IHS Economics Series
Working Paper 295
April 2013

Spatial Chow-Lin Models for Completing Growth Rates in Cross-sections

Wolfgang Polasek

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Impressum

Author(s):
Wolfgang Polasek

Title:
Spatial Chow-Lin Models for Completing Growth Rates in Cross-sections

ISSN: Unspecified

295

Reihe Ökonomie

Economics Series

Spatial Chow-Lin Models for Completing Growth Rates in Cross-sections

Wolfgang Polasek

295

Reihe Ökonomie

Economics Series

Spatial Chow-Lin Models for Completing Growth Rates in Cross-sections

Wolfgang Polasek

April 2013

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Wolfgang Polasek
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, AUSTRIA
☎: +43/1/599 91-155
email: polasek@ihs.ac.at
and
University of Porto
Rua Campo Alegre
Portugal

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share "work in progress" in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschafts-wissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

Abstract

Growth rate data that are collected incompletely in cross-sections is a quite frequent problem. Chow and Lin (1971) have developed a method for predicting unobserved disaggregated time series and we propose an extension of the procedure for completing cross-sectional growth rates similar to the spatial Chow-Lin method of Liano et al. (2009). Disaggregated growth rates cannot be predicted directly and requires a system estimation of two Chow-Lin prediction models, where we compare classical and Bayesian estimation and prediction methods. We demonstrate the procedure for Spanish regional GDP growth rates between 2000 and 2004 at a NUTS-3 level. We evaluate the growth rate forecasts by accuracy criteria, because for the Spanish data-set we can compare the predicted with the observed values.

Keywords

Interpolation, missing disaggregated values in spatial econometrics, MCMC, Spatial Chow-Lin methods, predicting growth rates data, spatial autoregression (SAR), forecast evaluation, outliers

JEL Classification

C11, C15, C52, E17, R12

Comments
Preprint submitted to Elsevier.

Contents

1. Introduction

Completing data sets at a disaggregated level when only aggregated values can be observed can be done by the Chow and Lin (1971) method. The original method was proposed for time series, but in Polasek and Sellner (2010) this method was extended for cross-sectional data based on a spatial autoregressive model, for panel data and for spatial flow models. An implicit assumption of the Chow-Lin approach is the summability of disaggregated variables to aggregated variables, a property that holds for so-called intensive variables. This paper shows how to extend the spatial Chow-Lin approach for cross-sectional data to non-extensive or intensive variables, like growth rates. In Physics "an extensive variable is one that is additive for independent, non-interacting subsystems" (Wikipedia, Feb. 6th, 2013, $http://en.wikipedia.org/wiki/Extensive_quantity$).

Many data are collected by Eurostat via the individual EU member states using common rules and methods. But not all member states have started at the same time their data collection, and therefore data series are often incomplete. In 1995 Eurostat introduced the harmonized European national accounting system. This leads to inhomogeneous data quality and sometimes to holes in the database if smaller regional units are needed. In order to apply many modern panel estimation methods one has to complete such data sets. While the simplest (deterministic) method to repair data holes is deterministic interpolation, this does not always give satisfactory results, and we prefer to use model based stochastic (imputation) methods for the missing disaggregated values.

For spatial data, this paper focuses on completing data sets that are growth rates or are extensive (i.e. summable) cross-sectional variables and we discuss two extensions of the Chow and Lin (1971) method: We will use spatial econometrics (see e.g. Anselin (1988)) and also the Bayesian MCMC approach as e.g. in LeSage and Pace (2009).

The paper is organized as follows. Section 2 outlines the classical estimation and prediction in the spatial and non-spatial system Chow-Lin (CL) model. The classical (BLUE) estimator for the spatial autoregressive model (SAR) is derived, 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. Section 4 describes the Bayesian approach for the spatial system Chow-Lin method together with the MCMC algorithms and we show how the numerical predictive densities for the missing disaggregated values can be obtained by simulating from the conditional density in the system approach. Furthermore we show that the method can also be used in the presence of outliers in the system Chow-Lin model. An example for completing the growth rates are given in section 5. We apply the spatial Chow-Lin method to Spanish NUTS-2 and NUTS-3 data. Because for Spain we can observe all data on the disaggregated level, we will evaluate the

quality of the spatial Chow-Lin method by comparing the predicted values for GDP at the NUTS-3 level to their observed values and calculate the usual forecast accuracy criteria. A final section concludes.

1.1. Eurostat and the regional data base for Europe

Eurostat publishes regional data on a range of different statistical topics, collected by the 27 member states, but also from candidate countries and by the four EFTA states. Usually, this information is collected at different spatial levels based on the nomenclature of territorial units for statistics (NUTS).
NUTS data are collected by the individual member states using common rules and methods. However, not all member states have developed the same level and speed of skills, especially after 1995 when the harmonized European economic account system started. This can lead to inhomogeneous data quality and sometimes to holes in the data base, especially if it comes to smaller regional units where never had been data collected before. Thus, although in 2003 the NUTS system was acquired as a basis for a regional EU data base, it is common to find that the data at the lowest levels of disaggregation (NUTS-3) is missing for some countries and indicators. Moreover, periodical changes in the NUTS regulation occur since the regional classification adapts to the new administrative boundaries or economic circumstances. Consequently, these changes lead to additional disconnections in the time series, which can lead to breaks in the information at the lowest spatial units under consideration.

Sometimes it is difficult to obtain a complete set of panel data of all EU regions at the NUTS-3 level covering even the most basic indicators referred to demographics, labor markets, infrastructure, prices or productivity. For example, if one downloads the Eurostat information for regional GDP at the NUTS-3 level for the EU 27, including EFTA countries and EU candidate countries for the period 1995-2005, one would find that 15% of the numbers are missing. On top of that, the problems of data restriction at the NUTS-3 level increases for more disaggregated components of the regional accounts, either from the supply (Gross Value Added by industries), the demand (investments, public or public expenses) or the income side (salaries or capital remuneration). Finally, as it has been described above, it could also be the case that the right spatial level for analyzing a specific economic phenomenon requires the use of data even at a lower level of aggregation as the presently available NUTS-3 data.

LeSage and Pace (2004) use spatial econometric techniques to estimate missing dependent data. They predict unobserved house prices by using the information of sold and unsold houses to increase the estimation efficiency. LeSage and Pace (2004) predict unobserved spatially dependent data with observable data at the same regional level. The goal is to predict unobserved dependent variables for all regions.

2. The Chow-Lin method for non-summable (intensive) variables

Chow-Lin (1971) developed a method to forecast ("construct") quarterly times series observations from yearly observations, by using appropriate "indicators" or auxiliary regressors for the quarterly series. This approach can be extended for constructing disaggregated observations in the spatial context if only aggregated observations are available again using indicator variables in the forecasting equation, as it was shown in Polasek and Sellner (2010). As a basis for the subsequent analysis we first review the Chow-Lin method as proposed in Llano et al. (2009).

2.1. The basic Chow-Lin method

Disaggregate (or high frequency) time series are occasionally needed since they offer valuable information for policy makers. However, such data on a monthly or quarterly basis are often not available- for various reasons. Attempts have been made to interpolate missing high frequency data by using related series that are known. Friedman (1962) suggested relating the series in a linear regression framework. The three problems in connection of missing data are known by statisticians as interpolation, extrapolation and the distributional problem of time series by related series. Interpolation is used to generate higher frequency level (or stock) data, while extrapolation extends a given series outside the sample period, and in the distribution framework one allocates lower frequency flow data, such as GDP (see Fernandez, 1981), to higher frequency observations. The path-breaking paper by Chow and Lin (1971) embedded the missing data problem to a predictive system framework of aggregate and disaggregate data, leading to a boost in research on this topic.

We assume a linear relationship for the high frequency (disaggregate) data y_d and the indicators X_d, i.e.

$$y_d = X_d\beta_d + \varepsilon_d \quad with \quad \varepsilon_d \sim \mathcal{N}[0, \sigma^2\Omega], \tag{1}$$

where y_d is a $(n \times 1)$ vector of unobserved disaggregate variables, but X_d is a $(n \times k)$ matrix of observed regressors. β_d is a $(k \times 1)$ vector of regression coefficients, and ε_d is a vector of random disturbances, with mean $\mathbb{E}(\varepsilon) = 0$ and covariance matrix $\mathbb{E}(\varepsilon_d\varepsilon_d') = \sigma^2\Omega$, Chow and Lin (1971) showed that the BLUE for the regression parameter $\hat{\beta}_d$ and the disaggregated (or unobserved high frequency) data $\hat{y}_d$ are given by

$$\hat{\beta}_d = (X_d'C'(C\Omega C')^{-1}CX_d)^{-1}X_d'C'(C\Omega C')^{-1}y_a \tag{2}$$

$$\hat{y}_d = X_d\hat{\beta}_d + \Omega C'(C\Omega C')^{-1}(y_a - CX_d\hat{\beta}_d), \tag{3}$$

where $y_a = Cy_d$ is the observed dependent variable at the aggregated level (while y_d is unobserved at the disaggregated level), and C is a $N \times n$ (with $n \geq N$) aggregation matrix consisting of 0's and 1's, indicating which cells have to be aggregated together. The essential part in the equation 2 and 3 is the residual covariance matrix Ω, which has to be estimated. The Chow-Lin procedure for the BLUE requires the knowledge or assumptions about this error covariance matrix. In the literature assumptions like random walk, white noise, Markov random walk or autoregressive process of order one have been suggested and tested (e.g. Fernandez, 1981; Di Fonzo, 1990; Litterman, 1983; Pavia-Miralles et al., 2003). Some authors extended the framework for the multivariate case (e.g. Rossi, 1982; Di Fonzo, 1990) covering time and space for example (e.g. Pavia-Miralles and Cabrer-Borras, 2007). Usually, constraints are imposed to make sure that the predicted unobserved series adds up to the observed lower frequency series, e.g. by specifying penalty functions (e.g. Denton, 1971). In this case, the discrepancy between the sum of the predicted high frequency observations and the corresponding low frequency observation is divided up over the high frequency data through some other assumptions.

There are important practical problems to solve if the Chow-Lin procedure is applied. First, one has to find a suitable set of observed disaggregated indicators. The Chow-Lin data completion are predictions of the model and totally rely on the indicators chosen and the fit of the forecasting model. Another important feature is the structure of the residual covariance matrix, which becomes important for the spatial and the system extension of the Chow-Lin method.

We summarize the structure of any Chow-Lin data completion (= fine-forecasting) method in the following 4 steps:

1. First, decide on a forecasting or base model with only intensive (or aggregable) regression variables for the unobserved data at the disaggregated level.
2. Decide on an aggregation matrix C that aggregates the disaggregated model into a fully observed aggregated model.
3. Estimates the disaggregated parameters using the aggregated reduced form of the base model.
4. Compute the disaggregated Chow-Lin forecasts based on known regression indicators in the base model.

These basic 4 steps can be adapted to more complex Chow-Lin models and form the basis for different estimation methods (classical or Bayesian, etc.) for the parameters and predictions of the disaggregated

model. In this paper we will show how the Chow-Lin method can be extended for the case where the dependent variable in the base model is intensive (non-summable over sub-units as e.g. growth rates).

Note that the Chow-Lin method is a (conditional) forecasting method for disaggregated data and can be eventually evaluated by forecast criteria if disaggregated data could be observed. In general, a good Chow-Lin model is in first line a predictive model and follows the advices and rules of how to build good forecasting models and is in second line an inference model. The goal is to get a good fit at an aggregated level, which in turn should lead to good forecasts at the disaggregate level.

2.2. Assumptions for the Chow-Lin forecasting procedure

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

Assumption 1. *Structural similarity: The aggregated model for* $\mathbf{y}_c$ *and the disaggregated model for* $\mathbf{y}_d$ *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 on both aggregation levels are similar. In the system approach we are assuming that the correlation structure between first differences and levels are similar on an aggregated and on a disaggregated level.*

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.*

2.3. Some properties of the Chow-Lin forecasts

This section discusses the structure of the Chow-Lin forecasts and analyzes some properties. First, the gain-in-mean term $Q\hat{\varepsilon}_a$ can be seen as a cutting or 'spatial smearing out' of the aggregated residual vector $\hat{\varepsilon}_a$ to the simple disaggregate forecasts $\hat{y}_d$. In case of $\rho = 0$ or $R = I_n$ we find the gain to be a simple 'reverse projection' or allocator matrix $Q = C'(CC')^{-1}$: in this case each aggregated residual $\hat{\varepsilon}_{a,i}$ is divided by n_i and is equally distributed over the n_i disaggregated sub-units.

It is interesting to note that G is a right generalized inverse of C (i.e. is orthogonal to the aggregation matrix C), because of $CG = I_N$ and the aggregated Chow-Lin forecasts have the property

$$C\hat{\hat{y}}_d = C\hat{y}_d + \hat{\varepsilon}_a, \quad or \quad agg.CL.forecast \; = \; agg.plain \; + agg.residual. \tag{4}$$

That means that the aggregated Chow-Lin forecasts are equal to the aggregated naive forecasts plus the

aggregated residuals. We like to note the following statistical properties of the Chow-Lin forecasts:

- The first property that follows from (4) is that on average the Chow-Lin forecasts and the plain forecasts are equal (just post-multiply (4) by a vector of 1's).

$$C\ Ave(\hat{\hat{y}}_d) = C\ Ave(\hat{y}_d).$$

- The second property is that the aggregated Chow-Lin forecasts have a larger variance than the aggregated naive forecasts:

$$\hat{\hat{y}}'_d C' C \hat{\hat{y}}_d > \hat{y}'_d C' C \hat{y}_d.$$

- The third property is based on

$$\hat{y}_d = X_d \hat{\beta}_d + Q\hat{\varepsilon}_a$$

with the 'reverse projection' or allocator matrix $Q = \Omega C'(C\Omega C')^{-1}$ and leads to the following error sum of squares (ESS) decomposition

$$\begin{aligned} ESS_{CL} &= ESS_{plain} + ESS_{alloc} + noise \quad or \\ \hat{y}'_d \hat{y}_d &= \hat{\beta}'_d X'_d X_d \hat{\beta}_d + \hat{\varepsilon}'_a Q' Q \hat{\varepsilon}_a + noise. \end{aligned} \tag{5}$$

$\hat{\varepsilon}_q = Q\hat{\varepsilon}_a$ is the allocation residual for the disaggregated units, which is the gain term that stems from the allocation of the aggregated residual $\hat{\varepsilon}_a$ using the allocator Q. ESS_{CL} is the error sum of squares of the Chow-Lin forecasts $\hat{y}_d$, ESS_{plain} is the error sum of squares of the plain or reduced form (RF) forecasts and ESS_{alloc} is the error sum of squares of the allocation residuals or gain-in-mean term. The relative decomposition takes the form

$$1 = \frac{\hat{\beta}'_d X'_d X_d \hat{\beta}_d}{\hat{y}'_d \hat{y}_d} + \frac{\hat{\varepsilon}'_a Q' Q \hat{\varepsilon}_a}{\hat{y}'_d \hat{y}_d} + rest\ . \tag{6}$$

where the 'rest' is the remainder of the decomposition that adds up to 1.

For the special case that $\Omega = I_n$ we find for the allocator product $Q'Q = (CC')^{-1} = D_N^{-1}$, but in the general case the allocator product is $Q'Q = (C\Omega C')^{-1}C\Omega^2C'(C\Omega C')^{-1}$.

Therefore the Chow-Lin point forecasts for the disaggregated observations y_d are forecasts 'with gain', where the average size of the gain – or the improvements to the naive forecasts – comes from the size of the aggregated residuals. The dispersion of the Chow-Lin forecasts are smaller due to the reduction of the variance of the gain-in-variance term G in (38).

3. The system Chow-Lin method for completing growth rates

3.1. The non-spatial intensive Chow-Lin procedure

To see the need for a different method for intensive (non-additive or non-aggregable) variables, consider the growth rates in 2 disaggregated regions: $\frac{\Delta y_1}{y_1}$ and $\frac{\Delta y_2}{y_2}$, which have to be combined to the growth rate of the aggregated region: $\frac{\Delta y_1+\Delta y_2}{y_1+y_2}$. The growth rate has to be understood as made up by the usual temporal difference between 2 periods, i.e. $\Delta y_1 = y_{1t} - y_{1,t-1}$ and $\Delta y_2 = y_{2t} - y_{2,t-1}$. Since this is a non-linear operation we have to aggregate the numerator and the denominator separately.

This leads to the system Chow-Lin model for a disaggregated $n \times 1$ cross-sectional model with differences and levels:

Definition [The bivariate system Chow-Lin (biCL) model]

$$\begin{pmatrix}\Delta y_d \\ y_d\end{pmatrix} = \begin{pmatrix} X_{d1} & 0 \\ 0 & X_{d2}\end{pmatrix}\begin{pmatrix}\beta_{d1} \\ \beta_{d2}\end{pmatrix} + \begin{pmatrix}\varepsilon_1 \\ \varepsilon_2\end{pmatrix} \quad with \quad \begin{pmatrix}\varepsilon_1 \\ \varepsilon_2\end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, \Sigma \otimes I_n\right], \tag{7}$$

where we assume that different sets of regression indicators in X_{d1} and X_{d1} explain nominator and denominator. Σ is a 2×2 covariance matrix where the off-diagonal element σ_{12} contains the correlation parameter between the levels and the first differences. In compact notation these two equations in (7) are called 'system' or 'bivariate Chow-Lin' model and can be written compactly as

$$\tilde{y}_d = \tilde{X}_d\tilde{\beta}_d + \tilde{\varepsilon} \quad with \quad \tilde{\varepsilon} \sim \mathcal{N}[0, \widetilde{\Sigma} = \Sigma \otimes I_n], \tag{8}$$

where $\tilde{y}_d = \binom{\Delta y_d}{y_d}$, $\tilde{X}_d = \binom{X_{d1}}{X_{d2}}$, $\tilde{\beta}_d = \binom{\beta_{d1}}{\beta_{d2}}$, and $\tilde{\varepsilon} = \binom{\varepsilon_1}{\varepsilon_2}$.

Now we have to apply the aggregation matrix $C : N \times n$ for both equations separately or use the system

aggregation matrix $\widetilde{C} = diag(C, C) = I_2 \otimes C$. As before, we obtain the aggregated reduced form (ARF)

$$\begin{aligned} \widetilde{C}\tilde{y}_d &= \tilde{C}\tilde{X}_d\tilde{\beta}_d + \tilde{C}\tilde{\varepsilon} \quad with \quad \widetilde{C}\widetilde{\varepsilon} \sim \mathcal{N}[0, \Omega = \widetilde{C}\widetilde{\Sigma}\widetilde{C}'] \quad or \\ \tilde{y}_a &\sim \mathcal{N}[\tilde{X}_a\tilde{\beta}_d, \Omega] \quad with \quad \Omega = \Sigma \otimes D_N, \end{aligned} \tag{9}$$

because $CC' = D_N$ and the observed aggregates are $\tilde{X}_a = \widetilde{C}\tilde{X}_d$ and $\tilde{y}_a = \widetilde{C}\tilde{y}_d$. For the estimated regression coefficients $\widehat{\tilde{\beta}}_d$ in the non-spatial eCL model (7) we get the GLS estimate

$$\widehat{\tilde{\beta}}_d = (\tilde{X}_a'(\widehat{\Sigma} \otimes D_N)^{-1}\tilde{X}_a)^{-1}\tilde{X}_a'(\widehat{\Sigma} \otimes D_N)^{-1}y_a. \tag{10}$$

Since the covariance matrix is not known we need to estimate them from the LS estimates of the system equation:

$$\widehat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} \\ ./. & \hat{\sigma}_{22} \end{pmatrix} = \begin{pmatrix} Var(\hat{\varepsilon}_1) & Cov(\hat{\varepsilon}_1, \hat{\varepsilon}_2) \\ ./. & Var(\hat{\varepsilon}_2) \end{pmatrix} \tag{11}$$

with $\hat{\sigma}_{11} = Var(\hat{\varepsilon}_1)$, $\hat{\sigma}_{22} = Var(\hat{\varepsilon}_2)$, and $\hat{\sigma}_{12} = Cov(\hat{\varepsilon}_1, \hat{\varepsilon}_2)$. The estimated aggregated residuals are $\hat{\varepsilon}_{a1} = \Delta y_a - X_{a1}\hat{\beta}_{d1}$ and $\hat{\varepsilon}_{a2} = y_a - X_{a2}\hat{\beta}_{d2}$ with the GLS estimates

$$\hat{\beta}_{d1} = (X_{a1}'(\widehat{\Sigma} \otimes D_N)^{-1}X_{a1})^{-1}X_{a1}'(\widehat{\Sigma} \otimes D_N)^{-1}\Delta y_a \tag{12}$$

$$\hat{\beta}_{d2} = (X_{a2}'(\widehat{\Sigma} \otimes D_N)^{-1}X_{a2})^{-1}X_{a2}'(\widehat{\Sigma} \otimes D_N)^{-1}y_a. \tag{13}$$

The plain system forecasts of the growth rate model in the non-spatial case are given by (7)

$$\widehat{\tilde{y}}_{d,0} = \widetilde{X}_d\widehat{\tilde{\beta}}_d \quad and \quad \widehat{y}_{d,\%} = X_{d1}\hat{\beta}_{d1}./.X_{d2}\hat{\beta}_{d2} \tag{14}$$

with $\widehat{\tilde{\beta}}_d = \binom{\hat{\beta}_{d1}}{\hat{\beta}_{d2}}$ and ./. denotes element-wise division.

In a diagonal system we can separate the 2 β coefficient estimates into $\hat{\beta}_{di} = (X_{ai}'D_N^{-1}X_{ai})^{-1}X_{ai}'D_N^{-1}y_{ai}$, because the variances cancel out and $CC' = D_N = diag(n_1, ..., n_N) : N \times N$, where the n_i are the number of sub-units in each aggregated unit and $y_{a1} = \Delta y_a$ and $y_{a2} = y_a$.

Finally, the non-additive or intensive Chow-Lin forecasts $\hat{y}_d$ (for the unobserved disaggregated y_d in the

non-spatial model is given by

$$\widehat{\tilde{y}}_d \quad = \quad \tilde{X}_d\widehat{\tilde{\beta}}_d + \widehat{\tilde{\Sigma}}\tilde{C}'(\tilde{C}\widehat{\tilde{\Sigma}}\tilde{C}')^{-1}(\tilde{y}_a - \tilde{X}_a\widehat{\tilde{\beta}}_d) \tag{15}$$

with $\tilde{\Sigma} = \widehat{\Sigma} \otimes D_N$ already given in (10).

Finally, the disaggregate forecasts of the growth rates vector $r(y_d)$ are given by the ratio of the Chow-Lin forecasted nominator and denominator

$$\hat{r}(y_d) = \widehat{\Delta y}_d./.\widehat{y}_d, \tag{16}$$

where ./. denotes element-wise division, and the Chow-Lin forecast vectors $\widehat{\Delta y}_d$ and $\widehat{y}_d$ given in (3). The non-summable or intensive Chow-Lin forecasts are computed by

$$\widehat{\tilde{y}}_d \quad = \quad \tilde{X}_d\widehat{\tilde{\beta}} + \tilde{\Sigma}\tilde{C}'(\tilde{C}\tilde{\Sigma}\tilde{C}')^{-1}\tilde{\varepsilon}_a, \quad \tilde{\varepsilon}_a = \tilde{y}_a - \tilde{X}_a\widehat{\tilde{\beta}}_d, \tag{17}$$

and the system allocator $\tilde{C}_{\Cap}$ can be simplified by

$$\tilde{C}_{\Cap} = \tilde{\Sigma}\tilde{C}'(\tilde{C}\tilde{\Sigma}\tilde{C}')^{-1} = (\tilde{\Sigma} \otimes I_n)(I_2 \otimes C')((I_2 \otimes C)(\tilde{\Sigma} \otimes I_n)(I_2 \otimes C'))^{-1} = (I_2 \otimes C'(CC')^{-1}) = I_2 \otimes C_{\Cap} \tag{18}$$

with $CC' = D_N$ and $C'(CC')^{-1} = C_{\Cap}$ being the univariate allocator of residuals. This leads to the surprising result that in the system Chow-Lin model the Chow-Lin forecasts can be made independently for both equations:

$$\begin{aligned} \widehat{\Delta y}_d \quad &= \quad X_{d1}\hat{\beta}_{d1} + C'(CC')^{-1}\hat{\varepsilon}_{a1}, \quad \hat{\varepsilon}_{a1} = \Delta y_a - X_{a1}\hat{\beta}_{d1}, \\ \hat{y}_d \quad &= \quad X_{d2}\hat{\beta}_{d2} + C'(CC')^{-1}\hat{\varepsilon}_{a2}, \quad \hat{\varepsilon}_{a2} = y_a - X_{a2}\hat{\beta}_{d2}. \end{aligned} \tag{19}$$

Thus the correlation of the components of the non-additive or intensive Chow-Lin model for growth rates forecast have no influence on the Chow-Lin predictions. The Chow-Lin point forecasts for the (disaggregated) growth rates are given by:

$$\widehat{\hat{y}}_{d0} = \widehat{\Delta y}_d./.\hat{y}_d. \tag{20}$$

3.2. The spatial extension of the system Chow-Lin (SAR-SCL) model

Consider a 'bivariate' cross-sectional Chow-Lin model of n regions as in (7) where we fit a spatial autoregressive (SAR) model for the system of 2 equations

$$\widetilde{y}_d = diag(\rho_{1d}, \rho_{2d})\widetilde{W}_d\widetilde{y}_d + \widetilde{X}_d\widetilde{\beta}_d + \tilde{\varepsilon}_d, \quad \tilde{\varepsilon}_d \quad \sim \quad \mathcal{N}[0, \Sigma_2 \otimes I_n] \tag{21}$$

where Σ_2 is the covariance matrix between the 2 equations and has to be estimated as in (11), $\widetilde{X}_d = diag(X_{d1}, X_{d2})$, $\widetilde{\beta}_d = \binom{\beta_{1d}}{\beta_{2d}}$ and $\widetilde{W}_d = diag(W_{1d}, W_{2d})$. ρ_1 and ρ_2 are the spatial correlation coefficients associated with spatial lag variables $\Delta y_{d,\odot 1} = W_1 \Delta y_{1d}$ and $y_{d,\odot 1} = W_2 y_{2d}$, where the index $\odot 1$ stands for the first order spatial neighbor and the neighborhood matrix $\widetilde{W}_d$ is row normalized.

This has the advantage that the SAR model restricts the spatial correlation coefficients to the interval $\rho_{id} \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$, where λ_{min} and λ_{max} $(= 1$ because of the row normalizing) are the extreme eigenvalues of $W_i, i = 1, 2$. The reduced form model is obtained by the spread matrix $\widetilde{R} = diag(I_n - \rho_1 W_1, I_n - \rho_2 W_2) = diag(R_1, R_2)$ for an appropriately chosen weight matrices $W_i : n \times n$ for $i = 1, 2$.

$$\tilde{y}_d = \tilde{R}^{-1}\widetilde{X}_d\widetilde{\beta}_d + \tilde{R}^{-1}\tilde{\varepsilon}_d, \quad with \quad \tilde{R}^{-1}\tilde{\varepsilon}_d \quad \sim \quad \mathcal{N}[0, \Omega = (\tilde{R}'\tilde{\Sigma}^{-1}\tilde{R})^{-1}]. \tag{22}$$

For $\tilde{\Sigma} = \Sigma \otimes I_n$ we find $\Omega^{-1} = R_1'\Sigma^{-1}R_1 \otimes R_2'R_2$. In case $\Sigma = diag(\sigma_1, \sigma_2)$ is diagonal we get $\Omega = \sigma_1(R_1'R_1)^{-1} \otimes \sigma_2(R_2'R_2)^{-1}$.

The spread matrix $\widetilde{R}$ has to be positive definite to be inverted and this imposes another feasibility condition on the parameter space of the ρ_i's:

$$\widetilde{R} > 0 \ (pos.def.) \quad if \quad Det(\widetilde{R}) > 0 \quad . \tag{23}$$

In a Bayesian estimation procedure this condition is easy to implement: After the draws from the full conditional distributions we just have to check this condition. see LeSage and Pace (2004).

We rewrite the intensive CL system (21) as a SAR(2) model in the following way

$$\widetilde{y}_d = \rho_{1d}\widetilde{W}_{1d}\widetilde{y}_d + \rho_{2d}\widetilde{W}_{2d}\widetilde{y}_d + \widetilde{X}_d\widetilde{\beta}_d + \tilde{\varepsilon}_d, \quad \tilde{\varepsilon}_d \quad \sim \quad \mathcal{N}[0, \widetilde{\Sigma} = \Sigma \otimes I_n] \tag{24}$$

with $\widetilde{W}_{1d} = \begin{pmatrix} W_1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\widetilde{W}_{2d} = \begin{pmatrix} 0 & 0 \\ 0 & W_2 \end{pmatrix}$.

Note: The aggregation of the intensive Chow-Lin SAR(2) model is obtained by multiplying equation (24) with the $2N \times 2n$ matrix $\widetilde{C}$ and produces

$$\begin{aligned} \widetilde{C}\widetilde{y}_d &= \rho_{1d}\widetilde{C}\widetilde{W}_{1d}\widetilde{y}_d + \rho_{2d}\widetilde{C}\widetilde{W}_{2d}\widetilde{y}_d + \widetilde{C}\widetilde{X}_d\widetilde{\beta}_d + \widetilde{C}\tilde{\varepsilon}_d, \quad \widetilde{C}\tilde{\varepsilon}_d \sim \mathcal{N}[0, C\Sigma C' \otimes CC'] \quad or \\ \widetilde{y}_a &= \rho_{1d}\widetilde{W}_{1C}\widetilde{y}_d + \rho_{2d}\widetilde{W}_{2C}\widetilde{y}_d + \widetilde{X}_a\widetilde{\beta}_d + \tilde{\varepsilon}_a, \quad \tilde{\varepsilon}_a = \widetilde{C}\tilde{\varepsilon}_d \sim \mathcal{N}[0, C\Sigma C' \otimes D_N], \end{aligned} \tag{25}$$

where $D_N = CC'$ is a diagonal matrix and $\widetilde{W}_{1C} = \widetilde{C}\widetilde{W}_{1d}$ and $\widetilde{W}_{2C} = \widetilde{C}\widetilde{W}_{2d}$ are left-aggregated W_i matrices. This aggregation of the SAR(2)-formulation of the intensive Chow-Lin cannot be used to estimate the $\widetilde{\beta}_d$ coefficients, so we need the aggregated reduced form (ARF).

Note that the aggregation of the differences Δy_d has the commutation property $C\Delta y_d = \Delta C y_d$ as

$$C\Delta y_d = Cy_d - Cy_{d,-1} = y_a - y_{a,-1} = \Delta y_a. \tag{26}$$

The aggregated reduced form (ARF) model is obtained by multiplying the reduced form equation (22) with the $2N \times 2n$ matrix $\widetilde{C}$

$$\begin{aligned} \tilde{C}\tilde{y}_d &= \widetilde{C}\tilde{R}^{-1}\tilde{X}_d\widetilde{\beta}_d + \tilde{C}\tilde{R}^{-1}\tilde{\varepsilon}_d, \quad with \quad \widetilde{C}\tilde{R}^{-1}\tilde{\varepsilon}_d \sim \mathcal{N}[0, \widetilde{\Omega} = \tilde{C}\Omega\tilde{C}'] \\ &\quad or \\ \tilde{y}_a &= \tilde{X}_{a\rho}\widetilde{\beta}_d + \tilde{\varepsilon}_{a\rho} \quad with \quad \widetilde{C}\tilde{R}^{-1}\tilde{\varepsilon}_a \sim \mathcal{N}[0, \widetilde{\Omega}] \end{aligned} \tag{27}$$

with $\tilde{y}_a = \tilde{C}\tilde{y}_d = \binom{C\Delta y_d}{Cy_d}$, $\tilde{X}_{a\rho} = \widetilde{C}\tilde{R}^{-1}\tilde{X}_d = diag(CR_1X_{d1}, CR_2X_{d2})$ the 'sprawled' regressors, and $\tilde{\varepsilon}_{a\rho} = \widetilde{C}\tilde{R}^{-1}\tilde{\varepsilon}_d = \binom{CR_1\varepsilon_{d1}}{CR_2\varepsilon_{d2}}$. The variance-covariance matrix Ω of the 'sprawled' residuals $\tilde{R}^{-1}\tilde{\varepsilon}_d$ is given by

$$Cov(\tilde{R}^{-1}\tilde{\varepsilon}_d) = \Omega = \tilde{R}^{-1}\widetilde{\Sigma}\tilde{R}'^{-1} = (\tilde{R}'\widetilde{\Sigma}^{-1}\tilde{R})^{-1}.$$

The precision matrix is

$$\begin{aligned}\Omega^{-1} &= \tilde{R}'\widetilde{\Sigma}^{-1}\tilde{R} = diag(R_1', R_2')\begin{pmatrix} \sigma^{11} & \sigma^{12} \\ ./. & \sigma^{22} \end{pmatrix} diag(R_1, R_2) = \\ &= \begin{pmatrix} \sigma^{11}(R_1'R_1) & \sigma^{12}(R_1'R_2) \\ ./. & \sigma^{22}(R_2'R_2) \end{pmatrix}. \end{aligned} \tag{28}$$

In case of a diagonal $\Sigma = diag(\sigma_{11}, \sigma_{22})$ matrix we find $\Omega = diag((R_1'R_1)^{-1}/\sigma_{11}, (R_2'R_2)^{-1}/\sigma_{22})$.

Thus, the $2N \times 2N$ covariance matrix $\widetilde{\Omega}$ of the aggregated residuals $\tilde{\varepsilon}_a$ takes the form

$$\begin{aligned}\widetilde{\Omega} &= \tilde{C}\Omega\tilde{C}' = \\ &= (I_2 \otimes C)\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ ./. & \Omega_{22} \end{pmatrix}(I_2 \otimes C') = \\ &= \begin{pmatrix} \sigma^{11}C(R_1'R_1)^{-1}C' & \sigma^{12}C(R_1'R_2)^{-1}C' \\ ./. & \sigma^{22}C(R_2'R_2)^{-1}C' \end{pmatrix} = \begin{pmatrix} \widetilde{\Omega}_{11} & \widetilde{\Omega}_{12} \\ ./. & \widetilde{\Omega}_{22} \end{pmatrix}. \end{aligned} \tag{29}$$

Based on the aggregated reduced form (27), the GLS estimate of $\widetilde{\beta}_d$ for known ρ_1, ρ_2 and Ω can be computed as

$$\widetilde{\beta}_{GLS} = (\tilde{X}_a'\Omega^{-1}\tilde{X}_a)^{-1}\tilde{X}_a\Omega^{-1}\tilde{y}_a \tag{30}$$

and the feasible GLS estimate $\widehat{\widetilde{\beta}}_{GLS}$ replaces Ω with an estimate $\widehat{\Omega}$. Denote the partitioned inverse by

$$\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ ./. & \Omega_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} \\ ./. & \Omega^{22} \end{pmatrix} : 2N \times 2N \tag{31}$$

then the GLS estimates are given by the $2k \times 1$ vector

$$\begin{aligned}\widetilde{\beta}_d &= \left[diag\,(X_{a1}', X_{a2}')\begin{pmatrix} \Omega^{11} & \Omega^{12} \\ ./. & \Omega^{22} \end{pmatrix} diag\,(X_{a1}, X_{a2}) \right]^{-1} diag\,(X_{a1}', X_{a2}')\begin{pmatrix} \Omega^{11} & \Omega^{12} \\ ./. & \Omega^{22} \end{pmatrix}\begin{pmatrix} y_{a1} \\ y_{a2} \end{pmatrix} = \\ &= \begin{pmatrix} X_{a1}'\Omega^{11}X_{a1} & X_{a1}'\Omega^{12}X_{a2} \\ X_{a2}'\Omega^{21}X_{a1} & X_{a2}'\Omega^{22}X_{a2} \end{pmatrix}^{-1}\begin{pmatrix} X_{a1}'(\Omega^{11}y_{a1} + \Omega^{12}y_{a2}) \\ X_{a2}'(\Omega^{21}y_{a1} + \Omega^{22}y_{a2}) \end{pmatrix} \end{aligned} \tag{32}$$

In case the ρ_i's have to be estimated we refer to this procedure as feasible GLS (FGLS) estimation. Based

on the coefficients estimate of the aggregated model we can forecast the missing values at the disaggregate level. This is possible in two ways: the first way neglects the system framework of the Chow-Lin method, i.e. the seemingly unrelated correlation of the aggregated and the disaggregated model and is therefore the usual univariate regression forecasts, in this paper called Chow-Lin without gain. This plain or 'no-gain' forecast in the reduced form is the usual point forecast at the observed disaggregated (low-frequency) indicator X_d and is given by

$$\begin{pmatrix} \widehat{\Delta y}_d \\ \hat{y}_d \end{pmatrix} = \widehat{\tilde{y}}_d = \widehat{R}^{-1}\widetilde{X}_d\widehat{\tilde{\beta}}_d = \begin{pmatrix} \hat{R}_1^{-1}X_{d,1}\hat{\beta}_{d1} \\ \hat{R}_2^{-1}X_{d,2}\hat{\beta}_{d2} \end{pmatrix}, \tag{33}$$

with the estimated spread matrix $\widehat{R} = diag(\hat{R}_1, \hat{R}_2)$ and $\hat{R}_i = I_n - \hat{\rho}_i W$. For the plain prediction, all the regressor variables in $\tilde{X}_d$ at the disaggregated level have to be known for all n regions. The second method uses the spatial correlation structure between the aggregated and the disaggregated model and we obtain forecasts with the gain, i.e. conditional normal estimates, where we condition the disaggregated forecasts on the known values of the aggregated model.

Note the dependency of the covariance matrix on the parameters ρ_1, ρ_2 that is part of the spread matrix R. In the Chow-Lin framework, the aggregated model is almost always given by completely observed data. Therefore, we can estimate $\widetilde{\beta}_d$ by GLS or maximum likelihood methods, although the estimates can become quite unreliable because only fewer observations are available for estimation on an aggregate level.
The joint distribution of disaggregates and aggregates uses the reduced form of the aggregated (27) and the disaggregated model (22) is given by

$$\begin{pmatrix} \tilde{y}_d \\ C\tilde{y}_d \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \tilde{\mu}_d = \tilde{y}_d = \tilde{R}^{-1}\widetilde{X}_d\widetilde{\beta}_d \\ \tilde{\mu}_a = C\tilde{\mu}_d \end{pmatrix}, \begin{pmatrix} \Omega & \Omega\tilde{C}' \\ \tilde{C}\Omega & \tilde{C}\Omega\tilde{C}' \end{pmatrix} \right] \tag{34}$$

with Ω given in (27). The conditional mean $\widehat{\tilde{y}}_d$ for the disaggregated observations given the aggregated data $\tilde{y}_a = \tilde{C}\tilde{y}_d$ have to be calculated by the partitioned inverse rule.[1]

[1]For the partitioned normal distribution

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma_{xx}} & \mathbf{\Sigma_{xy}} \\ \mathbf{\Sigma'_{xy}} & \mathbf{\Sigma_{yy}} \end{pmatrix} \right]$$

the conditional distribution is given by $\mathcal{N}[\mu_{x|y}, \Sigma_{x|y}]$ with

$$\mu_{x|y} = \mu_{\mathbf{x}} + \mathbf{\Sigma_{xy}}(\mathbf{\Sigma_{yy}})^{-\mathbf{1}}(\mathbf{y} - \mu_{\mathbf{y}})$$

This leads to the forecasting formula (3) for $\hat{y}_d$ that is common to all Chow-Lin methods, see Polasek and Sellner (2010).

$$\widehat{\hat{y}}_d = \widehat{\tilde{R}}^{-1} \tilde{X}_d \widehat{\tilde{\beta}}_d + \widehat{\tilde{g}}_d = \hat{y}_{plain} + \hat{y}_{gain}, \tag{35}$$

where the $\widehat{\tilde{g}}_d$ is the 'gain-in-mean' term of the Chow-Lin forecasts, because it is an improvement over the plain or reduced form forecast of the not observed y_d values in (33).

Thus, the aggregated reduced form (ARF) of the spatial regression system is the important model basis to make Chow-Lin forecasts and is structurally similar to the univariate spatial model (1) - in order to apply the Chow-Lin forecast formula. Using the covariance matrix Ω in (29) of the reduced form model of the spatial Chow-Lin system for growth rates (point forecasts) are given by

$$\widehat{\tilde{y}}_d = \tilde{X}_d \widehat{\tilde{\beta}} + \Omega \tilde{C}' (\tilde{C} \Omega \tilde{C}')^{-1} (\tilde{y}_a - \tilde{C} \tilde{X}_d \widehat{\tilde{\beta}}_d), \tag{36}$$

and the gain-in-mean term $\widetilde{\tilde{g}}_d$ plays the role of an allocator (of the residuals), where the estimated aggregated residual is given by

$$\widehat{\tilde{\varepsilon}}_a = \tilde{y}_a - \tilde{C} \tilde{R}^{-1} \tilde{X}_d \hat{\tilde{\beta}}_d$$

and the gain-in-mean term is

$$\widehat{\tilde{g}}_d = \Omega \tilde{C}' (\tilde{C} \Omega \tilde{C}')^{-1} \hat{\tilde{\varepsilon}}_a = \begin{pmatrix} \Omega_{11} C' & \Omega_{12} C' \\ ./. & \Omega_{22} C' \end{pmatrix} \begin{pmatrix} C\Omega_{11} C' & C\Omega_{12} C' \\ ./. & C\Omega_{22} C' \end{pmatrix}^{-1} \begin{pmatrix} \hat{\varepsilon}_{a1} \\ \hat{\varepsilon}_{a2} \end{pmatrix} \tag{37}$$

and the 'gain-in-variance' matrix $\tilde{G}$, which was first used by Goldberger (1962), is given by

$$\tilde{G} = \Omega \tilde{C}' (\tilde{C} \Omega \tilde{C}')^{-1} \tilde{C} \Omega. \tag{38}$$

3.3. A two step feasible GLS (FGLS) estimation

Based on the above system extension of the Chow-Lin method we suggest the following 2-step (feasible GLS) estimation for a spatial system Chow-Lin procedure to complete growth rates.

,

$$\Sigma_{x|y} = \mathbf{\Sigma_{xx} - \Sigma_{xy} (\Sigma_{yy})^{-1} \Sigma_{yx}}.$$

Procedure 1 (Two step feasible GLS (FGLS) estimation).

1. *First, we estimate* $\hat{\rho}_1, \hat{\rho}_2$ *by ML (or LS) in the SAR model for first differences and in levels.*
2. *Get the LS residuals from the SAR models and estimate the simple covariance matrix* $\widehat{\Sigma} = \Sigma(\hat{\rho}_1, \hat{\rho}_2)$.
3. *Compute the feasible system estimate* $\tilde{\beta}_{FGLS,d}$ *using* $\hat{\Omega}$.
4. *Compute the vector of system Chow-Lin forecasts as in (19).*
5. *Compute the vector of growth rates from the Chow-Lin forecasts* $\Delta\hat{y}_d./.\hat{y}_d$.

This procedure can be easily implemented along the existing statical program packages that allow SAR model estimation.

4. A Bayesian Chow-Lin model for completing growth rates

This section describes the estimation Bayesian system SAR-CL model, which builds upon the C-aggregation of the reduced form as given (27). The prior distribution for the parameters of the SAR-CL model $\theta = (\tilde{\beta}_d, \Sigma^{-1}, \rho_1, \rho_2)$ is proportional to

$$p(\tilde{\beta}_d, \Sigma^{-1}, \rho_1, \rho_2) \quad \propto \quad p(\tilde{\beta}_d)\ p(\Sigma^{-1}) = \mathcal{N}[\tilde{\beta}_d \mid \tilde{\beta}_*, H_*]\ \mathcal{W}[\Sigma^{-1} \mid \Sigma_*^{-1}, n_*],$$

where $\mathcal{W}$ stands for the Wishart distribution and where we assume a uniform prior for $\rho_i \sim U[-1, 1], i = 1, 2$. The joint distribution of $\theta_d = (\tilde{\beta}_d, \rho_1, \rho_2, \Sigma^{-1})$ in the Bayesian SAR-CL model is given by

$$p(\theta_d \mid \tilde{y}_d) = \mathcal{N}[\tilde{y}_d \mid \tilde{C}\tilde{R}^{-1}\tilde{X}_d\widetilde{\beta}_d, \sigma^2\Sigma]\ \mathcal{N}[\tilde{\beta}_d \mid \beta_*, H_*]\ \mathcal{W}[\Sigma^{-1} \mid S_*, n_*]. \tag{39}$$

Consider the Bayesian SAR system that we use to model the intensive Chow-Lin (SAR-eCL) model. and let us denote the 3 types of conditional distributions by $p(\rho \mid y_a, \theta^c), p(\beta_d \mid y_a, \theta^c)$, and $p(\Sigma^{-1} \mid y_a, \theta^c)$, where $\theta = (\rho, \beta_d, \Sigma^{-1})$ denotes all the parameter of the model, and θ^c denotes the complementary parameters (to the current daily new argument) of the full conditional distribution (fcd), respectively. The Markov Chain Monte Carlo (MCMC) procedure consists of 3 blocks of sampling, as is shown in the next theorem:

Theorem 1 (MCMC for the non-additive or intensive Chow-Lin (eCL-SAR) model).

The MCMC estimation for the system SAR Chow-Lin model (21), with the joint distribution defined in (39) involves the following iteration steps:

Step 1: *Draw* $\tilde{\beta}_d$ *from* $\mathcal{N}\left[\tilde{\beta}_d \mid \tilde{b}_{**}, \tilde{H}_{**}\right]$;
Step 2: *Draw* ρ_1 *and* ρ_2 *by griddy Gibbs;*
Step 3: *Draw* Σ^{-1} *from* $\mathcal{W}[\Sigma^{-1} \mid S_{**}, n_{**}]$;
Step 4: *Repeat until convergence.*

Proof 1.

(a) *The full conditional for the* $\tilde{\beta}_d$ *regression coefficients is*

$$\begin{aligned} p(\tilde{\beta}_d \mid \tilde{y}_a, \theta^c) &= \mathcal{N}[\tilde{\beta}_d \mid \tilde{b}_*, \tilde{H}_*]\, \mathcal{N}[\tilde{C}\tilde{y}_d \mid \tilde{C}\tilde{R}^{-1}\widetilde{X}_d\widetilde{\beta}_d, \Omega] \\ &= \mathcal{N}\left[\tilde{\beta}_d \mid \tilde{b}_{**}, \tilde{H}_{**}\right] \end{aligned} \tag{40}$$

with $\tilde{C}\tilde{y}_d = \tilde{y}_a$,

$$\widetilde{\Sigma} = \Sigma \otimes I_n \quad \textit{and} \quad \widetilde{\Omega} = \tilde{C}(\tilde{R}'\tilde{\Sigma}^{-1}\tilde{R})^{-1}\tilde{C}'. \tag{41}$$

The hyper-parameters are

$$\begin{aligned} \tilde{H}_{**}^{-1} &= \tilde{H}_*^{-1} + \tilde{X}_d'\tilde{R}'^{-1}\tilde{C}'\widetilde{\Omega}^{-1}\tilde{C}\tilde{R}^{-1}\tilde{X}_d\tilde{H}_*^{-1} + \tilde{H}, \\ \tilde{b}_{**} &= \tilde{H}_{**}[\tilde{H}_*^{-1}\tilde{b}_* + \tilde{X}_d'\tilde{R}'^{-1}\tilde{C}\widetilde{\Omega}^{-1}\tilde{y}_a]. \end{aligned}$$

Denote the partitioned inverse Ω *matrix as in (29) by*

$$\widetilde{\Omega}^{-1} = \begin{pmatrix} \widetilde{\Omega}^{11} & \widetilde{\Omega}^{12} \\ ./. & \widetilde{\Omega}^{22} \end{pmatrix}, \tag{42}$$

then the variance-covariance part $\widetilde{H} = \tilde{X}_d'\tilde{R}'^{-1}\tilde{C}'\widetilde{\Omega}^{-1}\tilde{C}\tilde{R}^{-1}\tilde{X}_d$ *in the hyper-parameters is*

$$\widetilde{H} = \begin{pmatrix} X_{d1}'R_1'C'\widetilde{\Omega}^{11}CR_1X_{d1} & X_{d1}'R_1'C'\widetilde{\Omega}^{12}CR_2X_{d2} \\ ./. & X_{d2}'R_2'C'\widetilde{\Omega}^{22}CR_2X_{d2} \end{pmatrix} \quad \textit{and} \quad \widetilde{h} = \begin{pmatrix} X_{d1}'R_1'C'\widetilde{\Omega}^{11}y_{a1} + X_{d1}'R_1'C'\widetilde{\Omega}^{12}y_{a2} \\ X_{d2}'R_2'C'\widetilde{\Omega}^{21}y_{a1} + X_{d2}'R_2'C'\widetilde{\Omega}^{22}y_{a2} \end{pmatrix}. \tag{43}$$

(b) *For the fcd of the inverse variance matrix* Σ^{-1} *of the residuals we find from the reduced form in (22)*

$$\begin{aligned} p(\Sigma^{-1} \mid \tilde{y}_a, \theta^c) &\propto \mathcal{W}[\Sigma^{-1} \mid S_*, n_*]\, \mathcal{N}[y_a \mid \mu, \widetilde{\Omega} = \tilde{C}\tilde{\Sigma}\tilde{C}'] \\ &\propto \mathcal{W}[\Sigma^{-1} \mid S_{**}, n_{**}] \end{aligned} \tag{44}$$

with the (hyper-)parameters $n_{**} = n_* + n$ *and* $n_{**}S_{**} = n_*S_* + n\hat{S}$. *The error sum of squares matrix* $n\hat{S}$ *is given by*

$$n\hat{S} = \begin{pmatrix} \varepsilon_1'\varepsilon_1 & \varepsilon_1'\varepsilon_2 \\ ./. & \varepsilon_2'\varepsilon_2 \end{pmatrix} \tag{45}$$

because of

$$vec\ E = \varepsilon = \tilde{z}_a - \widetilde{X}_a\widetilde{\beta}_d \quad with \quad \tilde{z}_a = \widetilde{y}_a - \rho_{1d}\widetilde{W}_{1C}\widetilde{y}_d + \rho_{2d}\widetilde{W}_{2C}\widetilde{y}_d$$

we find for the quadratic form

$$(\tilde{z}_a - \widetilde{X}_a\widetilde{\beta}_d)'(I_2 \otimes C)(\Sigma \otimes I_n)(I_2 \otimes C')(\tilde{z}_a - \widetilde{X}_a\widetilde{\beta}_d) = \varepsilon_a'(\Sigma \otimes D_n)\tilde{C}'\varepsilon = tr(\tilde{E}_a\Sigma\tilde{E}_a'D_n).$$

with $D_n = CC'$, $\tilde{\varepsilon}_a = vec\tilde{E}_a$ *and the residual matrix* $E_a = [\varepsilon_{1a} : \varepsilon_{2a}] : n \times 2$ *with the columns being the residuals* $\varepsilon_{1a} = \Delta y_a - X_{a,1}\beta_{1d}$ *and* $\varepsilon_{2a} = y_a - X_{a,2}\beta_{2d}$.

(c) *The full conditional distribution (fcd) for the spatial correlation coefficients* $\boldsymbol{\rho}$ *is done with griddy Gibbs in 2 steps and the rho-grid for both spatial correlation coefficients have to be set-up independently. We propose to use a symmetric grid around an initial simple rho estimate.*
1) For the fcd for ρ_1 *we use as dependent variable* $\mathbf{z}_2 = \tilde{\mathbf{y}}_a - \rho_2\tilde{\mathbf{W}}_2\tilde{\mathbf{y}}_a$

$$\begin{aligned} p(\rho_1) &\propto exp\left[-\frac{1}{2\sigma_y^2}(\mathbf{z}_2 - \rho_1\tilde{\mathbf{W}}_1\tilde{\mathbf{y}}_a)'(\mathbf{z}_2 - \rho_1\tilde{\mathbf{W}}_1\tilde{\mathbf{y}}_a)\right] \propto exp\left[-\frac{1}{2\sigma_y^2}(\rho_1 - \hat{\rho}_1)^2 S_2\right] \\ &\propto N\left[\rho_1 \mid \hat{\rho}_1, \sigma_y^2/S_2\right] \end{aligned} \tag{46}$$

with $\hat{\rho}_1 = \mathbf{y}'\mathbf{W}_1'\mathbf{z}_2/S_2$ *and* $S_2 = \mathbf{y}'\mathbf{W}_1'\mathbf{W}_1\mathbf{y}$.
2) For the fcd for ρ_2 *we use as dependent variable* $\mathbf{z}_1 = \tilde{\mathbf{y}}_a - \rho_1\tilde{\mathbf{W}}_1\tilde{\mathbf{y}}_a$

$$\begin{aligned} p(\rho_2) &\propto exp\left[-\frac{1}{2\sigma_y^2}(\mathbf{z}_1 - \rho_2\tilde{\mathbf{W}}_1\tilde{\mathbf{y}}_a)'(\mathbf{z}_1 - \rho_2\tilde{\mathbf{W}}_1\tilde{\mathbf{y}}_a)\right] \propto exp\left[-\frac{1}{2\sigma_y^2}(\rho_2 - \hat{\rho}_1)^2 S_1\right] \\ &\propto N\left[\rho_2 \mid \hat{\rho}_2, \sigma_y^2/S_1\right] \end{aligned} \tag{47}$$

with $\hat{\rho}_2 = \tilde{\mathbf{y}}_a'\tilde{\mathbf{W}}_2'\mathbf{z}_1/S_1$ *and* $S_1 = \tilde{\mathbf{y}}_a'\tilde{\mathbf{W}}_2'\tilde{\mathbf{W}}_2\tilde{\mathbf{y}}_a$.

Note: In case of a diagonal Σ matrix there have 2 residual σ_i^2 variances to be estimated. Draw

$$\sigma_1^{-2} \sim \Gamma[\sigma_1^{-2} \mid s_{1**}^2, n_{1**}] \quad and \quad \sigma_2^{-2} \sim \Gamma[\sigma_2^{-2} \mid s_{2**}^2, n_{2**}]$$

with $n_{i**} = n_{i*} + N, i = 1, 2$ and $s_{i**}^2 n_{i**} = s_{i*}^2 n_{i*} + ESS_i$. The error sum of squares (ESS) are given by $ESS_1 = (\Delta y_a - \rho_1 CW_1 y_d - X_{a1}\beta_{d1})'(\Delta y_a - \rho_1 CW_1 y_d - X_{a1}\beta_{d1})$ and $ESS_2 = (y_a - \rho_2 CW_2 y_d - X_{a2}\beta_{d2})'(y_a - \rho_2 CW_2 y_d - X_{a2}\beta_{d2})$.

From the MCMC simulation we obtain a numerical sample of the posterior distribution $\Theta_{MCMC} = p(\beta_d^{(j)}, \rho_1^{(j)}, \rho_2^{(j)}, \Sigma_j^{-1} \mid y_d)$, which is used to make the predictions for the missing disaggregate observations y_d.

4.1. The Bayesian Chow-Lin predictions of growth rates

In Bayesian inference, we obtain the posterior predictive distribution for $\widetilde{y}_p$ in the following way, by integrating over the conditional predictive distribution with the posterior distribution $p(\widetilde{\beta}_d, \rho, \Sigma^{-1} \mid \tilde{y}_a)$

$$p(\widetilde{y}_p \mid \tilde{y}_d) = \int\int\int p(\tilde{y}_p \mid \widetilde{\beta}, \rho, \Sigma^{-1})\; p(\widetilde{\beta}, \rho, \Sigma^{-1} \mid \tilde{y})\; d\widetilde{\beta}\; d\rho\; d\Sigma^{-1}$$

with $\rho = (\rho_1, \rho_2)$ and where the posterior normal-gamma density $p(\widetilde{\beta}_d, \rho, \Sigma^{-1} \mid \tilde{y}_a)$ is found numerically by the MCMC sample, yielding a posterior sample of the θ parameters: $\Theta_{MCMC} = \{(\widetilde{\beta}_d^{(j)}, \rho_{j,1}, \rho_{j,2}, \Sigma_j^{-1}), \quad j = 1, ..., J\}$.

Next, we compute a numerical predictive sample of the unknown vector $\tilde{y}_d$ by drawing from the reduced form using the known regressors $\tilde{X}_d$ for $j = 1, ..., J$

$$\tilde{y}_d^{(j)} \sim \mathcal{N}[\tilde{R}_j^{-1}\tilde{X}_d\tilde{\beta}_d^{(j)} + \tilde{g}_j, \Omega_j - \tilde{G}_j] \quad \text{with} \quad \tilde{R} = diag(R_1, R_2), \tag{48}$$

where Ω given in (29) and the spread matrix $\tilde{R} = diag(R_1, R_2)$ depends on the spread matrices $R_j = I_n - \rho_j W,\ j = 1, ..., J.$ $\tilde{g}$ is the gain-in-mean vector as in (37) and $\tilde{G}$ is the gain-in-variance matrix as in (38) for the mean and covariance matrix of the predictions, which are computed by

$$\begin{aligned} \tilde{G}_j &= \Omega_j\tilde{C}'(\tilde{C}\Omega_j\tilde{C}')^{-1}\tilde{C}\Omega_j \\ \tilde{g}_j &= \Omega_j\tilde{C}'(\tilde{C}\Omega_j\tilde{C}')^{-1}\tilde{e}_{a,j} \\ \Omega_j &= \tilde{R}_j^{-1}(\Sigma_j \otimes I_n)\tilde{R}'^{-1} = \begin{pmatrix} \Omega_{11}^{(j)} & \Omega_{12}^{(j)} \\ ./. & \Omega_{22}^{(j)} \end{pmatrix}, \end{aligned} \tag{49}$$

using the covariance structure as in (29)

$$\Omega = (\sigma^{ik}(R_i'R_k)^{-1})_{i=1,2;k=1,2} = \begin{pmatrix} \sigma^{11}(R_1'R_1)^{-1} & \sigma^{12}(R_1'R_2)^{-1} \\ ./. & \sigma^{22}(R_2'R_2)^{-1} \end{pmatrix}.$$

The σ^{ik} are the inverse elements of Σ^{-1} and we use the aggregated residuals $\tilde{e}_{a,j} = \tilde{y}_a - \tilde{\mu}_{a,j}$ since the current aggregate fit is $\tilde{\mu}_{a,j} = \tilde{C}\tilde{R}_j^{-1}\tilde{X}_d\widetilde{\beta}_d^{(j)}$.

Finally, the disaggregate forecasts of the growth rates vector $r(y_d)$ is given by the ratio of the Chow-Lin forecasted nominator and denominator similar to (16). The forecast sample of the $n \times 1$ vectors of growth rates r is

$$\hat{r}_{mcmc}(y_d) = \{\Delta y_d^{(j)}./.y_d^{(j)} = y_{d1}^{(j)}./.y_{d2}^{(j)}, j = 1, ..., J\}, \tag{50}$$

from where we can compute numerically the mean vector $Ave(\hat{r}_{mcmc}(y_d))$ of the draws and the interval predictions (e.g. by deciles, quantiles, etc.) for all of the n sub-units. A simpler way is found by applying

the method of Doucet (2010). We start from the joint distribution as in (34)

$$\begin{pmatrix} \tilde{y}_d \\ \tilde{y}_a \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \tilde{\mu}_d^{(j)} = \tilde{R}_{(j)}^{-1} \widetilde{X}_d \widetilde{\beta}_d^{(j)} \\ \tilde{\mu}_a^{(j)} = \tilde{C} \tilde{\mu}_d^{(j)} \end{pmatrix}, \begin{pmatrix} \Omega^{(j)} & \Omega^{(j)} \tilde{C}' \\ \tilde{C}\Omega^{(j)} & \tilde{C}\Omega^{(j)}\tilde{C}' \end{pmatrix} \right] \tag{51}$$

with Ω given in (29). We propose the following procedure to predict the disaggregate growth rates r_d by MCMC:

Procedure 2 (Fine-forecasting growth rates $r_{mcmc} = \{r_d^j\}$ by MCMC).

1. *Draw* $\binom{\widetilde{Y}_d^{(j)}}{\widetilde{Y}_a^{(j)}}, j = 1, ..., J$ *from the joint density (51);*
2. *Compute the conditional draws given the observed aggregate values* y_a

$$\widetilde{Y}_d^{(j)} | y_a = Y_d^{(j)} - g_d^{(j)}$$

with the gain-in-mean term $g_d^{(j)} = \Omega^{(j)} \tilde{C}' (\tilde{C}\Omega^{(j)}\tilde{C}')^{-1} (Y_a^{(j)} - y_a)$ *as in (37) and compute*

$$\tilde{g}_d^{(j)} = \Omega^{(j)} \tilde{C}' (\tilde{C}\Omega^{(j)}\tilde{C}')^{-1} \tilde{\varepsilon}_a^{(j)} = \begin{pmatrix} \Omega_{11}^{(j)} C' & \Omega_{12}^{(j)} C' \\ ./. & \Omega_{22}^{(j)} C' \end{pmatrix} \begin{pmatrix} C\Omega_{11}^{(j)} C' & C\Omega_{12}^{(j)} C' \\ ./. & C\Omega_{22}^{(j)} C' \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon_{a1}^{(j)} \\ \varepsilon_{a2}^{(j)} \end{pmatrix} \tag{52}$$

with $\Omega_{ik}^{(j)} = \sigma_{(j)}^{ik} (R_j' R_j)^{-1}$ *and* $\tilde{\varepsilon}_a^{(j)} = \tilde{y}_a - \tilde{C}\tilde{R}_j^{-1}\tilde{X}_d\widetilde{\beta}_d^{(j)} = \binom{\varepsilon_{a1}^{(j)}}{\varepsilon_{a2}^{(j)}}$.

3. *Compute the MMC sample of growth rates by the ratio* $r^{(j)} = Y_{d1}^{(j)} ./. Y_{d2}^{(j)}$ *from* $\widetilde{Y}_d^{(j)} | y_a = \binom{Y_{d1}^{(j)}}{Y_{d2}^{(j)}}$ *for all j.*

4.2. Sampling a conditional r.v. from a joint distribution

Recently Doucet (2010) has proposed a possibly faster way as how to simulate a conditional random variable from a joint normal distribution. A draw from the conditional density $\mathcal{N}[\mu_{x|y}, \Sigma_{x|y}]$ based on the joint density

$$\mathcal{N}[\mu, \Sigma] = \mathcal{N}\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}' & \Sigma_{yy} \end{pmatrix} \right] \tag{53}$$

can be obtained in the following way:

1. Draw the bivariate r.v. $\binom{\dot{X}}{\dot{Y}}$ from $\mathcal{N}[\mu, \Sigma]$;
2. Compute the conditional r.v. $\dot{X}|y = \dot{X} - \Sigma_{xy}\Sigma_{yy}^{-1}(\dot{Y} - y)$ for a known y-value.

In our case for the prediction of the disaggregate observations leads to the procedure

1. Draw $\binom{\dot{Y}_d}{\dot{Y}_a}$ from the joint density (51);

2. Compute the Chow-Lin forecasts by the conditional $\dot{Y}_d|\hat{y}_d = \dot{Y}_d - (R'R)^{-1}C'(C(R'R)^{-1}C')^{-1}(\dot{Y}_a - C\hat{y}_d)$ evaluated at the plain forecasts $\hat{y}_d$ from (48), which can be embedded into the MCMC iteration.

The conditional system or panel forecasts are made in the same way.

The estimation of the SAR system (21) with two ρ's is equivalent to the estimation of the SAR(2) model, which is outlined in the appendix. Note that we can write the SAR regression system as

$$\begin{aligned}
\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \rho_1 W_1 & 0 \\ 0 & \rho_2 W_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \varepsilon \quad = \\
&= \rho_1 \begin{pmatrix} W_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \rho_2 \begin{pmatrix} 0 & 0 \\ 0 & W_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \varepsilon \quad or \\
\tilde{y} &= \rho_1 \, \tilde{W}_1 \tilde{y} + \rho_2 \, \tilde{W}_2 \tilde{y} + \varepsilon. \qquad (54)
\end{aligned}$$

Note that the 0-augmented W_i matrices act like a selection matrix for the x and the y component when modeling the dependent variable $\tilde{y}$.[2]

4.3. Model selection by marginal likelihood

The marginal likelihood of model $\mathcal{M}$ is computed by the harmonic mean formula

$$\hat{m}(\mathbf{y} \mid \mathcal{M})^{-1} = \frac{1}{n_{rep}} \sum_{j=1}^{n_{rep}} \left(\sum_{i=1}^{n} l(\mathcal{D}_i \mid \mathcal{M}, \theta_j) \right)^{-1} \qquad (55)$$

where $\mathcal{D}_i = (\Delta y_i, y_i)$ is the i-th data observation and with the likelihood given in (21). We also use the 1% trimmed harmonic estimator.

5. 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[3]. Spain has 18 regions (NUTS-2) and 52 provinces (NUTS-3). The associated C matrix is constructed from the knowledge of the hierarchical structure of the NUTS-2 to NUTS-3 regions. Note that, in contrast to the temporal Chow-Lin method where each

[2]The LS estimate in the model $\binom{x}{y} = \beta\binom{0}{z}$ is $\hat{\beta} = ((0, z')\binom{0}{z})^{-1}(0, z')\binom{x}{y} = (z'z)^{-1}z'y$.

[3]All data and the hierarchical C-Matrix for Spanish provinces are available from the authors upon request.

aggregated period (year) has the same number of disaggregated stretches (4 quarters, 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 1 and 9, and 7 regions are single unit regions, having just 1 province. This heterogeneity in terms of size and administrative structure makes Spanish regions a real challenge and testing ground for spatial Chow-Lin methods.

The Figures 5 a) and b) show the outcome of the Chow-Lin forecasting method for aggregated (NUTS-2) and disaggregated (NUTS-3) regions. The forecast evaluations are given in Table 1.

Table 1: Chow-Lin Prediction Accuracy: Classical vs. Bayesian estimates

growth	rates		CORR[a]	MAE[b]	MAPE[c]	RMSE[d]
Classic	simple	(no gain)	-0.005	1.666	0.109	0.340
		with gain	0.204	0.5146*)	0.03475*)	0.05115
	spatial	with gain	0.217*)	0.5180	0.03501	0.05067*)
Bayesian		gain	0.101	0.377695	0.025228	.038274
MCMC		no gain	0.200	0.327792	0.021307	.035276
	spatial	with gain	0.211*)	0.308276*)	0.020109*)	.033292*)

*)... best value

[a]Correlation of predicted and observed
[b]Mean Absolute Error
[c]Mean Absolute Percentage Error
[d]Root Mean Squared Error

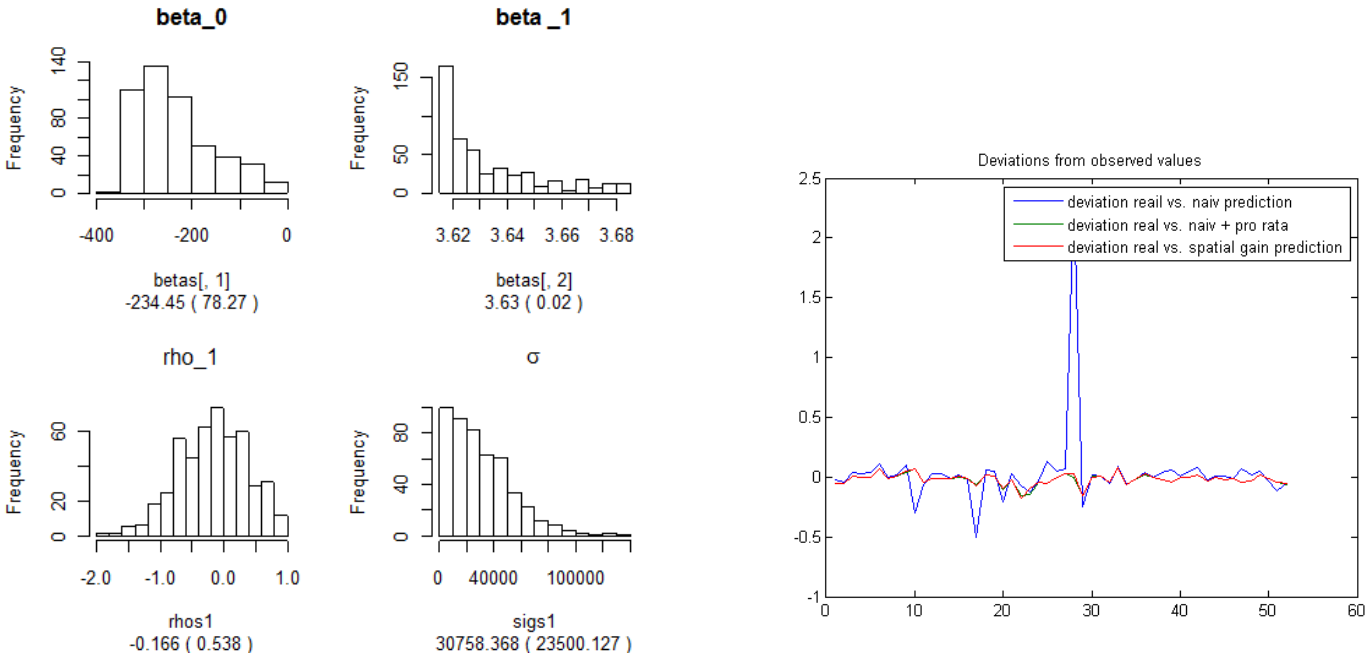

Figure 1: a) Spanish regional CL forecasts: aggregated growth rates; b) disaggregated growth rates

6. Outliers in the aggregate equation

Since the aggregate equation is an auxiliary equation in the Chow-Lin forecasting problem, we can use all the usual techniques to eliminate extreme and outlying observations from the aggregate equation. Also, since the number of observation is considerably smaller in the aggregate equation, an outlier will have a larger effect on the regression coefficients of the indicator variables that are the primary targets and of direct interests. Thus we can estimate as aggregate equation the model

$$y = X\beta + Z\gamma + u,$$

where the matrix Z could contain all nuisance variables that might potentiality alter or distort the effects of the indicator variables, β. The matrix Z could contain any other external variables or fixed effects that is irrelevant to the forecasting process. For the Chow-Lin forecasting in the disaggregate equation we are only using the β coefficients since we are interested in the mean effects of the disaggregate units and we not trying to forecast idiosyncratic effects. It could be argued that the outlier in the aggregate equation is stemming from a certain unit in the area of the aggregate unit. But unless we don't know the origin of the outliers with certainty we can neglect this effect for the purpose of forecasting, which is computed always as a mean effect, and also the forecasting evaluation measures should reflect only the average behavior of the forecasting procedure.

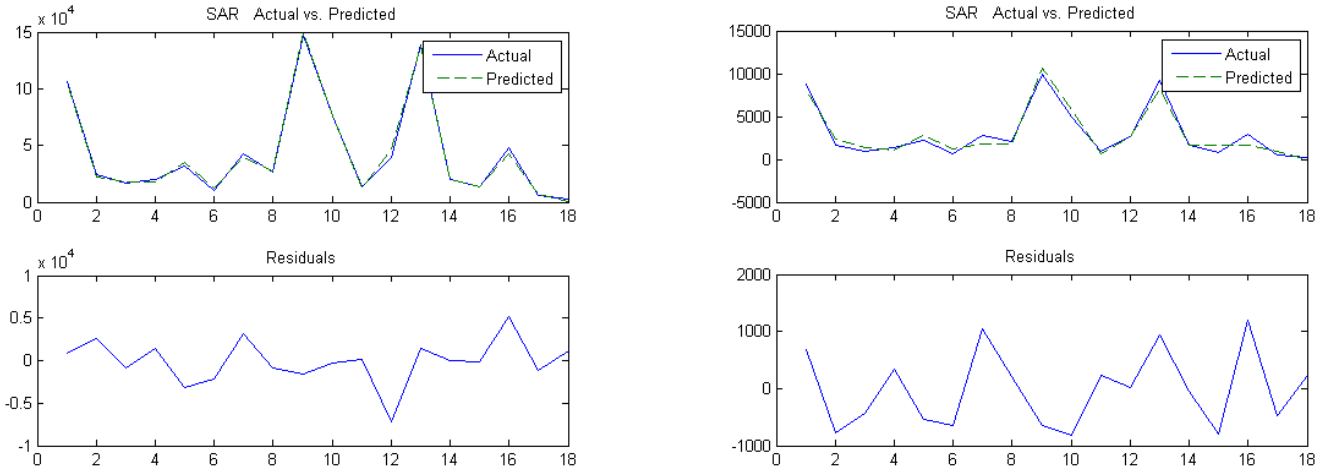

Figure 2: a) Residuals from agg. level data; b) Resid of agg. differenced data

The relative decomposition as in (6) takes the form

$$1 = \frac{\hat{\tilde{\beta}}_d' \tilde{X}_d' \tilde{X}_d \hat{\tilde{\beta}}_d}{\hat{\tilde{y}}_d' \hat{\tilde{y}}_d} + \frac{\hat{\varepsilon}_a' Q' Q \hat{\varepsilon}_a}{\hat{\tilde{y}}_d' \hat{\tilde{y}}_d} + rest \ . \tag{56}$$

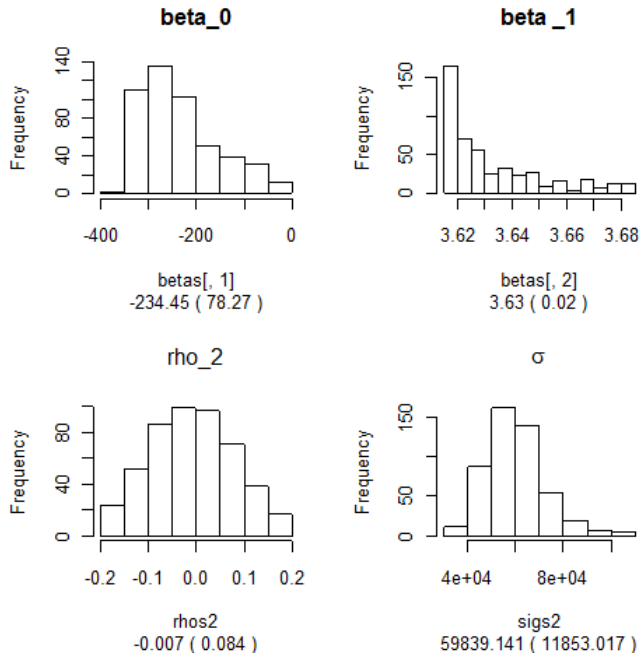

Figure 3: GDP level 2004

7. Conclusions

In regional science the task of fine-casting, i.e. the prediction of variables for quite small spatial units when only larger units can be observed, is called the Chow-Lin forecasting method. Surprisingly, the procedure needs an extension if non-summable variables have to be predicted. For spatial or regional analyses we sometimes need to predict growth rates for smaller, disaggregated units and this paper has shown that the spatial Chow-Lin procedure can be used to make forecasts (better fine-forecasts) for these missing subunits. The new procedure is demonstrated for growth rates, which is an example of non-summable random variables, uses the regression-indicators at the disaggregated regional level to predict the disaggregated unobserved dependent variable, conditional on the complete aggregated observed model. Spatial econometric applications have become increasingly important for many regional economic policy questions, especially through the integration process of the European Union, which focusses on quite small regional units, because

no EU region should be left behind.

The new approach 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 model. An important condition for finding a good model is the existence of good indicator variables and some good modeling skills to find the appropriate weight matrix to estimate the spatial effects. In future research we will explore these modeling possibilities in more heterogenous environments, especially if it comes to the question what is the best way of aggregating smaller to larger units, and we could 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.

8. References

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

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.

Denton, F., 1971. Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. Journal of American Statistical Association 66, 99–102.

Di Fonzo, T., 1990. The estimation of m disaggregate time series when contemporaneous and temporal aggregates are known. The Review of Economics and Statistics 71, 178–182.

Doucet, A., 2010. A note on efficient conditional simulation of gaussian distributions,, departments of Computer Science and Statistics, University of British Columbia, Vancouver, BC, Canada.

Fernandez, R. B., 1981. A methodological note on the estimation of time series. The Review of Economics and Statistics 53, 471–478.

Friedman, M., 1962. The interpolation of time series by related series. Journal of American Statistical Association 57, 729–757.

LeSage, J. P., Pace, R. K., 2004. Models for spatially dependent missing data. Journal of Real Estate Finance and Economics 29, 233–254.

LeSage, J. P., Pace, R. K., 2009. Introduction to Spatial Econometrics. Chapman & Hall/CRC.

Litterman, R. B., 1983. A random walk, markov model for the distribution of time series. Journal of Business and Economic Statistics 1, 169–173.

Llano, C., Polasek, W., Sellner, R. P., 2009. Bayesian methods for completing data in space-time panel models. Economic Series 241, Institute for Advanced Studies.

Pavia-Miralles, J. M., Cabrer-Borras, B., 2007. On estimating contemporaneous quarterly regional gdp. Journal of Forecasting 26, 155–170.

Pavia-Miralles, J. M., Vila-Lladosa, L.-E., Valles, R. E., 2003. On the performance of the chow-lin procedure for quarterly interpolation of annual data: Some monte-carlo analysis. Spanish Economic Review 5, 291–305.

Polasek, W., Sellner, R., 2010. Spatial chow-lin methods for data completion in econometric flow models. Economic Series 255, Institute for Advanced Studies.

Rossi, N., 1982. A note on the estimation of disaggregate time series when the aggregate is known. Review of Economics and Statistics 64, 695–696.

9. APPENDIX: Higher order SAR models

9.1. MCM: Griddy Gibbs for the SARX model

We consider the SAR(1)X model

$$y = \rho W y + X\beta + \varepsilon, \quad \varepsilon \sim N[0, \sigma_y^2] \tag{57}$$

with the informative prior $p(\theta) = \mathcal{N}[\tilde{C}\tilde{R}^{-1}\tilde{X}\widetilde{\beta}_d, \sigma^2\Sigma]\ \mathcal{N}[\beta_d \mid \beta_*, H_*]\ \mathcal{W}[\Sigma^{-1} \mid S_*, n_*]$.

Theorem 2 (MCMC for the SARX model).
The MCMC estimation for the SARX model (57), with the joint distribution $p(\theta, y) = \mathcal{N}[\rho W y + X\beta, \sigma_y^2 I_n]\ p(\theta)$ *involves the following iteration steps:*

Step 1: *Starting value* $\rho = 0$*;*
Step 2: *Draw* β *from* $\mathcal{N}[\beta \mid b_{**}, H_{**}]$*;*
Step 3: *Draw* σ^{-2} *from* $\Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}]$*;*
Step 4: *Draw* ρ *by griddy Gibbs;*
Step 5: *Repeat until convergence.*

Proof 2.

(a) *The full conditional for the* β *regression coefficients follows for known* ρ *a simple linear regression model with dependent variable* $z = y - \rho W y$

$$p(\beta \mid y, \theta^c) \quad = \quad \mathcal{N}[\beta \mid b_*, H_*]\ \mathcal{N}[z \mid X\beta, \sigma_y^2 I_n] = \mathcal{N}[\beta \mid b_{**}, H_{**}]$$

with the hyper-parameters

$$\begin{aligned} H_{**}^{-1} &= H_*^{-1} + \sigma^{-2} X'X \\ b_{**} &= H_{**}[H_*^{-1} b_* + \sigma^{-2} X'z]. \end{aligned}$$

(b) *The fcd for the residual variance we find*

$$p(\sigma^{-2} \mid \mathbf{y}_a, \boldsymbol{\theta}^c) = \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}] \tag{58}$$

with $n_{**} = n_* + n$ *and* $s_{**}^2 n_{**} = s_*^2 n_* + (z - X\beta)'(z - X\beta)$.

(c) *The fcd for the spatial* ρ *we look at the kernel of the normal fcd, which is given with* $z = y - X\beta$ *by*

$$\begin{aligned} p(\rho) &\propto exp\left[-\frac{1}{2\sigma_y^2}(z - \rho W y)'(z - \rho W y)/\sigma^2\right] \propto exp\left[-\frac{1}{2\sigma_y^2}(\rho - \hat{\rho})^2 S_y\right] \\ &\propto \mathcal{N}\left[\rho \mid \hat{\rho}, \sigma_y^2/S_y\right] \end{aligned} \tag{59}$$

with $\hat{\rho} = y'W'z/S_y$ *and* $S_y = y'W'Wy$*. Using the griddy Gibbs concept of Ritter and Tanner (1992) we generate from the normal pdf a grid of 100 points and draw draw according to the discrete ordinates of the pdf. Note that the spatial rho is now estimated after* y *is purged from the influence of* X*.*

From the MCMC simulation we obtain a numerical sample of the posterior distribution $\Theta_{MCMC} = p(\beta_d^{(j)}, \rho_j, \Sigma_j^{-1} \mid y_d)$. The $ml = 1.403897$ for EMP model.

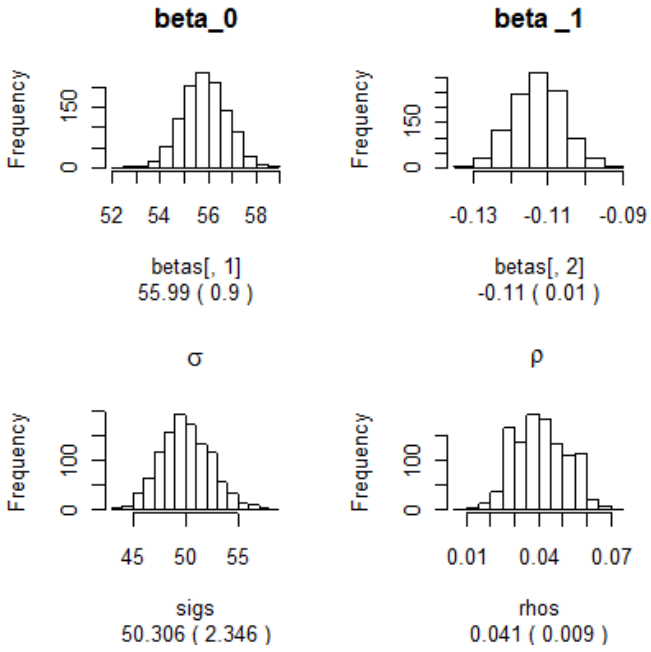

Figure 4: SAR(1)X with X=EMP data

9.2. The SAR(2) model

Using two types of neighborhood matrices W_1 and W_2 of the of a cross-sectional variable y, we can define the second order spatial SAR(2) model in the following formulate a 'structural' form:

$$\mathbf{y} = \mathbf{X}\beta + \rho_1 \mathbf{W}_1\mathbf{y} + \rho_2 \mathbf{W}_2\mathbf{y} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}\left[\mathbf{0}, \sigma^2 \mathbf{I}_n\right], \tag{60}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, and ρ_1 and ρ_2 are the first and second order spatial correlation parameter, respectively.

Next, we obtain a reduced form if we take all $\mathbf{y}$ variables on the left hand side:

$$\mathbf{z} = \mathbf{y} - \rho_1 \mathbf{W}_1\mathbf{y} - \rho_2 \mathbf{W}_2\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim \mathcal{N}\left[\mathbf{0}, \sigma^2 \mathbf{I}_n\right], \tag{61}$$

where $\mathbf{z} = (\mathbf{I}_n - \rho_1 \mathbf{W}_1 - \rho_2 \mathbf{W}_2)\mathbf{y}$ is a transformed dependent variable that can be easily computed if the ρ's in $\boldsymbol{\rho} = (\rho_1, \rho_2)$ would be known. Now we denote the inverse of the spatial transformation by the matrix $\mathbf{R}$ using the neighborhood polynomial:

$$\mathbf{R}_\rho^{-1} = (\mathbf{I}_n - \mathbf{W}(\boldsymbol{\rho}))^{-1} \quad with \quad \mathbf{W}(\boldsymbol{\rho}) = \rho_1 \mathbf{W}_1 + \rho_2 \mathbf{W}_2.$$

Therefore we can write the spatial regression model in its transformed form as

$$\mathbf{R}_\rho \mathbf{y} \sim \mathcal{N}\left[\mathbf{X}_d \boldsymbol{\beta}, \sigma^2 \mathbf{I}_n\right]. \tag{62}$$

Note that the reduced form is normally distributed with

$$\mathbf{y} \sim \mathcal{N}\left[\mathbf{R}_\rho^{-1} \mathbf{X}_d \boldsymbol{\beta}, \sigma^2 (\mathbf{R}_\rho^\top \mathbf{R}_\rho)^{-1}\right], \tag{63}$$

because $Var(\mathbf{R}\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{R}\mathbf{R}^\top$. The prior distribution for the parameter $\theta = (\boldsymbol{\beta}_d, \sigma^{-2}, \boldsymbol{\rho})$ is given by the product of (independent) blocks of normal and gamma distributions:

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{\rho}) &= p(\boldsymbol{\beta}_d)\ p(\sigma^{-2})\ \mathcal{U}[\rho_1 \mid -1, 1]\ \mathcal{U}[\rho_2 \mid -1, 1] \\ &= \mathcal{N}\left[\boldsymbol{\beta}_d \mid \mathbf{b}_*, \mathbf{H}_*\right]\ \Gamma[\sigma^{-2} \mid s_*^2, n_*]\ \frac{1}{4}, \end{aligned} \tag{64}$$

where $\mathcal{U}[-1, 1]$ stands for a uniform distribution in the interval $(-1, 1)$. Because of restrictions the interval of feasible ρ's depends on λ_{min} and λ_{max}, the minimum and maximum eigenvalue of $\mathbf{W}$. Because of $\lambda_{min}^{-1} < 0$ and $\lambda_{max}^{-1} > 0$ and therefore the ρ_i's must lie between these bounds. Thus, we restrict the prior space of the ρ's to the interval $(\lambda_{min}^{-1}, \lambda_{max}^{-1})$.

The joint distribution for $\mathbf{y}_d$ and the parameter $\theta = (\boldsymbol{\beta}_d, \sigma^2, \rho)$ is

$$p(\boldsymbol{\beta}_d, \sigma^{-2}, \rho, \mathbf{y}) \propto \mathcal{N}\left[\mathbf{y}_d \mid \mathbf{X}_d \boldsymbol{\beta}_d, \sigma^2\right]\ \mathcal{N}\left[\boldsymbol{\beta}_d \mid \mathbf{b}_*, \mathbf{H}_*\right]\ \Gamma[\sigma^{-2} \mid s_*^2, n_*].$$

Theorem 3 (Griddy Gibbs sampling in the Normal*Gamma SAR(2) model).

We consider the SAR(2) model as in (60) with prior (64)

1. *Starting values: set* $\boldsymbol{\beta}_d = \boldsymbol{\beta}_{OLS}$ *and* $\rho = 0$ *in the aggregated model*
2. *Draw* σ^{-2} *from* $\Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}]$
3. *Draw* $\boldsymbol{\beta}_d$ *from* $\mathcal{N}\left[\boldsymbol{\beta}_d \mid \mathbf{b}_{**}, \mathbf{H}_{**}\right]$
4. *Draw* $\boldsymbol{\rho}$ *by a griddy Gibbs step for* $(-1, 1) \times (-1, 1)$
5. *Repeat until convergence.*

The full conditional distributions are:

1. For the regression coefficients $\boldsymbol{\beta}_d$ we use the transformation of the dependent variable y to z as in (61)

since $\boldsymbol{\rho}$ is conditionally known

$$p(\boldsymbol{\beta}_d \mid \sigma^{-2}, \boldsymbol{\rho}, \mathbf{y}) \propto \mathcal{N}\left[\boldsymbol{\beta}_d \mid \mathbf{b}_{**}, \mathbf{H}_{**}\right], \tag{65}$$

and by combining quadratic forms in the usual way we find for the hyper-parameters

$$\begin{aligned} \mathbf{H}_{**}^{-1} &= \mathbf{H}_{*}^{-1} + \sigma^{-2}\mathbf{X}^{\top}\mathbf{X}, \\ \mathbf{b}_{**} &= \mathbf{H}_{**}\left[\mathbf{H}_{*}^{-1}\mathbf{b}_{*} + \sigma^{-2}\mathbf{X}^{\top}\mathbf{z}\right]. \end{aligned}$$

2. The fcd for the inverse residual variance σ^2 is

$$p(\sigma^{-2} \mid \boldsymbol{\beta}, \rho, \mathbf{y}) \propto \Gamma[\sigma^{-2} \mid s_{**}^2, n_{**}],$$

a gamma distribution with the parameters $n_{**} = n_* + n$ and

$$n_{**}s_{**}^2 = n_*s_{**}^2 + (\mathbf{R}_\rho\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\top}(\mathbf{R}_\rho\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{66}$$

where $\mathbf{R}_\rho\mathbf{y} = \mathbf{y} - \rho_1\mathbf{W}_1\mathbf{y} - \rho_2\mathbf{W}_2\mathbf{y}$.

3. The full conditional distribution for the spatial correlation coefficients $\boldsymbol{\rho}$ is done with griddy Gibbs in 2 steps and the ρ-grid for both spatial correlation coefficients have to be set-up independently. We propose to use a symmetric grid around an initial simple rho estimate.

   1) For the fcd for ρ_1 we use as dependent variable $\mathbf{z}_2 = \mathbf{y} - \rho_2\mathbf{W}_2\mathbf{y}$

$$\begin{aligned} p(\rho_1) &\propto exp\left[-\frac{1}{2\sigma_y^2}(\mathbf{z}_2 - \rho_1\mathbf{W}_1\mathbf{y})'(\mathbf{z}_2 - \rho_1\mathbf{W}_1\mathbf{y})\right] \propto exp\left[-\frac{1}{2\sigma_y^2}(\rho_1 - \hat{\rho}_1)^2 S_1\right] \\ &\propto \mathcal{N}\left[\rho_1 \mid \hat{\rho}_1, \sigma_y^2/S_1\right] \end{aligned} \tag{67}$$

   with $\hat{\rho}_1 = \mathbf{y}^{\top}\mathbf{W}_1^{\top}\mathbf{z}_2/S_2$ and $S_1 = \mathbf{y}^{\top}\mathbf{W}_1^{\top}\mathbf{W}_1\mathbf{y}$.

   2) For the fcd for ρ_2 we use as dependent variable $\mathbf{z}_1 = \mathbf{y} - \rho_1\mathbf{W}_1\mathbf{y}$

$$\begin{aligned} p(\rho_2) &\propto exp\left[-\frac{1}{2\sigma_y^2}(\mathbf{z}_1 - \rho_2\mathbf{W}_2\mathbf{y})^{\top}(\mathbf{z}_1 - \rho_2\mathbf{W}_2\mathbf{y})\right] \propto exp\left[-\frac{1}{2\sigma_y^2}(\rho_2 - \hat{\rho}_1)^2 S_2\right] \\ &\propto \mathcal{N}\left[\rho_2 \mid \hat{\rho}_2, \sigma_y^2/S_2\right] \end{aligned} \tag{68}$$

   with $\hat{\rho}_2 = \mathbf{y}^{\top}\mathbf{W}_2^{\top}\mathbf{z}_1/S_2$ and $S_2 = \mathbf{y}^{\top}\mathbf{W}_2^{\top}\mathbf{W}_2\mathbf{y}$. Because these two fcd for ρ_1 and ρ_2 are structural

identical to the fcd of ρ_1 in the SAR(1) model, we can easily use the griddy Gibbs or the multiple try MH algorithm for the MCMC estimation of the SAR(2) model.

The marginal likelihood is given by the harmonic MDL formula

$$p_{\mathcal{N}}(\mathbf{y} \mid SAR(2)...)^{-1} = \frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} p_{\mathcal{N}}(\mathbf{y} \mid \theta_{(i)}, \mathbf{X}, \mathbf{W})^{-1}$$

with the parameters for simulation i given by $\theta_{(i)} = (\boldsymbol{\beta}_{(i)}, \sigma^{-2}_{(i)}, \boldsymbol{\rho}_{(i)})$ and the likelihood function is

$$p_{\mathcal{N}}(\mathbf{y} \mid \theta_{(i)}, \mathbf{X}, \mathbf{W}) \quad = \quad (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_y^2}(\mathbf{R}_\rho \mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{R}_\rho \mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]. \tag{69}$$

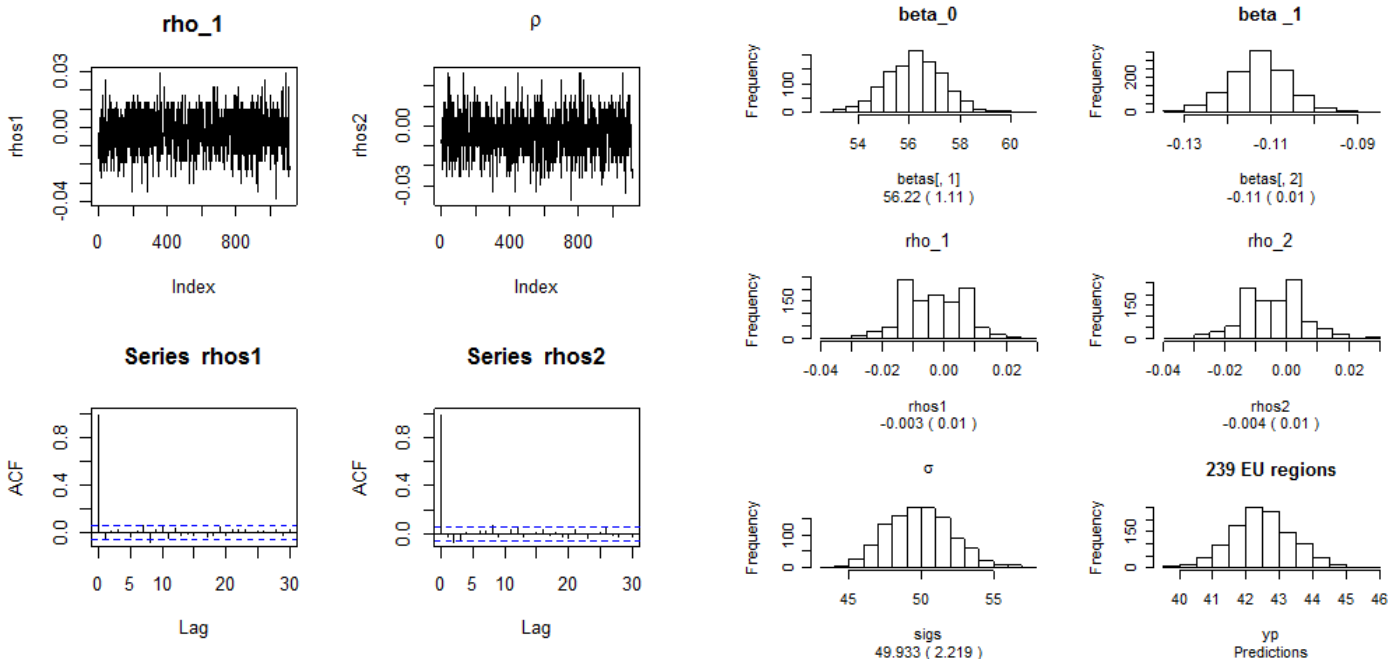

Figure 5: a) Paths and ACF of the rho coefficients; b) SAR(2): estimated coefficients

Author: Wolfgang Polasek

Title: Spatial Chow-Lin Models for Completing Growth Rates in Cross-sections

Reihe Ökonomie / Economics Series 295

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Michael Reiter (Macroeconomics), Selver Derya Uysal (Microeconomics)

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996