Regional convergence and spatial dependence: A worldwide perspective*

Marcos Sanso-Navarro[§a] and María Vera-Cabello[b]

[a]Departamento de Análisis Económico, Universidad de Zaragoza, Spain

[b]Centro Universitario de la Defensa de Zaragoza, Spain

marcossn@unizar.es, mvera@unizar.es

January 29, 2017

Abstract

This paper incorporates spatial dependence into a neoclassical regional growth framework with imperfect factor mobility. Using a sub-national global data set, the empirical analysis consists of the implementation of multiple imputation techniques to the estimation of a spatial Durbin panel model. Our results show that accounting for spatial effects increases the estimated regional convergence rate. This provides an explanation for puzzling findings in the related literature. Further, we obtain evidence of a nonlinear relationship between the levels of national income and financial development and the regional speed of convergence.

JEL classification: C23, O47, R11.

Keywords: Regional convergence, spatial Durbin model, panel data, multiple imputation.

*This paper was originally written while the first author was visiting the Department of Economics (Institute of Macroeconomics) at the Vienna University of Economics and Business (WU), whose members acknowledge their hospitality and friendliness. The authors have benefited from the valuable comments of Jesús Crespo Cuaresma. Financial support from Gobierno de Aragón (S16-ADETRE Research Group), Ministerio de Educación, Cultura y Deporte (José Castillejo Grant CAS16/00010) and Universidad de Zaragoza (Project JIUZ-2016-SOC-14) is also acknowledged.

§Corresponding author. Address: Facultad de Economía y Empresa. Departamento de Análisis Económico. Gran Vía 2. 50005 Zaragoza, Spain. Tel.: (+34) 876 554 629; Fax: (+34) 976 761 996.

1 Introduction

There is a recent trend in the economic growth and development literature towards adopting a worldwide sub-national perspective that has resulted in the emergence of new models and databases (Gennaioli et al., 2013, 2014; Mitton, 2016). Although these studies put forward interesting hypotheses, they do not seriously take into account the presence of spatial dependence as they use corrections that do not address underlying problems with the regression specification (Breinlich et al., 2014). Nevertheless, regional interdependence deserves special attention both from a theoretical and an empirical point of view because not controlling for existing externalities across regions may produce biased results and, hence, lead to misleading conclusions (Fingleton and López-Bazo, 2006; Benos et al., 2015).

The first representative of these new contributions to the regional development literature is the work by Gennaioli et al. (2013) who develop a so called 'Lucas-Lucas' model. It consists of a standard migration framework with both talent allocation between entrepreneurship and work and *within-region* human capital externalities. Sanso-Navarro et al. (2016) introduced technological interdependence *à la* Ertur and Koch (2007) into this model, achieving an expression for regional income per capita with spatial effects. This result corroborates the necessity of accounting for the presence of spatial dependence when studying the determinants of regional development. In doing so, these authors provide evidence of *between-regions* human capital externalities. That is to say, a higher stock of human capital in a region entails not only a higher technological level for that economy but also additional technological flows into its neighbors.

Given the homogeneity of the units under scrutiny, the explanatory power of the neoclassical growth model can be better assessed at the sub-national level. With this aim, Gennaioli et al. (2014, GLLS hereafter) incorporated a stylized process of limited mobility of physical and human capital into this theoretical framework, concluding that the speed of convergence between regions decreases with the intensity of the frictions. Moreover, factor mobility is expected to increase the influence of national income on regional growth. In order to empirically test for these predictions, a new data set with an extensive coverage both in the cross-sectional and temporal dimensions was constructed.

GLLS find that there is limited regional convergence because of institutional barriers to the mobility of resources and raise the puzzle of a slow convergence speed at the sub-national level. These authors also point out the need to shed more light on the role played by technological diffusion between regions and productive externalities in this context. Taking these arguments into account, we try to disentangle the extent to which this puzzling result regarding the convergence rate is driven by not explicitly controlling for spatial dependence. Initially, we contribute to the literature that establish theoretical foundations to the inclusion of this feature in convergence models (Ertur and Koch, 2007; Fischer, 2011; Pfaffermayr, 2012). Both this theoretical extension and the structure of the global data set constructed by GLLS lead us to pose an empirical analysis based on the estimation of a spatial Durbin panel model.

Due to the considerable amount of missing data at sub-national level, we are implementing a multiple imputation approach (Rubin, 1976, 1987). By proceeding in this way, we follow Belotti et al. (2017) who suggest using this simulation-based statistical technique as a way to deal with unbalancedness in spatial panel data. Although multiple imputation consists of replacing missing information by a number of sets of plausible values, our results are robust to changes in the model specification, the choice of the spatial weights matrix and the number of imputations. In a nutshell, we find that the estimated regional convergence speed increases when spatial dependence is taken into account. Further, we provide evidence of a nonlinear relationship between the levels of national income and financial development and the regional rate of convergence.

The rest of the paper is structured as follows. The theoretical framework and the empirical strategy are laid out in sections 2 and 3, respectively. Section 4 describes the data and shows some preliminary analysis. Section 5 presents the main results and Section 6 concludes.

2 A neoclassical regional growth model with imperfect factor mobility

Our point of departure is the theoretical framework of regional convergence and migration developed by GLLS. This model considers discrete time periods ($t = 0, 1, ...$) and a country with a measure one of regions, $i \in [0, 1]$. These regions are characterized by their levels of total factor productivity (TFP) A_i, population L_i and initial endowment of capital per capita $\hat{h}_{i,0}$. The latter includes both physical and human capital and, for a given period of time, is different to the amount of capital employed $h_{i,t}$. Technology is represented by a Cobb-Douglas production function with diminishing returns where the inputs are capital and labor. Therefore, output per capita is given by:

$$y_{i,t} = A_i h_{i,t}^{\alpha} \tag{1}$$

where α is the regional income share remunerating capital, expected to be close to unity as most labor productivity is determined by capital and skills. Consequently, $(1 - \alpha)$ is the share of income that rewards labor.

There is no TFP growth and more productive regions display higher levels of A_i. The competitive remuneration of capital is $w_{i,t} = \alpha A_i h_{i,t}^{\alpha - 1}$. If there is perfect mobility of capital, it will shift towards regions with higher $w_{i,t}$ until this remuneration equates across them. This will imply that:

$$\bar{h}_{i,t} = \hat{A}_i h_t \tag{2}$$

with $\hat{A}_i = \dfrac{A_i^{\frac{1}{1-\alpha}}}{\int A_i^{\frac{1}{1-\alpha}} di}$ and $h_t = \int \hat{h}_{i,t} di$.

It is considered that, at a given time period, each region invests the same share s of income in education and in physical capital. Moreover, capital fully depreciates in each period and there is no population growth. Under these assumptions, the initial endowment of capital at period $t + 1$ in region i is:

$$\hat{h}_{i,t+1} = s A_i h_{i,t}^{\alpha} \tag{3}$$

The relationship between this initial capital endowment and the level of employment will be determined by migration, which takes place before production and after the generation of new capital. Only highly skilled workers migrate, making human capital to move across regions (Gennaioli et al., 2013). On the one hand, each region will employ its capital endowment ($h_{i,t+1} = \hat{h}_{i,t+1}$) under infinite mobility costs. On the other hand, and taking into account expression (2), it will be obtained that $h_{i,t+1} = \hat{A}_i h_{t+1}$ under perfect mobility of capital; where $h_{t+1} = s \int A_i h_{i,t}^{\alpha} di$ denotes the resulting aggregate capital endowment at $t+1$.

Intermediate degrees of mobility will be allowed by assuming that:

$$h_{i,t+1} = v_{t+1} \left(\hat{h}_{i,t+1}\right)^{\tau} \left(\hat{A}_i h_{t+1}\right)^{1-\tau} \tag{4}$$

where $\tau \in [0,1]$ reflects mobility costs and

$$v_{t+1} = \frac{h_{t+1}^{\tau}}{\int \left(\hat{h}_{i,t+1}\right)^{\tau} \left(\hat{A}_i\right)^{1-\tau} di} \tag{5}$$

is a normalization factor.

According to (1), the growth rate of region i between t and $t+1$ is given by $\left(\frac{h_{i,t+1}}{h_{i,t}}\right)^{\alpha}$. Following equations (2) and (3), it is obtained that per capita income growth is determined by capital employment growth:

$$\frac{h_{i,t+1}}{h_{i,t}} = v_{t+1} h_{i,t}^{\alpha\tau-1} \left(sA_i\right)^{\tau} \left(\hat{A}_i s \int A_j h_{j,t}^{\alpha} dj\right)^{\tau-1} \tag{6}$$

This expression implies that region i's growth rate increases with the savings rate, TFP and aggregate investment. In addition, and due to diminishing returns, the initial level of capital stock reduces growth. Therefore, the economy evolves dependent on the capital endowment and migration which, in turn, determine the aggregate levels of capital endowment and output. Diminishing returns make regional income to converge to a unique steady state with non-zero income and an absence of migration.

After some algebraical manipulations, it is concluded that regional convergence takes place according to:

$$\ln\left(\frac{y_{i,t+1}}{y_{i,t}}\right) = a_{t+1} + b_i - (1-\alpha\tau)\ln(y_{i,t}) + \alpha(1-\tau)\ln(y_t) + \epsilon_{i,t+1} \tag{7}$$

where

$$b_i = [1 + \alpha(1-\alpha)(1-\tau)]\ln(A_i) \tag{8}$$

captures regional productivity, whose transitory shocks are reflected in the random term $\epsilon_{i,t+1}$.

Expression (7) reflects the standard convergence result of neoclassical growth models that a higher initial level of income per capita reduces subsequent economic growth. Furthermore, and as a new aspect in this theoretical framework, the the speed of convergence decreases with the intensity of mobility frictions. This model also predicts that factor mobility increases the extent to which income at the country level promotes regional growth.

GLLS point out that OLS estimations of the convergence equation in (7) may be biased due to an omitted variables problem, as long as not all productivity determinants are accounted for. Nonetheless, this may not be a serious issue at the regional level because institutional and cultural factors are similar within countries. It has also been widely acknowledged in the regional growth and convergence literature that outcomes in a given region are related to the outcomes and characteristics of its neighbors (Fingleton and López-Bazo, 2006). This implies that the possible presence of spatial dependence between regions is another potential source of bias in this context that should be taken into consideration.

3 Empirical strategy

3.1 Spatial Durbin panel model

Spatial dependence has been incorporated into a neoclassical growth framework in Ertur and Koch (2007) and Fischer (2011) by considering technological spillovers across economies which, in the end, generate spatial externalities. Using a similar setting, Pfaffermayr (2012) highlighted the influence of knowledge spillovers on the estimated speed of convergence. In all these studies, the empirical counterpart of the neoclassical growth model with technological interdependence is a spatial Durbin model[1] (SDM). The distinghishing characteristic of this model is that spatial effects affect both the endogenous and the explanatory variables.

In what follows, we are adopting a static panel SDM version of (7) which, in matrix form for N regions and a given time period t, can be expressed as:

[1]This may explain its extensive use to capture spatial dependence in empirical analyses of regional convergence, see Rey and Le Gallo (2009) and Le Gallo and Fingleton (2014).

$$
\begin{aligned}
Y_t &= \delta\iota_N + \rho W Y_t + X_t\beta + W X_t\gamma + U_t \\
U_t &= \mu + E_t, \qquad t = 1, \ldots, T
\end{aligned}
\tag{9}
$$

where Y_t is the N-dimensional vector for the dependent variable, ι_N is a Nx1 vector of ones and δ is the intercept. W denotes the spatial weights matrix - that determines the interaction scheme between the regions - and ρ is the spatial autoregressive parameter. X_t refers to the Nxk matrix of explanatory variables and WX_t to its spatial lag. β and γ are their corresponding k-dimensional parameter vectors. U_t is a Nx1 vector of error terms, including the time-invariant regional effects in μ and the zero-mean i.i.d. innovation terms in E_t.

Depending on the distributional assumption for the disturbance term, the SDM can be estimated using maximum likelihood (Elhorst, 2003, 2010) or the quasi-maximum likelihood approach (Belotti et al., 2017). In the present paper, we are following this second alternative (i.e., assuming non-normal errors). For the sake of comparability of our results, given that the empirical analysis carried out by GLLS includes time-invariant regressors, expression (9) has been estimated considering the regional effects as random[2]. By proceeding in this way, it is being assumed that there is no correlation between the regional effects and the explanatory variables and that the regions in the sample are representative of a larger population.

3.2 Multiple imputation

In spite of the frameworks developed by Pfaffermayr (2009) and Wang and Lee (2013) to handle unbalanced spatial panels, there is no general approach to cope with this issue in the related literature yet. We are implementing the strategy proposed by Belotti et al. (2017) grounded on the use of multiple imputation (Rubin, 1976, 1987). This simulation-based statistical technique is one of the most commonly applied methods for dealing with missing data which, in short, consists of replacing missing values by multiple sets of plausible values. Rather than filling in a single value, the distribution of the observed data is used to estimate multiple values, reflecting the uncertainty regarding the true value.

[2]The random effects panel data model is usually taken as a point of departure in the spatial econometrics literature because it (i) provides an intermediate solution to the 'all-or-nothing' way of exploting the cross-sectional dimension of the data, and (ii) avoids losing degrees of freedom when the number of the units in the panel is large (Elhorst, 2014).

Imputed values are not intended to represent the 'real' values but to reproduce the variance-covariance structure that would have been observed in the absence of missing information. In doing so, multiple imputation handles missing data in a way resulting in valid statistical inference. Although the theoretical foundations of this methodology were derived under the bayesian paradigm, they are valid from a frequentist point of view. The bayesian approach is used to create the imputations and underlies the combination of the estimated parameters. Once an imputation model is selected, M complete data sets are generated to which the analysis of interest is performed. Finally, the results from these M analyses are combined into a single multiple-imputation result. At this estimation step, coefficients and standard errors are adjusted for the variability between imputations following the combination rules proposed by Rubin (1987).

Multiple imputation is preferable to alternative methodologies - like listwise deletion, pairwise deletion, mean imputation or single imputations - because it only requires that the missing data mechanism is ignorable. In other words, that it is possible to disregard the process that causes missing information, assuming that the data is missing at random[3] (MAR). This implies that the probability that the information is missing does not depend on unobserved data, but may depend on observed variables. For this reason, the larger the number of predictors included in the imputation step the more plausible the MAR assumption. Under the latter, iterative Markov chain Monte Carlo (MCMC) imputation methods are used to simulate imputed values from the posterior predictive distribution of missing data given observed data.

The simulation error will decrease when the number of imputations increases, especially with high fractions of missing information. In the present paper, simulations have been carried out running multiple independent chained equations (MICE), which obtain univariate conditional distributions for each variable from a fully conditional specification of the prediction equations. Even though this technique lacks a rigorous theoretical justification, its flexibility has made of it a frequently encountered choice in practice. More specifically,

[3] Another possibility is to assume that the data are missing completely at random (MCAR). In this case, the probability that information is missing does not depend on observed or unobserved variables, implying that missing values are a random sample of all data values. If the missing data were not at random (MNAR), the reasons for missingness should be accounted for in the imputation model to obtain valid results.

MICE is similar to the Gibbs sampler, a popular MCMC method for simulating data from complicated multivariate distributions[4].

Considering that the complete data $\tilde{Y}$ is a partially observed random sample from a multivariate distribution $P\left(\tilde{Y}|\theta\right)$, it is assumed that the latter is completely specified by the unknown vector of parameters θ (van Buuren and Groothuis-Oudshoorn, 2011). In order to obtain the multivariate distribution of θ, the MICE algorithm samples iteratively, on a variable-by-variable basis, from the p univariate conditional distributions $P\left(\tilde{Y}_1|\tilde{Y}_{-1},\theta_{-1}\right),\dots,P\left(\tilde{Y}_p|\tilde{Y}_{-p},\theta_{-p}\right)$[5]. Variables will be imputed from the most observed to the least observed. Taking a simple drawn from the observed marginal distributions as the starting point, the κth iteration is a Gibbs sampler that draws:

$$\begin{aligned}
\theta_1^{*(\kappa)} &\sim P\left(\theta_1|\tilde{Y}_1^{obs},\tilde{Y}_2^{\kappa-1},\dots,\tilde{Y}_p^{\kappa-1}\right)\\
\tilde{Y}_1^{*(\kappa)} &\sim P\left(\tilde{Y}_1|\tilde{Y}_1^{obs},\tilde{Y}_2^{\kappa-1},\dots,\tilde{Y}_p^{\kappa-1},\theta_1^{*(\kappa)}\right)\\
&\vdots\\
\theta_p^{*(\kappa)} &\sim P\left(\theta_p|\tilde{Y}_p^{obs},\tilde{Y}_1^{\kappa},\dots,\tilde{Y}_{p-1}^{\kappa}\right)\\
\tilde{Y}_p^{*(\kappa)} &\sim P\left(\tilde{Y}_p|\tilde{Y}_p^{obs},\tilde{Y}_1^{\kappa},\dots,\tilde{Y}_p^{\kappa},\theta_p^{*(\kappa)}\right)
\end{aligned} \tag{10}$$

where $\tilde{Y}_l^{(\kappa)}=\left(\tilde{Y}_l^{obs},\tilde{Y}_l^{*(\kappa)}\right)$ denotes the lth imputed variable.

4 Data

4.1 Description

With the aim of empirically testing the predictions of the model presented in Section 2, GLLS constructed a database covering 1,528 regions of 83 countries during the period 1950 – 2010. This data set[6] includes information at the sub-national level of GDP per

[4]See Lee and Carlin (2010) for a comparison of MICE with the iterative multivariate normal (MVN) method. The latter ensures that imputed values are drawn from a specific distribution. Despite MVN may be more attractive from a theoretical point of view, it may not be suitable to formulate a joint model for general data structures. That is to say, this technique is appropriate if the multivariate normal distribution is a reasonable description of the data. Panzera et al. (2016) have recently developed a procedure to deal with missing information in spatial data that combines bayesian interpolation and multiple imputation. However, this new method also relies on the assumption that the distribution of the underlying spatial process is normal.

[5]$\theta_1,\dots,\theta_p$ are specific to their corresponding conditional densities and not necessarily the product of a factorization of $P\left(\tilde{Y}|\theta\right)$.

[6]Available at https://static-content.springer.com/esm/art%3A10.1007%2Fs10887-014-9105-9/MediaObjects/10887_2014_9105_MOESM1_ESM.xlsx

capita (in constant 2005 PPP dollars) and years of schooling for the population aged 15 and older. Regional differences in productivity levels are captured using variables related to geography (area, population density, latitude, average distance to the coast and an indicator of whether the region contains the capital city of the country), natural resource endowments (cumulative oil and gas production until the year 2000) and the disease environment (Kiszewski et al., 2004, 'malaria index'). The database also includes GDP per capita at country level. At this point, it is worth noting that we are handling all this information trying to take advantage of both its cross-sectional and temporal dimensions (i.e., its panel structure). To do so, the data has been grouped in five-year intervals, ruling out consecutive observations over periods longer than seven years. In addition, regional real GDP per capita growth rates have been calculated as annual averages according to their time span.

As noted before, regions should not be considered as independent from a spatial perspective in a regional growth and convergence context. The reason is that technological interdependence generates spatial interactions and spillover effects. Therefore, we incorporate the spatial dimension into the data set constructed by GLLS with a shapefile containing regional boundaries. This geospatial information has been extracted from the GADM database of global administrative areas. Furthermore, individual country shapefiles were merged into a single one where the coordinate reference system is latitude/longitude and the WGS84 datum.

[Insert Table 1 around here]

The main difficulty faced to estimate a spatial panel model with this global sub-national data set is the presence of missing information, which prevents us from having a balanced panel. This problem is encountered in both regional and national GDP per capita levels, population density and years of schooling. Table 1 reports the missing information rates of these variables in four alternative sample periods. These figures show that the percentage of missing data is over 49 per cent during the period 1950 – 2010, what can be considered a high rate so as to pose a reliable multiple imputation analysis. Nonetheless, the incidence of missing data decreases when more recent initial years are taken into account. The highest percentage of missing observations for the sample period 1980 – 2010 is displayed by the years of schooling for the population aged 15 and older (40.7 per cent). These facts lead us to estimate the spatial Durbin panel model with data starting in 1980.

4.2 Preliminary analysis

This subsection study the extent to which changes in data handling and the sample period analyzed and the use of the multiple imputation technique alter standard OLS regression results. With this aim, Table 2 presents cross-sectional estimates of the convergence equation (7) in four alternative samples. The first column of results reproduce those reported by GLLS[7], to be compared to the estimated parameters under our proposed treatment of the data, displayed in the second column. Although the explanatory power is now slightly higher, the number of observations is smaller and different coefficients are obtained.

[Insert Table 2 around here]

The more important disparities are found for national GDP per capita and the indicator reflecting if the region includes the capital city, whose estimated parameters are negative and statistically significant. Despite coefficients of a similar magnitude are obtained for latitude, average distance to the coast, the 'malaria index' and years of schooling, this is not the case for oil and gas production and population density. These variables also change their statistical significance, being now more favorable for population density. Last, but not least, the parameter for the initial level of regional GDP per capita implies an even lower convergence speed than that reported by GLLS.

The explanatory power, estimated coefficients and, as a consequence, implied values for α and τ are stable across samples. The greatest difference is experienced by the variable that reflects the regional disease environment, as its corresponding parameter is not only higher but also statistically significant. In absolute values, the coefficients for the capital city indicator and the proxy for human capital are higher in the most recent period. These findings allow us to conclude that restricting the analysis to the years 1980 – 2010 does not alter the underlying relationship between regional growth and its determinants in this neoclassical framework.

[7]Table 5, specification (4); page 282.

For each sample period analyzed, the last two columns show the estimated parameters when the multiple imputation technique is implemented, fixing $M = 20$. The difference between these pairs of set of results relies on whether or not the panel structure underlying the data has been explicitly taken into account. Together with the growth determinants not affected by the missing information problem, the longitude coordinate of the regional centroid and country and time dummies have also been considered as explanatory variables in the imputation step. It can be stated that the use of this method does not modify the sign nor the statistical significance of the estimated parameters. This is especially the case when the panel nature of the data is controlled for, but at a cost of a higher sampling variance due to missing information, as measured by the average relative variance increase (RVI) and the largest fraction of missing information (FMI).

[Insert Figure 1 around here]

In order to give a visual impression of how the multiple imputation method works in practice, and for the sample covering the years from 1980 to 2010, Figure 1 plots kernel estimates of the density function for the regional growth determinants with missing data. While the black lines correspond to the distributions of the observed data, the grey lines also include the imputed values. The density functions of both the observed and imputed data sets are similar for population density, what might explain that the estimated coefficient for this variable does not change when the presence of missing information is accounted for. More interestingly, both national and regional GDP per capita levels and years of education display higher frequencies in their lower values when imputed values are considered. This is reflecting that the unbalancedness of the panel is mainly driven by the missing information problems suffered by the regions in countries with lower levels of development.

5 Results

Knowledge spillovers and their productivity effects tend to be geographically concentrated (Fischer et al., 2009). Following this argument, spatial dependence has been modelled by means of a spatial weights matrix reflecting geographical proximity. The presence of a non-trivial number of islands in the data hampers the use of a contiguity matrix. As an alternative, we are considering a binary matrix that assigns a value of one to the five nearest neighboring regions, according to their great-circle distances between regional centroids. This matrix has been row-standardized so that each neighbor receives the same weight, enhancing our understanding of the estimated coefficients - because the spatial lags of the variables are the weighted average of neighboring observations - and allowing us to obtain comparable parameters across specifications.

Table 3 shows the estimation results from a panel SDM version of (7) for the period 1980 – 2010. A random effects model has been fitted because, otherwise, time-invariant growth determinants - latitude, distance to the coast, disease environment, oil and gas producion and the capital city indicator - would have been dropped from the regression. In spite of relying on some restrictive assumptions (Elhorst, 2014), the random effects model permits to take the estimates already established both in the literature and in the previous subsection as a frame of reference. The first column of results display the coefficients from the specification where spatial effects concern all explanatory variables. Taking cross-sectional regressions as a baseline, the consideration of a panel data model with spatial dependence makes distance to the coast, the 'malaria index' and population density to lose their relevance in explaining regional growth differences worldwide. Nonetheless, estimated coefficients for latitude, national income per capita and years of schooling remain almost unchanged. We also find that the spatial lags of these three regressors, as well as that of the endogenous variable, are statistically significant.

[Insert Table 3 around here]

The result that the growth rate in a region is affected by the GDP per capita levels and growth rates in its neighbors reflects spatial dependence in regional economic activity. Further, and in line with recent related studies (Sanso-Navarro et al., 2016), our findings provide evidence that human capital generates spatial spillovers. The estimated coefficients under this first specification imply lower values for the regional income share that remunerates capital and for mobility costs, more easily satisfying the constraints imposed in the theoretical model. It is worth emphasizing that the parameter that corresponds to the lag of regional GDP per capita is, in absolute value, more than 3.5 times higher. This result corroborates our initial suspicion that the low convergence rates found by GLLS may be determined by not properly accounting for the presence of spatial dependence in a neoclassical regional growth context.

The second column of results in Table 3 reports the estimation results when only time-varying regressors are considered to generate spatial effects. Although the regional disease environment is again significant, latitude and the capital city indicator are not different to zero from a statistical point of view. Neither the spatial lags nor the speed of convergence are affected by this change in the model specification. The third and fourth columns show the results obtained when country and time fixed effects are, consecutively, also included as explanatory variables. The changes mainly affect latitude and population density - with negative and statistically significant coefficients - and the spatial lag of human capital. In absolute values, the parameters for the spatial lags of regional income per capita levels and growth rates are now lower. Similarly to related studies (Barro, 2015), the introduction of country fixed effects increases estimated regional convergence rates. As a novelty, this seems to be also the case of time fixed effects when the panel structure of the data is taken into account.

The interpretation of parameter estimates in spatial regression models is more complicated than in standard OLS regressions due to the feedback effects generated by the dependence relationships in the spatial lag terms. Nevertheless, this is a valuable feature of spatial models that permits the quantification of spillover effects which, in the case of the SDM, are global. The direct impact of a variation in an explanatory variable in region i includes the effects that this change exerts on its growth rate as well as on that of its neighbors. This direct impact also reflects that these effects in neighboring regions affect region i. These 'own' spillover effects are heterogenous in the presence of spatial autocorrelation because the interaction terms in the spatial weights matrix are different. However, the magnitude of these feedback effects are relatively small compared to the magnitude of the corresponding parameter for that variable in β. The indirect impact of a change in a regressor in a given region reflects the effects that this variation has on the growth rate of its neighbors.

[Insert Table 4 around here]

Given that the effects generated by changes in the regressors will differ across regions, LeSage and Pace (2009) propose summarizing the impacts through the average values of the direct and indirect effects from changes in all regions. Table 4 shows the marginal effects of regional growth determinants obtained from the panel SDM estimation calculated using this method. These estimates do not depend in a great extent on whether or not it is considered that all regional growth determinants induce spatial effects. In addition, the sign of the average direct effects tends to coincide with that of the estimated parameters[8]. The sign and magnitude of the direct impact for the initial level of regional GDP per capita corroborate that the speed of convergence is higher when spatial effects are controlled for. Although the average direct effect for national income is small, its positive sign reflects that regional convergence is higher is less developed countries. Years of education is also found to be a relevant variable to explain regional growth differences. Furthermore, the indirect impact of this proxy for human capital suggests that it generates positive spillovers. This seems to be also the case of the initial level of regional GDP per capita. That is to say, regions gain growth benefits from being surrounded by richer regions. The disease environment and population density also display positive indirect effects. Nevertheless, their statistical significance for population density disappear when spatial effects are considered to be caused only by time-varying variables.

[Insert Table 5 around here]

One criticism of spatial econometric models is that the selection of the weights matrix is somehow arbitrary. Bearing in mind that there is no clear guidance on how to specify this matrix, it is a common practice to check whether the estimation results are robust to its choice. Table 5 reports the results obtained when the binary contiguity matrix reflects the three or seven nearest neighbors. The coefficients present fewer changes when spatial effects are only related to time-varying regressors. In any case, shifts in the estimated parameters neither affect our main variables of interest - national and regional initial income per capital levels and human capital - nor their spatial lags. It can also be observed that the coefficient for the endogenous spatial lag increases with the number of neighbors. On the contrary, spatial effects from population density are statistically significant only when a

[8]The differences between parameter estimates and direct impact estimates represent the feedback effects passing through neighboring regions and back to the origin itself.

small number of regions is taken into account. Furthermore, an increase in the number of imputations is expected to reduce the sampling error. Figures displayed in Table 5 show that generating 50 or 100 imputed data sets do not alter parameter estimates. All these findings lead us to conclude that our results are robust to both the specification of the spatial weights matrix and the number of imputations.

Richer countries should experience higher regional convergence rates as they have better capital markets. In this line, GLLS point towards poorly developed financial markets as one of the elements slowing down factor mobility. Following these arguments, Table 6 reports panel SDM estimates by country groups. The first three columns of results distinguish the countries according to their income levels, using the World Bank classification[9] for the year 1995. Regardless of the income group, the estimated regional speed of convergence is higher when less heterogeneous subsamples of countries are considered. Although we find that regions in medium and high-income countries present faster convergence rates, this relationship is non-linear. Actually, our results suggest that there exists an inverted-U shape relationship between income and the speed of convergence. The endowment of natural resources exerts a great influence on regional growth in low-income countries. On the contrary, this is the only group where the initial level of GDP per capita in neighboring regions and years of schooling are not statistically significant. The highest parameter estimate for the endogenous spatial lag is found in the subsample of the more developed countries. The latter do not display significant spatial effects from human capital.

[Insert Table 6 around here]

Differences of the growth and convergence processes across regions in different subsamples can be further analyzed using the index proposed by Svirydzenka (2016). At country level, this new index of financial development[10] takes into account its multi-dimensional nature by combining information on the depth, access and efficiency of both financial markets and institutions. The last three columns of results in Table 6 show the estimated coefficients when countries are classified according to the value of the financial development index in the whole distribution for the year 1995. The theoretical framework adopted in

[9] Available at http://databank.worldbank.org/data/download/site-content/OGHIST.xls

[10] Available at http://www.imf.org/~/media/Websites/IMF/imported-datasets/external/pubs/ft/wp/2016/Data/_wp1605.ashx

this paper does not have a good fit in the regions whose countries have an index of financial development in the lower quartile. In particular, estimated parameters for the initial level of regional income and its spatial lag are not significant. Further, the non-linear relationship between the convergence rate and financial development is more apparent than that found when countries are distinguished by income levels.

6 Concluding remarks

This paper incorporates spatial dependence into the neoclassical regional growth framework developed by Gennaioli et al. (2014). The main aim is to study the role played by technological diffusion between regions and productive externalities. After incorporating the geographical dimension into their newly constructed database, the empirical analysis is based on the implementation of multiple imputation techniques to deal with the presence of missing observations. Our results from a spatial Durbin panel model show that accounting for spatial effects leads to a higher estimated convergence rate at the sub-national level. This finding is robust to changes in the model specification, the choice of the spatial weights matrix and the number of imputations. We also obtain evidence of an inverted-U shape relationship between the levels of national income and financial development and the regional speed of convergence. Therefore, not only spatial effects but also non-linearities should be incorporated into the neoclassical regional growth framework with imperfect factor mobility.

References

Barro RJ. 2015. Convergence and modernisation. *The Economic Journal* **125**: 911–942. doi: `10.1111/ecoj.12247`

Belotti F, Hughes G, Mortari AP. 2017. Spatial panel data models using Stata. *Stata Journal* : Forthcoming.

Benos N, Karagiannis S, Karkalakos S. 2015. Proximity and growth spillovers in European regions: The role of geographical, economic and technological linkages. *Journal of Macroeconomics* **43**: 124–139. doi: `10.1016/j.jmacro.2014.10.003`

Breinlich H, Ottaviano GI, Temple JR. 2014. Regional growth and regional decline. In Aghion P, Durlauf SN (eds.) *Handbook of Economic Growth*, volume 2. Amsterdam:

Elsevier, 683–779. doi: 10.1016/B978-0-444-53540-5.00004-5

Elhorst JP. 2003. Specification and estimation of spatial panel data models. *International Regional Science Review* **26**: 244–268. doi: 10.1177/0160017603253791

Elhorst JP. 2010. Spatial panel data models. In Fischer MM, Getis A (eds.) *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin: Springer, 377–407. doi: 10.1007/978-3-642-03647-7-19

Elhorst JP. 2014. *Spatial econometrics: From cross-sectional data to spatial panels*. Berlin: Springer. doi: 10.1007/978-3-642-40340-8

Ertur C, Koch W. 2007. Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics* **22**: 1033–1062. doi: 10.1002/jae.963

Fingleton B, López-Bazo E. 2006. Empirical growth models with spatial effects. *Papers in Regional Science* **85**: 177–198. doi: 10.1111/j.1435-5957.2006.00074.x

Fischer MM. 2011. A spatial Mankiw–Romer–Weil model: Theory and evidence. *The Annals of Regional Science* **47**: 419–436. doi: 10.1007/s00168-010-0384-6

Fischer MM, Bartkowska M, Riedl A, Sardadvar S, Kunnert A. 2009. The impact of human capital on regional labor productivity in Europe. *Letters in Spatial and Resource Sciences* **2**: 97–108. doi: 10.1007/s12076-009-0027-7

Gennaioli N, La Porta R, Lopez-de Silanes F, Shleifer A. 2013. Human capital and regional development. *Quarterly Journal of Economics* **128**: 105–164. doi: 10.1093/qje/qjs050

Gennaioli N, La Porta R, Lopez De Silanes F, Shleifer A. 2014. Growth in regions. *Journal of Economic Growth* **19**: 259–309. doi: 10.1007/s10887-014-9105-9

Kiszewski A, Mellinger A, Speilman A, Malaney P, Sachs SE, Sachs J. 2004. A global index representing the stability of malaria transmission. *The American Journal of Tropical Medicine and Hygiene* **70**: 486–498.

Le Gallo J, Fingleton B. 2014. Regional growth and convergence empirics. In Fischer MM, Nijkamp P (eds.) *Handbook of Regional Science*. Berlin: Springer. ISBN 978-3-642-23430-9, 291–315. doi: 10.1007/978-3-642-23430-9-17

Lee KJ, Carlin JB. 2010. Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology* **171**: 624–632. doi: 10.1093/aje/kwp425

LeSage J, Pace R. 2009. *Introduction to spatial econometrics.* Statistics: A Series of Textbooks and Monographs. Boca Raton: CRC Press.

Mitton T. 2016. The wealth of subnations: Geography, institutions, and within-country development. *Journal of Development Economics* **118**: 88–111. doi: 10.1016/j.jdeveco.2015.09.002

Panzera D, Benedetti R, Postiglione P. 2016. A bayesian approach to parameter estimation in the presence of spatial missing data. *Spatial Economic Analysis* **11**: 201–218. doi: 10.1080/17421772.2016.1102962

Pfaffermayr M. 2009. Maximum likelihood estimation of a general unbalanced spatial random effects model: A Monte Carlo study. *Spatial Economic Analysis* **4**: 467–483. doi: 10.1080/17421770903317645

Pfaffermayr M. 2012. Spatial convergence of regions revisited: A spatial maximum likelihood panel approach. *Journal of Regional Science* **52**: 857–873. doi: 10.1111/j.1467-9787.2012.00773.x

Rey SJ, Le Gallo J. 2009. Spatial analysis of economic convergence. In Mills TC, Patterson K (eds.) *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics.* London: Palgrave Macmillan UK, 1251–1290. doi: 10.1007/978-0-230-24440-5-27

Rubin DB. 1976. Inference and missing data. *Biometrika* **63**: 581–592. doi: 10.1093/biomet/63.3.581

Rubin DB. 1987. *Multiple imputation for nonresponse in surveys.* New York: Wiley. doi: 10.1002/9780470316696

Sanso-Navarro M, Vera-Cabello M, Ximénez-De-Embún DP. 2016. Human capital spillovers and regional development. *Journal of Applied Econometrics* : Early View. doi: 10.1002/jae.2541

Svirydzenka K. 2016. Introducing a new broad-based index of financial development. IMF Working Papers 16/5, International Monetary Fund.

van Buuren S, Groothuis-Oudshoorn K. 2011. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software* **45**: 1–67. doi: 10.18637/jss.v045.i03

Wang W, Lee LF. 2013. Estimation of spatial panel data models with randomly missing data in the dependent variable. *Regional Science and Urban Economics* **43**: 521–538. doi: 10.1016/j.regsciurbeco.2013.02.001

Tables and figures

Table 1: Missing observations (%) across alternative sample periods.

	1950-2010	1960-2010	1970-2010	1980-2010
L.lngdppc	54.9	48.4	41.2	31.3
L.lngdppcnat	54.9	48.4	41.2	31.3
lnpopdens	49.5	42.9	34.6	26.0
yearsed	59.3	52.6	46.1	40.7
growth	61.9	55.7	48.3	39.1

Table 2: Determinants of regional growth. Cross-sectional OLS estimation.

	1950-2010				1960-2010			1970-2010			1980-2010		
	GLLS2014	raw data	imputed	imputed panel	raw data	imputed	imputed panel	raw data	imputed	imputed panel	raw data	imputed	imputed panel
L.lngdppc	-0.0116***	-0.0074***	-0.0058***	-0.0076***	-0.0071***	-0.0050***	-0.0071***	-0.0071***	-0.0060***	-0.0070***	-0.0082***	-0.0076***	-0.0095***
	(0.0040)	(0.0010)	(0.0008)	(0.0010)	(0.0010)	(0.0008)	(0.0010)	(0.0010)	(0.0009)	(0.0010)	(0.0011)	(0.0009)	(0.0011)
L.lngdppcnat	0.0016	-0.0039***	-0.0047***	-0.0057***	-0.0034***	-0.0048***	-0.0056***	-0.0032***	-0.0043***	-0.0049***	-0.0041***	-0.0048***	-0.0046***
	(0.0022)	(0.0009)	(0.0008)	(0.0009)	(0.0009)	(0.0007)	(0.0009)	(0.0009)	(0.0009)	(0.0010)	(0.0010)	(0.0008)	(0.0010)
latitude	0.0001*	0.0003***	0.0002***	0.0004***	0.0003***	0.0002***	0.0004***	0.0003***	0.0002***	0.0003***	0.0004***	0.0002***	0.0003***
	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
invdistcoast	0.0397***	0.0347***	0.0312***	0.0403***	0.0320***	0.0267***	0.0398***	0.0308***	0.0271***	0.0302***	0.0299***	0.0281***	0.0197***
	(0.0116)	(0.0073)	(0.0060)	(0.0065)	(0.0074)	(0.0062)	(0.0065)	(0.0077)	(0.0064)	(0.0068)	(0.0083)	(0.0070)	(0.0072)
malaria	0.0001	0.0002	0.0002	0.0009***	0.0002	0.0003**	0.0008***	0.0004**	0.0003**	0.0010***	0.0005***	0.0003**	0.0014***
	(0.0005)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)
lnoilgas	0.1026**	0.0506	0.0553	0.0261	0.0529	0.0686	0.0232	0.0511	0.0568	0.0211	0.0491	0.0610	0.0401
	(0.0489)	(0.0482)	(0.0437)	(0.0448)	(0.0497)	(0.0467)	(0.0507)	(0.0536)	(0.0506)	(0.05070	(0.0603)	(0.0556)	(0.0605)
lnpopdens	0.0004	0.0016***	0.0012***	0.0017***	0.0014***	0.0012***	0.0016***	0.0012***	0.0010***	0.0013***	0.0016***	0.0011***	0.0016***
	(0.0005)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)
capital	0.0047**	-0.0053***	-0.0041***	-0.0044***	-0.0049***	-0.0038***	-0.0038***	-0.0047***	-0.0041***	-0.0039**	-0.0066***	-0.0054***	-0.0055***
	(0.0020)	(0.0016)	(0.0013)	(0.0016)	(0.0017)	(0.0014)	(0.0013)	(0.0018)	(0.0014)	(0.0015)	(0.0019)	(0.0015)	(0.0017)
yearsed	0.0017*	0.0012***	0.0013***	0.0019***	0.0010***	0.0012***	0.0016***	0.0011***	0.0013***	0.0018***	0.0020***	0.0022***	0.0031***
	(0.0010)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
constant	0.0931***	0.1009***	0.0981***	0.1091***	0.0956***	0.0918***	0.1063***	0.0932***	0.0966***	0.1009***	0.0997***	0.1063***	0.1090***
	(0.0344)	(0.0039)	(0.0030)	(0.0053)	(0.0039)	(0.0032)	(0.0035)	(0.0041)	(0.0034)	(0.0035)	(0.0042)	(0.0034)	(0.0041)
Observations	6,145	5,654	9,267	18,324	5,510	8,730	15,270	5,115	7,991	12,216	4,438	6,782	9,162
R^2	0.08	0.11			0.09			0.09			0.11		
Average RVI			0.3568	2.7910		0.3032	1.8221		0.3569	1.3626		0.2435	0.6711
Largest FMI			0.4262	0.8164		0.3139	0.7793		0.4144	0.7264		0.2869	0.5267
Implied α	0.9900	0.9887	0.9895	0.9867	0.9895	0.9902	0.9873	0.9897	0.9897	0.9881	0.9877	0.9876	0.9859
Implied τ	0.9984	1.0039	1.0047	1.0058	1.0034	1.0048	1.0057	1.0032	1.0043	1.0050	1.0042	1.0049	1.0047

Note: Robust standard errors are reported in parentheses. Missing observations have been controlled for using 20 multiple imputations. *** p<0.01; ** p<0.05; * p<0.10.

Table 3: Determinants of regional growth, 1980-2010. Panel SDM estimation.

	(1)	(2)	(3)	(4)
L.lngdppc	-0.0348***	-0.0342***	-0.0388***	-0.0550***
	(0.0028)	(0.0024)	(0.0016)	(0.0018)
L.lngdppcnat	-0.0024**	-0.0030***	-0.0050***	-0.0034***
	(0.0011)	(0.0010)	(0.0008)	(0.0007)
latitude	0.0005**	0.0000	-0.0003***	-0.0001**
	(0.0002)	(0.0000)	(0.0001)	(0.0001)
invdistcoast	0.0148	0.0078	-0.0047	-0.0040
	(0.0098)	(0.0081)	(0.0054)	(0.0050)
malaria	0.0003	0.0005***	0.0005***	0.0003**
	(0.0003)	(0.0001)	(0.0002)	(0.0002)
lnoilgas	-0.0316	0.0059	0.0096	-0.0082
	(0.0727)	(0.0635)	(0.0401)	(0.0369)
lnpopdens	-0.0001	-0.0003	-0.0009***	-0.0006***
	(0.0004)	(0.0003)	(0.0002)	(0.0002)
capital	-0.0029*	-0.0026	-0.0024*	-0.0011
	(0.0017)	(0.0019)	(0.0013)	(0.0012)
yearsed	0.0028***	0.0030***	0.0048***	0.0030***
	(0.0005)	(0.0004)	(0.0004)	(0.0004)
constant	0.0388***	0.0407***	0.2743***	0.4259***
	(0.0042)	(0.0041)	(0.0121)	(0.0151)
W*growth	0.7566***	0.7564***	0.6218***	0.5302***
	(0.0083)	(0.0079)	(0.0110)	(0.0115)
W*L.lngdppc	0.0303***	0.0299***	0.0126***	0.0081***
	(0.0027)	(0.0024)	(0.0015)	(0.0012)
W*L.lngdppcnat	0.0014	0.0016	-0.0041***	-0.0013
	(0.0014)	(0.0015)	(0.0012)	(0.0011)
W*latitude	-0.0005**			
	(0.0002)			
W*invdistcoast	-0.0267			
	(0.0163)			
W*malaria	0.0003			
	(0.0004)			
W*lnoilgas	0.0922			
	(0.2048)			
W*lnpopdens	0.0007	0.0006	-0.0000	-0.0002
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
W*capital	0.0005			
	(0.0040)			
W*yearsed	-0.0011**	-0.0012**	0.0028***	0.0004
	(0.0005)	(0.0005)	(0.0004)	(0.0004)
Observations	9,162	9,162	9,162	9,162
Country FEs	No	No	Yes	Yes
Time FEs	No	No	No	Yes
Average RVI	0.6376	0.6784	0.7669	0.9212
Largest FMI	0.8138	0.7477	0.7368	0.7265
Implied α	0.9628	0.9628	0.9562	0.9416
Implied τ	1.0025	1.0031	1.0052	1.0036

Note: Robust standard errors are reported in parentheses. Missing values have been controlled for using 20 multiple imputations. W is a five nearest neighbors row-standardized spatial weights matrix. *** $p<0.01$; ** $p<0.05$; * $p<0.10$.

Table 4: Marginal effects of regional growth determinants, 1980-2010. Panel SDM estimation.

	All variables (1)			Time-varying variables (2)		
	Direct	Indirect	Total	Direct	Indirect	Total
L.lngdppc	-0.0338***	0.0154***	-0.0184***	-0.0332***	0.0155***	-0.0177***
	(0.0027)	(0.0043)	(0.0045)	(0.0023)	(0.0042)	(0.0045)
L.lngdppcnat	-0.0025**	-0.0017	-0.0041	-0.0032***	-0.0027	-0.0059
	(0.0011)	(0.0040)	(0.0043)	(0.0010)	(0.0042)	(0.0044)
latitude	0.0005**	-0.0004	0.0001	0.0000	0.0001	0.0001
	(0.0002)	(0.0003)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
invdistcoast	0.0100	-0.0729	-0.0629	0.0090	0.0220	0.0310
	(0.0095)	(0.0511)	(0.0538)	(0.0095)	(0.0233)	(0.0329)
malaria	0.0005	0.0021***	0.0026***	0.0006***	0.0016***	0.0022***
	(0.0003)	(0.0006)	(0.0006)	(0.0002)	(0.0004)	(0.0006)
lnoilgas	-0.0095	0.1601	0.1506	0.0063	0.0151	0.0214
	(0.0864)	(0.8038)	(0.8536)	(0.0763)	(0.1871)	(0.2634)
lnpopdens	0.0000	0.0022**	0.0023*	-0.0001	0.0017	0.0016
	(0.0004)	(0.0011)	(0.0012)	(0.0003)	(0.0011)	(0.0012)
capital	-0.0034	-0.0069	-0.0104	-0.0032	-0.0077	-0.0109
	(0.0022)	(0.0154)	(0.0170)	(0.0022)	(0.0055)	(0.0077)
yearsed	0.0031***	0.0040***	0.0071***	0.0032***	0.0039***	0.0071***
	(0.0005)	(0.0007)	(0.0008)	(0.0004)	(0.0009)	(0.0008)

Note: Standard errors are reported in parentheses. The empirical distribution of these marginal effects have been obtained using the Monte Carlo procedure proposed by LeSage and Pace (2009) (100 replications). *** p<0.01; ** p<0.05; * p<0.10.

Table 5: Determinants of regional growth. Panel SDM estimation, robustness check.

	All variables (1)				Time-varying variables (2)			
	knn3	knn7	M = 50	M = 100	knn3	knn7	M = 50	M = 100
L.lngdppc	-0.0348***	-0.0337***	-0.0346***	-0.0343***	-0.0346***	-0.0334***	-0.0343***	-0.0340***
	(0.0025)	(0.0025)	(0.0030)	(0.0027)	(0.0025)	(0.0025)	(0.0030)	(0.0028)
L.lngdppcnat	-0.0031***	-0.0026***	-0.0028***	-0.0028***	-0.0030***	-0.0026***	-0.0028***	-0.0028***
	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0010)	(0.0009)	(0.0009)	(0.0010)
latitude	0.0004	0.0004*	0.0005**	0.0005***	0.0000	0.0000	0.0000	0.0000
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
invdistcoast	0.0197**	0.0096	0.0141	0.0147	0.0127	0.0028	0.0062	0.0061
	(0.0099)	(0.0098)	(0.0101)	(0.0100)	(0.0082)	(0.0083)	(0.0083)	(0.0083)
malaria	0.0007*	-0.0004	0.0003	0.0003	0.0006***	0.0004***	0.0005***	0.0005***
	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
lnoilgas	-0.0136	-0.0502	-0.0305	-0.0352	0.0116	0.0011	0.0108	0.0076
	(0.0708)	(0.0725)	(0.0734)	(0.0730)	(0.0639)	(0.0651)	(0.0653)	(0.0651)
lnpopdens	-0.0002	-0.0003	-0.0002	-0.0002	-0.0002	-0.0003	-0.0002	-0.0002
	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
capital	-0.0021	-0.0025	-0.0031	-0.0028	-0.0021	-0.0026	-0.0031	-0.0028
	(0.0018)	(0.0019)	(0.0020)	(0.0018)	(0.0018)	(0.0019)	(0.0020)	(0.0020)
yearsed	0.0030***	0.0030***	0.0030***	0.0029***	0.0030***	0.0030***	0.0031***	0.0030***
	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0004)
constant	0.0504***	0.0325***	0.0396***	0.0391***	0.0509***	0.0350***	0.0409***	0.0404***
	(0.0042)	(0.0042)	(0.0041)	(0.0042)	(0.0042)	(0.0041)	(0.0041)	(0.0041)
W*growth	0.6864***	0.7935***	0.7562***	0.7567***	0.6859***	0.7931***	0.7557***	0.7557***
	(0.0087)	(0.0075)	(0.0082)	(0.0082)	(0.0087)	(0.0075)	(0.0082)	(0.0080)
W*L.lngdppc	0.0294***	0.0307***	0.0302***	0.0300***	0.0291***	0.0300***	0.0297***	0.0295***
	(0.0024)	(0.0026)	(0.0029)	(0.0028)	(0.0023)	(0.0026)	(0.0029)	(0.0027)
W*L.lngdppcnat	0.0015	0.0009	0.0017	0.0015	0.0015	0.0010	0.0017	0.0016
	(0.0014)	(0.0016)	(0.0013)	(0.0015)	(0.0014)	(0.0015)	(0.0013)	(0.0014)
W*latitude	-0.0004	-0.0004*	-0.0005***	-0.0005**				
	(0.0003)	(0.0002)	(0.0003)	(0.0002)				
W*invdistcoast	-0.0193	-0.0265	-0.0248	-0.0257				
	(0.0149)	(0.0174)	(0.0164)	(0.0164)				
W*malaria	-0.0002	0.0009***	0.0003	0.0003				
	(0.0004)	(0.0004)	(0.0003)	(0.0003)				
W*lnoilgas	0.0603	0.2459	0.0974	0.1123				
	(0.1671)	(0.2556)	(0.2107)	(0.2098)				
W*lnpopdens	0.0008*	0.0007	0.0007	0.0008*	0.0007*	0.0007	0.0006	0.0006
	(0.0004)	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)
W*capital	0.0015	0.0026	0.0014	0.0011				
	(0.0033)	(0.0051)	(0.0041)	(0.0041)				
W*yearsed	-0.0010*	-0.0014***	-0.0013***	-0.0012***	-0.0010**	-0.0015***	-0.0014***	-0.0013***
	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0005)
Observations	9,162	9,162	9,162	9,162	9,162	9,162	9,162	9,162
Average RVI	0.6842	0.5941	0.5487	0.6624	0.7246	0.6380	0.6303	0.6537
Largest FMI	0.7434	0.7309	0.7789	0.7372	0.7435	0.7307	0.7801	0.7562
Implied α	0.9621	0.9637	0.9626	0.9629	0.9624	0.9640	0.9629	0.9632
Implied τ	1.0032	1.0027	1.0029	1.0029	1.0031	1.0027	1.0029	1.0029

Note: Robust standard errors are reported in parentheses. *** p<0.01; ** p<0.05; * p<0.10.

Table 6: Determinants of regional growth by country groups. Panel SDM estimation.

	Income groups (World Bank)			Financial development (Svirydzenka, 2016)		
	Low	Medium	High	Lower quartile	Medium	Upper quartile
L.lngdppc	-0.0535***	-0.0808***	-0.0558***	0.0062	-0.0891***	-0.0185***
	(0.0083)	(0.0052)	(0.0044)	(0.0188)	(0.0065)	(0.0023)
L.lngdppcnat	0.0032	-0.0019*	-0.0014	-0.0003	-0.0003	-0.0004
	(0.0022)	(0.0011)	(0.0011)	(0.0067)	(0.0013)	(0.0009)
latitude	0.0000	0.0012**	-0.0000	-0.0002	0.0014**	-0.0001
	(0.0008)	(0.0005)	(0.0001)	(0.0014)	(0.0006)	(0.0001)
invdistcoast	-0.0029	0.0557**	0.0024	-0.0654	0.0083	0.0024
	(0.0302)	(0.0223)	(0.0047)	(0.1178)	(0.0248)	(0.0039)
malaria	-0.0009	-0.0002	-0.0004	0.0004	-0.0002	-0.0003
	(0.0006)	(0.0006)	(0.0023)	(0.0008)	(0.0007)	(0.0002)
lnoilgas	11.3121	-0.1002	0.0276	0.3867	-0.1453	0.0173
	(6.9111)	(0.1383)	(0.0294)	(1.0334)	(0.1640)	(0.0397)
lnpopdens	-0.0012	-0.0006	0.0000	0.0003	-0.0001	-0.0000
	(0.0009)	(0.0005)	(0.0004)	(0.0030)	(0.0006)	(0.0003)
capital	-0.0104*	-0.0031	-0.0004	-0.0022	-0.0055	-0.0011
	(0.0056)	(0.0034)	(0.0011)	(0.0099)	(0.0039)	(0.0010)
yearsed	0.0006	0.0013**	0.0008**	0.0031	0.0018***	0.0009***
	(0.0009)	(0.0006)	(0.0004)	(0.0030)	(0.0006)	(0.0003)
constant	0.0308**	0.0992***	0.0466***	0.0448*	0.0737***	0.0218***
	(0.0133)	(0.0128)	(0.0064)	(0.0254)	(0.0126)	(0.0032)
W*growth	0.8161***	0.7777***	0.9014***	0.6259***	0.7499***	0.8725***
	(0.0159)	(0.0101)	(0.0077)	(0.0424)	(0.0114)	(0.0086)
W*L.lngdppc	0.0467	0.0710***	0.0527***	-0.0027	0.0785***	0.0166***
	(0.0089)	(0.0045)	(0.0048)	(0.0176)	(0.0063)	(0.0024)
W*L.lngdppcnat	-0.0001	-0.0012	-0.0004	-0.0078	-0.0007	-0.0000
	(0.0037)	(0.0017)	(0.0015)	(0.0076)	(0.0021)	(0.0011)
W*latitude	-0.0005	-0.0013**	0.0001	0.0007	-0.0016***	0.0001
	(0.0008)	(0.0005)	(0.0001)	(0.0016)	(0.0006)	(0.0001)
W*invdistcoast	-0.0794	-0.0647*	-0.0026	0.1754	-0.0726*	0.0033
	(0.0506)	(0.0343)	(0.0081)	(0.1486)	(0.0412)	(0.0070)
W*malaria	0.0010	0.0004	0.0029	-0.0002	0.0013	0.0004*
	(0.0006)	(0.0007)	(0.0042)	(0.0008)	(0.0008)	(0.0002)
W*lnoilgas	-1.8172	-0.4605	0.0400	-1.3164	-0.1191	0.0020
	(15.3326)	(0.4862)	(0.0796)	(3.0761)	(0.5897)	(0.0938)
W*lnpopdens	0.0013	0.0008	-0.0000	-0.0040	0.0009	0.0000
	(0.0011)	(0.0007)	(0.0004)	(0.0035)	(0.0007)	(0.0003)
W*capital	-0.0492***	-0.0095	-0.0020	0.0147	0.0038	-0.0024
	(0.0117)	(0.0071)	(0.0025)	(0.0175)	(0.0085)	(0.0024)
W*yearsed	0.0026**	0.0015**	-0.0006	-0.0037	0.0030***	-0.0006*
	(0.0012)	(0.0006)	(0.0004)	(0.0029)	(0.0007)	(0.0003)
Observations	1,506	5,436	2,220	702	4,620	3,840
Average RVI	0.7976	0.7670	0.4779	1.9880	0.8097	0.6946
Largest FMI	0.8183	0.7967	0.8008	0.9069	0.8806	0.8994
Implied α	0.9497	0.9173	0.9428	1.0059	0.9106	0.9811
Implied τ	0.9966	1.0021	1.0015	1.0003	1.0003	1.0004

Note: Robust standard errors are reported in parentheses. Missing values have been controlled for using 20 multiple imputations. W is a five nearest neighbors row-standardized spatial weights matrix. *** p<0.01; ** p<0.05; * p<0.10.

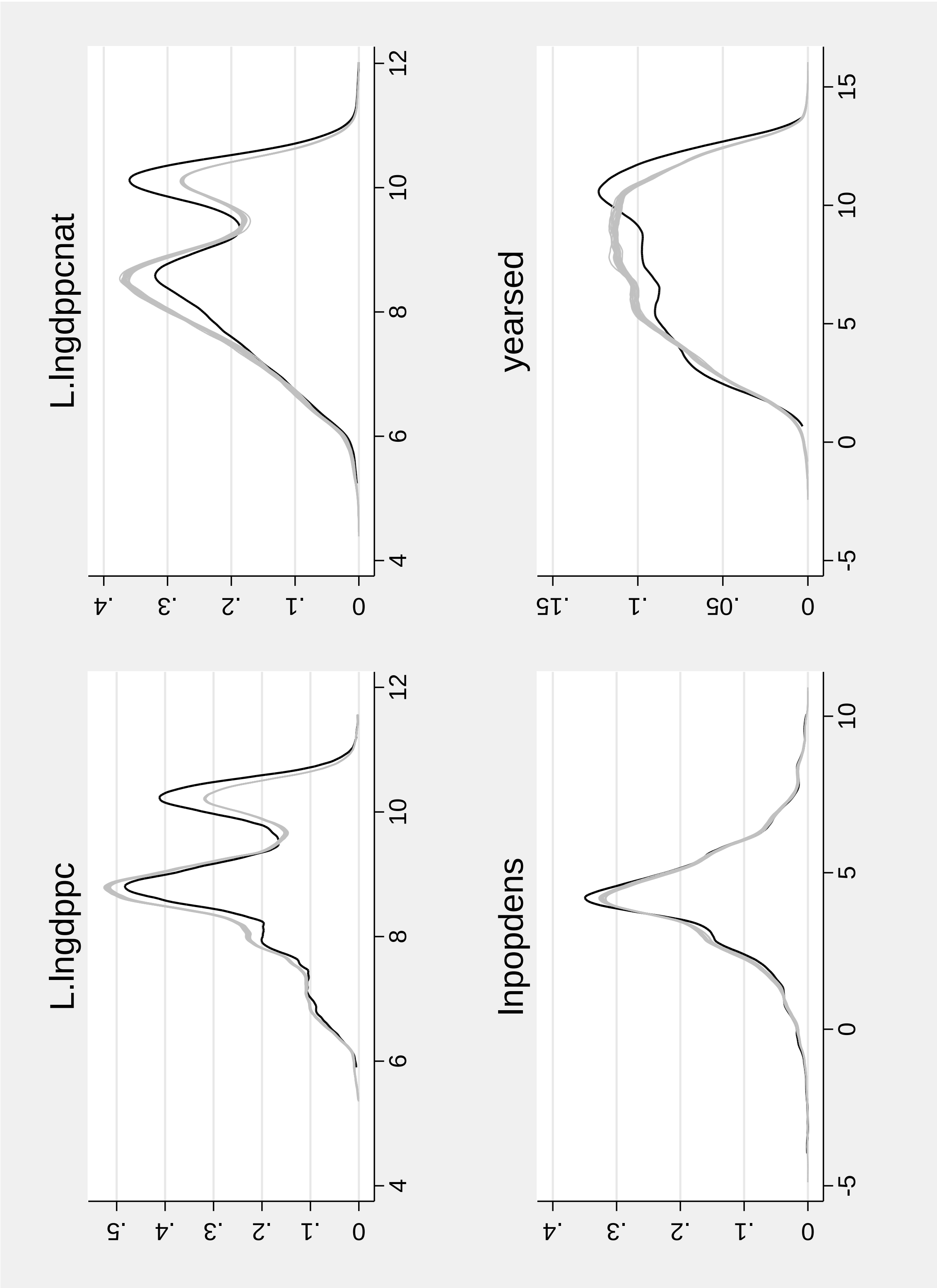

Figure 1: Kernel density estimation. Observed (black) and imputed values (light grey, 20 panel imputations) of regional growth determinants with missing information, 1980 – 2010.