compute some forecast criteria to evaluate the four different predictions. To evaluate the accuracy of the ML and Bayesian prediction we chose three criteria from the forecasting literature (see e.g. Chatfield, 2001): the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE)[5].

The results are shown in Table 3. According to the three criteria (RMSE, MAE and MAPE), the rankings of the models are the same. Moreover, the forecasts including the 'gain term', which is a function of the spatial autocorrelation, always outperform the equivalent methods 'without the gain'. According to these rankings, the best method is the Bayesian model 'with gain'. This shows that a spatial model will considerable improve the Chow-Lin forecasts for disaggregate data, while ignoring the spatial correlation - i.e. applying a conventional regression model instead - will lead to a considerable accuracy loss for the predicted data. Finally, to visualize the comparisons, Figures 2 and 3 show overlay plots of the classical and Bayesian Chow-Lin predictions for Model 1, with and without gain, together with the observed data, using the panel data specification. Figure 3 shows clearly that the Bayesian spatial Chow-Lin forecasts lie closer to the observed values than classical predictions or non-spatial methods (denoted as 'no gain') in Figure 2.

8. Conclusions

Regional econometric work in Europe has become increasingly important, especially since the integration process of the European Union puts a lot of weight on policies for regional coherence. For such evaluations NUTS data are the main source of information. They are collected by Eurostat and the individual member states using common rules and methods. But not all member states have developed the same level of skills, especially since 1995 after the harmonized European national accounting system has started. This leads to

[5]The formulas are $RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(y-\widehat{y})^2}$, $MAE = \frac{1}{N}\sum_{i=1}^{N}|y-\widehat{y}|$ and $MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{y-\widehat{y}}{y}|$ respectively.

inhomogeneous data quality and sometimes to holes in the database if smaller regional units are needed. In order to apply many modern panel methods one has to complete such data sets. While the simplest method is interpolation, this gives not always satisfactory results.

Based on the well known Chow-Lin method of temporal interpolation, and a recent spatial extension (see Polasek et al., 2009), we develop a new spatial interpolation procedure for panel data in this paper. The procedure uses the indicators at the disaggregated regional level to predict the disaggregated unobserved dependent variable, conditional on the complete aggregated observed model. We propose a spatial estimation procedure in a classical or Bayesian framework, where the latter is done by MCMC.

To evaluate the new method, we forecasted the GDP for the 52 Spanish provinces (at NUTS-3 level), but based only on the information for the 18 Spanish regions (i.e. NUTS-2 GDP as dependent variable), while the forecasts are based on high frequency socio-economic indicators at the NUTS-3 level. Then, to compare the results obtained with the actual series available at the NUTS-3 level, we computed forecast criteria. We point out that a significant spatial lag parameter leads to an improvement (through the so called gain term) in the spatial Chow-Lin prediction of the disaggregated data. The Bayesian MCMC method yield the best result among the models in the forecast experiment. Our new method has shown that it pays to get a good spatial model if one is interested in good predictions of missing data in a cross-sectional or panel model. A non-trivial condition for finding a good model is the existence of good indicators, the removal of outliers and the skill to find the appropriate weight matrix to estimate the spatial effects. In future research we will explore these modeling possibilities in more detail, and we extend the spatial Chow-Lin method to complete large blocks of data at the national and European level, including flow data such as inter-regional trade or migration flows.

Amrhein, C. G., Wong, D., 1996. Research on the maup: Old wine in a new bottle or real breakthrough? Geographical Systems 3, 73–76.

Baltagi, B. H., Li, D., 2006. Prediction in the panel data model with spatial correlation: the case of liquor. Spatial Economic Analysis 1 (2), 175–185.

Chatfield, C., 2001. Time-series Forecasting. Chapman & Hall.

Chow, G. C., Lin, A., 1971. Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. The Review of Economics and Statistics 53 (4), 372–375.

Fotheringham, A., Wong, D., 1991. The modifiable areal unit problem in statistical analysis. Environment and Planning A 23, 1025–1044.

Fotheringham, A. S., Brunsdon, C., Charlton, M., 2001. Scale issues and geographically weighted regression. In: Tate, N. J., Atkinson, P. M. (Eds.), Modelling Scale in Geographical Information Science. New York: Wiley and Sons, pp. 123–40.

Goldberger, A. S., 1962. Best linear unbiased prediction in the generalized linear regression model. American Statistical Association Journal 57, 369–375.

Holt, D., Steel, D., Tranmer, M., 1996. Area homogeneity and the modifiable areal unit problem. Geographical Systems 3, 181–200.

Jelinski, D. E., Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology.

King, G., 1995. A Solution to the Ecological Problem: Reconstructing Individual Behavior from Aggregate Data. N. J.: Princeton University Press., London, UK.

LeSage, J. P., 1997. Bayesian estimation of spatial autoregressive models. International Regional Science Review 20, 113–129.

Openshaw, S., 1984. The modifiable areal unit problem. CATMOG 38. GeoBooks, Norwich, England.

Openshaw, S., Taylor, P., 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley, N. (Ed.), Statistical Applications in the Spatial Sciences. Pion, London, pp. 127–144.

Openshaw, S., Taylor, P., 1981. The modifiable areal unit problem. In: Wrigley, N., Bennett, R. (Eds.), Quantitative Geography: A British View. Routledge and Kegan Paul, London., pp. 60–69.

Pavía, J. M., Larraz, B., Montero, J. M., 2008. Election forecasts using spatiotemporal models. Journal of the American Statistical Association 103 (483), 1050–1059.

Polasek, W., Sellner, R., Llano, C., 2009. Bayesian methods for completing data in spatial models, forthcoming.

Sui, D., 2000. New directions in ecological inference: An introduction. Annals of the Association of American Geographers 90 (3), 579–582.

Wong, D. W. S., 2001. Location-specific cumulative distribution function (lscdf): An alternative to spatial correlation analysis. Geographical Analysis 33, 76–93.

Wong, D. W. S., July 2003. Spatial decomposition of segregation indices: A framework toward measuring segregation at multiple levels. Geographical Analysis 35 (3), 179–194.

Wu, C., Murray, A. T., 2005. A cokriging method for estimating population density in urban areas. Computers, Environment and Urban Systems. 29 (5), 558–579.

Annex 1: Tables and Figures

Table 1: Description and source of the variables in the database

Variable	Description	Source
Area	Area of provinces in square km	INE[a]
Pop	Population by provinces in 1,000	INE
Emp	Employment by provinces in 1,000	INE
Kstock	Capital stock by provinces	FBBVA-IVIE[b]
Export	International exports of goods by provinces	AEAT[c]
Import	International imports of goods by provinces	AEAT
Vat	Value Added Tax revenue by provinces	AEAT
IncTax	Income tax revenue by provinces	AEAT
Income	IncTax by provinces per capita	Own calc.- INE
Trucks	Number of heavy trucks by provinces	La Caixa[d]
Banks	Number of banks in each province	La Caixa
Mad_Bar	Dummy for Madrid and Barcelona	Own calc.
Capi	Dummy for Madrid only	Own calc.
Caprov	Dummy: 1 for all capital provinces	Own calc.
Rforal	Dummy: 1 for provinces with special tax system	Own calc.

[a] www.ine.es
[b] www.fbbva.es,www.ivie.es
[c] www.aeat.es
[d] www.lacaixa.es

Table 2: Panel data SAR models: GLS and Bayesian estimates for GDP, 2000-2004

Models	Model 1		Model 2	
Estimation	Classic	Bayesian	Classic	Bayesian
R-squared	0.9995	0.9995	0.9996	0.9995
Rbar-squared	0.9995	0.9995	0.9995	0.9995
σ^2	0.2073		0.1782	
sige, ESS/(n-k)		0.2230		0.2003
ndraws,nomit		500,50		500,50
Nobs, Nvars	90, 4	90, 4	90, 5	90, 5
log-likelihood	-25.7614		-18.9705	
coefficients[a]				
constant	-3.7695	-3.4991	-4.1362	-4.0264
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
log(Emp)	0.4193	0.4066	0.4516	0.4873
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
log(Exports)	0.2392	0.2414	0.2321	0.2208
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
log(Imports)	0.2653	0.2662	0.2611	0.2576
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Mad_Bar			-0.5765	-0.6526
			(0.0001)	(0.0000)
ρ	0.1299	0.1223	0.1449	0.1443
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

[a]z-probabilities in parentheses

Table 3: Chow-Lin Prediction Accuracy: Classical vs. Bayesian estimates

			RMSE[a]	MAE[b]	MAPE[c]
Panel-data	Classical	gain	3.166	0.348	3.146
		no gain	3.209	0.352	3.187
	Bayesian	gain	0.822	*0.067*[d]	0.621
		no gain	3.100	0.340	3.078

[a]Root Mean Squared Error
[b]Mean Absolute Error
[c]Mean Absolute Percentage Error
[d]Minimum

Figure 1: Geographical distribution of GDP 2004 for the Spanish provinces (NUTS-3)

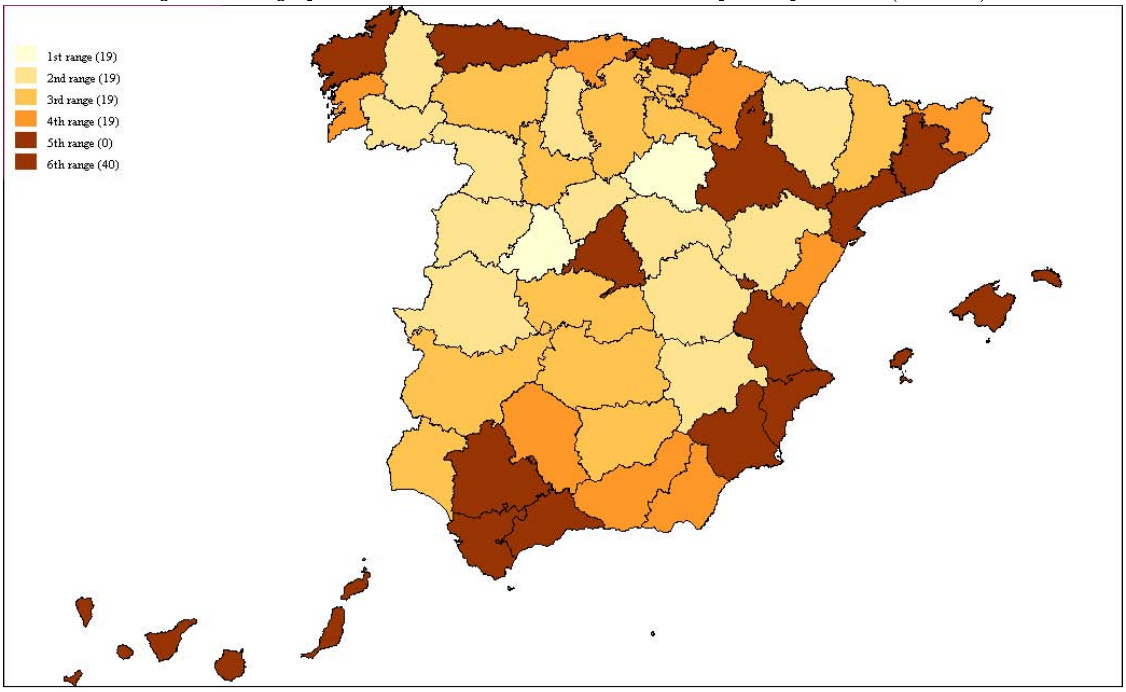


Figure 2: Overlay Comparison: Classical panel-data predictions with and without gain across NUTS-3 regions

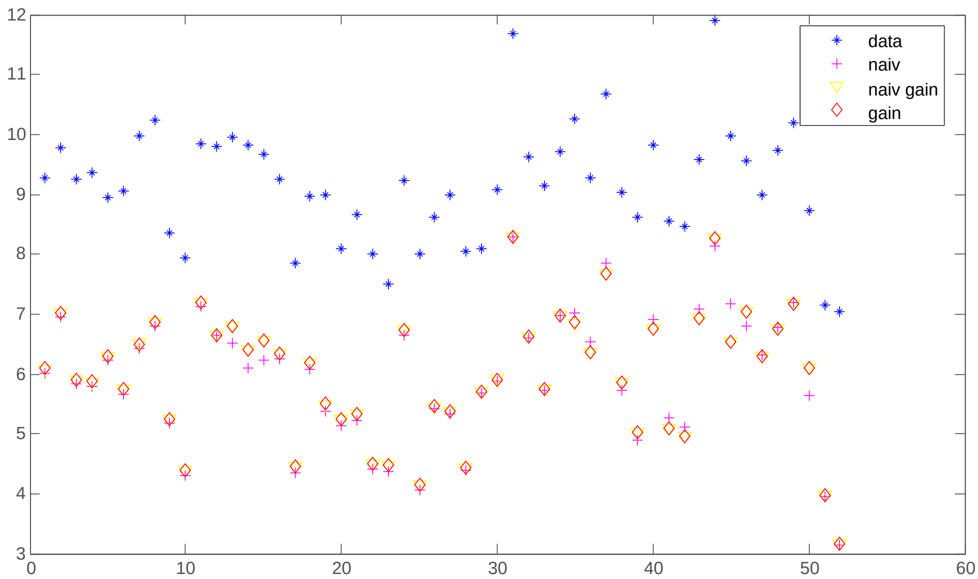
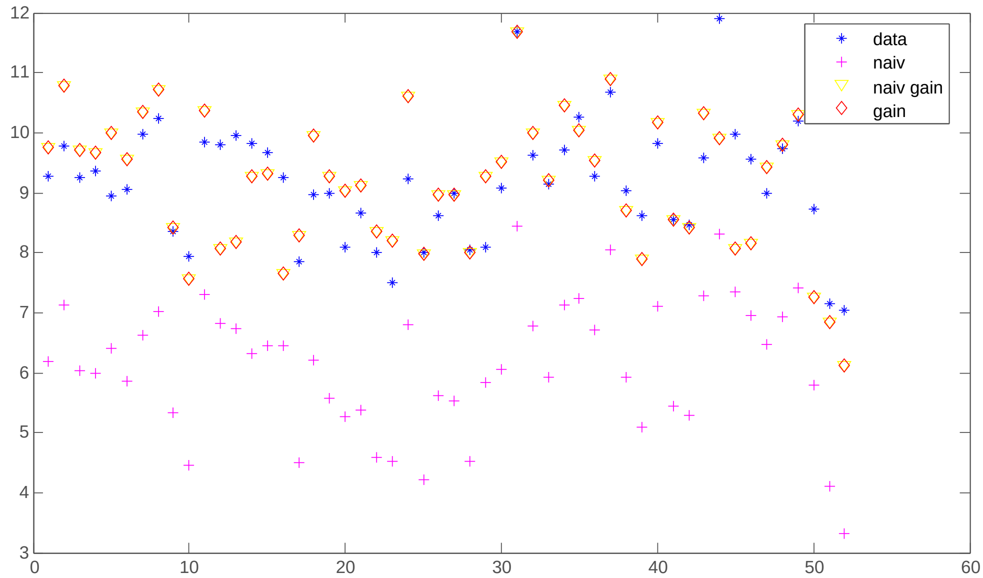


Figure 3: Overlay Comparison: Bayesian panel-data predictions with and without gain across NUTS-3 regions



Annex 2: Proofs

Proof of theorem (1):

a) The fcd for the beta regression coefficients y_d is

$$p(\beta_d \mid y_d, \theta^c) = \mathcal{N}[\beta \mid b_*, H_*] \cdot \mathcal{N}[Cy \mid CR^{-1}X\beta, \sigma^2 C(R'R)^{-1}C'] \tag{46}$$

$$= \mathcal{N}[\beta \mid b_{**}, H_{**}] \tag{47}$$

with the parameters

$$H_{**}^{-1} = H_*^{-1}b_* + \sigma^{-2}X^\top R'^{-1}C'\Omega(\rho)^{-1}CR^{-1}X,$$

$$b_{**} = H_{**}[H_*^{-1}b_* + \sigma^{-2}X'R'^{-1}C'\Omega(\rho)^{-1}Cy]$$

b) The fcd for the residual inverse variance we find

$$p(\sigma^{-2} \mid y, \theta^c) = \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}] \tag{48}$$

with $n_{**} = n_* + n$ and

$$s_{**}^2 n_{**} = s_*^2 n_* + ESS_\rho$$

where the error sum of squares ESS_ρ is given by

$$ESS_\rho = (Cy_d - CR^{-1}X_d\beta_d)'\Omega(\rho)^{-1}(Cy_d - CR^{-1}X_d\beta_d). \tag{49}$$

c) The fcd for the spatial correlation rho

For the ρ we use a Metropolis step:

$$\rho_{new} = \rho_{old} + \mathcal{N}(0, \tau^2)$$

with the acceptance ratio

$$\alpha = min\left[1, \frac{p(\rho_{new})}{p(\rho_{old})}\right],$$

where $p(\rho)$ is the (kernel of) the full conditional for ρ, in our case the kernel is just stemming from the likelihood function:

$$p(\rho) = |\Omega(\rho)|^{-\frac{1}{2}} exp(-\frac{1}{\sigma^2} ESS_\rho)$$

with ESS_ρ given in (49).