Market Potential, panel data, and aggregate fluctuations: All that glitters is not gold

Bruna, Fernando

Universidade da Coruña

29 August 2024

Market Potential, panel data, and aggregate fluctuations: All that glitters is not gold

Fernando Bruna*

* C&D Research Group, Department of Economics, Faculty of Sociology, Campus de A Coruña, 15071 A Coruña, Spain. E-mail: f.bruna@udc.es.
https://orcid.org/0000-0002-6433-0987

Abstract

The New Economic Geography (NEG) provides a historical explanation for the spatial agglomeration of economic activity. One of its predictions, the 'wage equation', relates regional income to market accessibility. Although the NEG is a long-term theory, empirical literature has tested it using panel data methods, which capture short-term relationships between temporal changes in variables. For a sample of European regions, I show that panel data estimations of the wage equation identify only potential spillover effects of the European business cycle on the synchronic evolution of regional per capita income. That is, the panel data results are not due to the mechanisms proposed by the NEG. The paper concludes with a cautionary note about misinterpretation of panel data estimations.

Keywords: NEG, agglomeration, wage equation, fixed effects, first differences, European cycle

JEL Classification: C18, C23, F12, R12, E32

Draft, August 26, 2024

1 Introduction

Krugman's (1991, 1993) development of the New Economic Geography (NEG) provided an explanation of economic agglomeration—that is, of the formation of clusters in the spatial distribution of economic activity. The so-called 'wage equation' of the NEG is a long-run prediction relating higher income to locations with higher Market Potential, an indicator of the accessibility and market size of the other regions. The cross-sectional form of the equation has been widely studied in the empirical literature (Redding 2011). More generally, location theory establishes "fundamental determinants" of economic activity (Redding and Venables 2004) based on long-term consequences of centrifugal and centripetal agglomeration forces. These historical explanations should thus be tested using cross-sectional data (Baltagi and Griffin 1984), not panels of data at intervals of one or a few years.

Starting with Hanson (2005) and Mion (2004), however, this equation has been estimated using panel data. The author who has studied the NEG wage-type equation most using panel data techniques is Bernard Fingleton (e.g., Fingleton (2008), Baltagi et al. (2014)). Other important articles on the empirics of the NEG have used panel data: Breinlich (2006), Boulhol and de Serres (2010), de Sousa and Poncet (2011), Head and Mayer (2011). Their results are not satisfactory. Panel data models are designed to capture short-term relationships. Why is an indicator of Market Potential statistically significant in panel data models if that indicator is designed to synthesize forces underlying location decisions over centuries?

The goal of this paper is to provide an explanation for this anomaly. It employs European regional data to study the properties of the time series of the time-demeaned data used to derive fixed effects panel data estimates. I illustrate the intuitions behind the black box estimations through a graphical and correlation analysis for a few regions with very different access to European markets. Finally, I compare estimation results obtained using the indicator of Market Potential and an artificial indicator calculated for the whole European sample, the evolution of which summarizes the European business cycle.

The conclusions are as follows. The evolution of income per capita in European regions displays high synchronicity (Giannone and Reichlin 2006; Kunroo 2023). The average changes of regional income per capita may be captured by either the average variations of regional Market Potential or the variations of an aggregated indicator with identical European data for all the regional observations of the same year. Average temporal changes in regional Market Potential thus capture only the common economic cycle, reflecting synchronicity, or global spillovers. For the case of Europe at least, the panel data results of a NEG equation should not be interpreted in terms of the NEG hypotheses. This result is another example of the 'Marshallian' (Duranton and Puga 2004) or 'observational' equivalence of the NEG (Head and Mayer 2004, p. 2,663; Bruna 2024a).

Some of these conclusions may be generalized to many studies in different fields, that ignore the statistical properties of the transformed data when interpreting results from panel data models. Fixed effects estimations are usually justified by the need to control for time-invariant regional factors. Researchers who attend only to this argument may forget that results from cross-sectional and panel models have different interpretation. Estimating an equation with data on first differences or time-demeaned data means studying the effects of temporal changes in the explanatory variable on temporal changes in the dependent variable. Although this is well known, it is not discussed in many papers comparing cross-sectional and panel data estimates.

The remainder of the paper is structured as follows. The second section introduces the wage equation and the literature estimating it with panel data. Section 3 reviews the interpretation of estimation results when using cross-sectional and panel data models. Section 4 presents the data and methodology, and Section 5 the results. The final section draws conclusions.

2 The empirical wage-type equation and panel data

A variety of NEG models focus on different mechanisms to explain agglomeration (Baldwin et al. 2003), even in the absence of any pure external economies. Using Krugman's (1993) model of metropolitan areas as a benchmark and assuming all other conditions equal, firms that have an incentive to concentrate production at a limited number of locations prefer locations with good access to markets. Yet access to markets will be good precisely where a large number of firms choose to locate. This positive feedback loop drives the formation of urban centers. It also implies that the location of such centers is not wholly determined by the underlying natural geography, that there are typically multiple locational equilibria. To capture this intuition, the formal model has three features. First, location matters because of transportation costs. Second, some immobile production factors provide a form of 'first nature' that constrains the possible spatial structure of the economy. Finally, economies of scale in the production of at least some goods provide an incentive for concentration. The existence of the metropolis thus creates a 'second nature' that drags the optimal location of firms with it.

Krugman's initial models "suggest an explanation for the nineteenth-century formation of real-world core-periphery patterns, notably the emergence of the United States' manufacturing belt and Europe's 'hot banana'" (Krugman 2011). Krugman recognizes, however, the increasing importance of technology and information spillovers: "Ever since the beginnings of New Economic Geography, and up until very recently, I and others have had a slightly guilty sense that we were talking about was the past, not the present, and much less the future (Krugman 2011). In sum, the NEG establishes "fundamental determinants" of economic activity (Redding and Venables 2004) based on long-term consequences of agglomeration forces.

The so-called 'wage equation' is a market-clearing condition of the basic NEG model in which labor is a unique production factor. I will now present a one-sector generalized form of this equation in which the dependent variable is not wages but marginal costs and thus encompasses many of the 'wage equations' previously derived in the literature (Combes et al. 2008, chap. 12; Bruna 2015). For a firm in region i ($i = 1, \dots, R$) with zero profit, the maximum value of marginal cost (m_i) the firm can afford to pay depends on its access to markets. Marginal cost is thus proportional to firm's (region's) Real Market Potential (RMP_i) (to use Head and Mayer's (2006) term) or Market Access (to use Redding and Venables' (2004) term), as follows:

$$m_i = Constant \cdot (RMP_i)^{\frac{1}{\sigma}} = Constant \cdot \left(\sum_{j}^{R} {T_{ij}}^{1-\sigma} \frac{E_j}{S_j} \right)^{\frac{1}{\sigma}} \qquad (1)$$

where, $\sigma > 1$ is the elasticity of substitution between any pair of varieties of goods in a love-of-variety utility function. RMP_i is a weighted sum of the market conditions in the other j regions, where T_{ij} is the trade cost from firm-or-region i to region j, and E_j is total expenditure in j. S_j is called the 'competition index' to stress that it measures the level of competition among varieties in j market, given consumers' characteristic tastes. The NEG's long-term prediction is that firms and regions with higher Market Potential tend to earn more profit and pay higher remuneration to production factors, resulting in higher regional income per capita.

If trade costs are proxied by physical distances (d_{ij}), the explanatory variable of Equation (1) becomes $RMP_i = \sum_j^R d_{ij}{}^{1-\sigma} \frac{E_j}{S_j}$. As in some previous literature, marginal cost (m_i) can be proxied by data on gross value added per capita ($GVApc$) and total expenditure (E_j) by data on GVA. Harris' (1954) index of accessibility to markets, in contrast, can be defined as $HMP_i = \sum_j^R d_{ij}{}^{-1} GVA_j$. Since a -1 trade elasticity to distance is an extremely robust empirical finding in the literature on gravity equations (Head and Mayer 2014), the major difference between RMP_i and HMP_i lies in S_j, which is not directly measurable in NEG theory. For samples of European regions, Breinlich (2006) and Head and Mayer (2006) obtained similar empirical results using both Harris' indicator and the more sophisticated procedure of Redding and Venables (2004) to proxy S_j. Bruna (2024a) shows that both approaches capture the core-periphery spatial patterns in the data in a similar way.

Moreover, when calculating Market Potential with areal data, the access of firms to markets also depends on the market size of their own region—that is, on so-called self-potential or Internal Market Potential. Not only does considering this potential in applied work add endogenous information, but the measurement of internal distances (d_{ii}) is controversial (Bruna 2024b). This study therefore avoids self-potential and uses External Market Potential, defined as $EMP_i = \sum_{j \neq i}^{R-1} d_{ij}{}^{-1} GVA_j$. Taking natural logarithms to equation (1) and replacing variables with my proxies, I thus obtain the following estimable cross-sectional equation:[1]

$$\log GVApc_i = C + \beta \log EMP_i + u_i \quad (2)$$

[1] Empirical NEG literature does not clearly distinguish pecuniary external economies due to market size from other possible spillovers due to knowledge or expectations. In Equation (2), regional gross value added (GVA) per capita is affected by the GVA of other regions, a condition compatible with a variety of explanations. See Duranton and Puga (2004), Bruna et al. (2016), Elhorst (2024), Bruna (2024a), and discussion below.

Hanson (2005) and Mion (2004) estimated the first panel data model of the wage equation and discussed the advisability of using the generalized method of moments (GMM) or nonlinear least squares. They also carefully discussed justification of a panel data version of Equation (2), including time-invariant individual effects, as in Equation (5) below. Breinlich (2006) justified these fixed effects to capture persistent factors such as institutional quality or climatic or other amenities of a region. Fingleton (2008) assumed that the dependent variable in a time-varying version of Equation (1) also depends on level of efficiency (A_{it})—a useful trick to model complexity by making A_{it} depend on past efficiency, the efficiency of neighboring regions (spillovers), and time-invariant regional characteristics.

For the NEG equation and areal data, Fingleton applied the fixed effects estimator and the Kapoor-Kelejian-Prucha (KKP) GMM estimator of random effects models, including serially and spatially autocorrelated disturbances (Fingleton 2008, 2009; Fingleton and Fischer 2010; Gómez-Antonio and Fingleton 2012; Fingleton and Palombi 2013). Amaral et al. (2010) and Wang and Haining (2017) have also used this methodology. Further, Fingleton was a coauthor in Baltagi et al.'s (2014) proposal of a KKP method to estimate a complex spatial econometric dynamic panel data model, proposal that is illustrated with a NEG wage-type equation. Using microdata, Fingleton and Palombi (2013) and Fingleton and Longhi (2013) estimated a fixed effects panel data model for individuals' wages. Some panel data literature has used the NEG framework to study other topics (e.g., Gómez-Antonio and Fingleton 2012, for public capital; de Sousa and Poncet 2011, for migration). For instrumental variables estimation, Boulhol and de Serres (2010) and Head and Mayer (2011) included time-invariant instruments and time dummies in the first stage regression.

Some of these studies have used samples of European regions (Breinlich 2006; Fingleton and Fischer 2010; Baltagi et al. 2014). Others used data for one European country—Mion (2004) for Italy, Gómez-Antonio and Fingleton (2012) for Spain, three of Fingleton's studies for the United Kingdom, and Rokicki and Cieślik (2023) for Poland.[2]

[2] The following papers use Harris' indicator of Market Potential: Breinlich (2006), Fingleton (2008), Amaral et al. (2010), Baltagi et al. (2014), Wang and Haining (2017), and Rokicki and Cieślik (2023).

Although many of these panel data studies find significant effects of Market Potential on income per capita, their results may be described as an anomaly. As the next section shows, panel data estimates capture short-term effects, so those significant effects are an unexpected result from a theory explaining the historical causes of agglomeration.

3 Interpreting panel data estimations of the NEG equation

To establish key ideas, it is useful to review interpretation of the coefficients in several econometric models. Readers who feel comfortable with these basics may skip to the end of the section.

The cross-sectional wage-type equation derived from the NEG produces an estimable equation for region i ($i = 1 \ldots R$) in a given period t, such as the following:

$$y_i = C + \beta x_i + u_i \tag{3}$$

Pooling data for T periods ($t = 1 \ldots T$)[3] in each region i (group), we get the following model:

$$y_{it} = C + \beta x_{it} + u_{it} \tag{4}$$

The fixed effects extension of Equation (4) includes unobserved time-invariant individual effects, u_i, as follows:

$$y_{it} = C + \beta x_{it} + u_i + u_{it} \tag{5}$$

where u_i collects omitted regional variables assumed to have a roughly constant role in explaining regional differences of y_{it}. Averaging Equation (5) over the T periods produces the following between-group model, capturing cross-sectional average relationships:

$$\bar{y}_i = C + \beta \bar{x}_i + u_i + \bar{u}_i \tag{6}$$

Subtracting Equation (6) from (5) produces the estimable fixed effects panel model, with the variables as deviations to the regional means:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i) \tag{7}$$

[3] For reasons explained in the following section, I do not include time effects in any of these models. Considering time effects does not change the results obtained in the paper (see online Supplementary Appendix).

This is the within-group model, and the transformed variables are temporal variations within each region. We can estimate the model using standard ordinary least squares (OLS) by pooling the demeaned data. The $\hat{\beta}$ estimate will be identical to that of Equation (4) after including (controlling out) $R-1$ regional dummy variables.

Boulhol and de Serres (2010)—and Acemoglu et al. (2008) in another field—summarize the standard argument in favor of defining the model as in equation (7). The major source of potential bias in a regression of Market Potential on income per capita is region-specific, historical factors that influence both Market Potential and economic development. If these omitted characteristics are, on first approximation, time-invariant (u_i), inclusion of fixed effects will remove them and this source of bias. Paying attention only to that possible unobserved individual heterogeneity has, however, led some researchers to mistaken conclusions when interpreting results, as described below.

The cross-sectional, pooled, and between models of Equations (3), (4), and (6), respectively, are about relative levels of regional variables. The estimate of these models, which I term $\hat{\beta}_c$, captures the effect of cross-sectional variations of x_i on cross-sectional variations of y_{it}. A cross-sectional wage equation thus implies that regions with higher Market Potential will be richer. Conversely, the estimate from the fixed effects Equation (7), which I term $\hat{\beta}_f$, captures the effect of time variations of x_{it} on time variations of y_{it}. In Acemoglu et al.'s (2008) terminology, a fixed effects panel model of the wage equation implies that regions with increasing Market Potential will become richer. This is not what the NEG predicts for a time span of either one or a few years.[4] The NEG's historical explanation should thus be studied using cross-sectional/pooled estimations ($\hat{\beta}_c$), while fixed effects estimations ($\hat{\beta}_f$) are more suitable for studying short-run relationships (Baltagi and Griffin 1984).

To better interpret the fixed effects estimator, it is useful to compare it to a model in first differences. The time lag of Equation (5) is the following:

$$y_{it-1} = C + \beta x_{it-1} + u_i + u_{it-1} \quad (8)$$

[4] I have repeated the calculations in this paper for a panel of seven time-observations defined for four-year data averages. The online Supplementary Appendix shows that the empirical results are very similar.

Subtracting Equation (8) from (5) produces the following first difference estimator:

$$y_{it} - y_{it-1} = \beta(x_{it} - x_{it-1}) + (u_{it} - u_{it-1}) \tag{9}$$

where the first differences for period $t = 1$ are lost. If we use the notation $\hat{\beta}_d$ for the first difference estimate, its magnitude will be similar to $\hat{\beta}_f$. $\hat{\beta}_d$ in Equation (9) makes clearer, however, that we are estimating the short-run effects of changes in x_{it} on changes in y_{it}.[5] Moreover, since y_{it} is the log of income per capita and x_{it} is the log of Market Potential, the first difference of the logarithm of a variable is the instantaneous growth rate of that variable, which is very similar to the discrete growth rate (g) of the variable. In sum, in a wage equation, $\hat{\beta}_f$ will be very similar to the estimate of a model for the growth rates of the levels of Market Potential and income per capita. Calling $X_{it} = e^{x_{it}}$ and $Y_{it} = e^{y_{it}}$, the fixed effects estimate will be similar to that obtained with the following pooled model of the variables in levels, for $t = 2 \dots T$:[6]

$$g_{Yit} = C + \beta g_{Xit} + u_{it} \tag{10}$$

As mentioned in Section 2, nothing in (long-term) location theory predicts that yearly growth rates of an indicator of regional Market Potential should have a significant impact on regional growth rates of income per capita. I, however, obtain positive significant $\hat{\beta}_f$ estimates below. Why?

4 Data and methodology

The sample includes Eurostat regional data for 233 regions from 25 countries (but not Switzerland), for the years 1995-2022. The explanatory variable is Harris' indicator of External Market Potential. As mentioned above, my estimation of a wage-type equation uses real $GVApc$ to proxy regional marginal costs and real GVA to proxy regional market size. The data come from the ARDECO database (see online Supplementary Appendix for

---

[5] When there is high serial correlation in the idiosyncratic random term, $\hat{\beta}_d$ is more efficient than $\hat{\beta}_f$ (Boulhol and de Serres 2010).

[6] Ottaviano and Pinelli (2006) propose another model derived from an interpretation of the wage equation assuming that regions fluctuate around a balanced growth path. These authors define the growth rate g_{Yit} within a time span of ten years and regress it on the logs of Market Potential and per capita income for the initial year. I do not study their model here. Bruna (2024a) shows that the spatial patterns of European EMP and a pure indicator of geographical centrality are almost identical. For 2019, the correlations of the log of EMP with the logs of GVApc and inverse mean distance to all the other regions are 0.46 and 0.88, respectively.

further details). Inter-regional distances (d_{ij}) are measured as great-circle distances between regional centroids.

I now begin to disentangle the reasons we obtain a significant fixed effects estimate of Market Potential in my European wage-type equation. I start by estimating the following cross-sectional and pooled panel models, expecting to find similar values for $\hat{\beta}$:

$$\log GVApc_i = C + \beta \log EMP_i + u_i \qquad \text{Model 1}$$

$$\log GVApc_{it} = C + \beta \log EMP_{it} + u_{it} \qquad \text{Model 2}$$

I then estimate the model in first differences, as in Equation (9), and the within-region model, as in Equation (7), expecting to find similar values for $\hat{\beta}$, as follows:

$$\begin{aligned} &\log GVApc_{it} - \log GVApc_{it-1} \\ &= C + \beta(\log EMP_{it} - log\, EMP_{it-1}) + (u_{it} - u_{it-1}) \end{aligned} \qquad \text{Model 3}$$

$$\begin{aligned} &\log GVApc_{it} - \overline{log\, GVApc_i} \\ &= C + \beta(\log EMP_{it} - \overline{log\, EMP_i}) + (u_{it} - \overline{u_i}) \end{aligned} \qquad \text{Model 4}$$

For a cross-sectional European sample, Bruna (2024a) shows that the estimation results of a wage-type equation are spurious because the data are spatially nonstationary. The spatial distribution of Market Potential and many other variables roughly matches the core-periphery spatial patten of European regional income, which cannot be used to confirm the NEG explanation. Inspired by this finding, I conjecture that the significant effect of Market Potential in the fixed effects estimation of Model 4 might be due to a common time pattern in the series of GVA per capita and EMP. To test this hypothesis, I calculate an artificial indicator of European Market Potential ($EuMP$). As with the definition of $EMP_{it} = \sum_{j\neq i}^{R-1} d_{ij}{}^{-1} GVA_{jt}$, I build the following indicator $EuMP_t = d_m{}^{-1} \sum_{j=1}^{R} GVA_{jt}$, where d_m is the median of the average distances from each region to each of the others.[7] Note that EMP_{it} changes by region and period, whereas $EuMP_t$ is the same for any region in the same period and captures the aggregate business cycle. Model 5 repeats the fixed effects

[7] With this definition, the GVA of each region is used to build the European MP, which is later used to explain each region's GVApc. The resulting endogeneity of the European variable is similar to the results obtained using regional income to measure Internal Market Potential in the NEG literature (Bruna 2024b). The consequences are minor, however. The median weight of GVA_{jt} in $\sum_{j=1}^{R} GVA_{jt}$ is 0.26%.

equation of Model 4 but replaces EMP_{it} with $EuMP_t$. Since $EuMP_t$ only varies by year, the estimation of Model 5 cannot include time effects. For reasons of comparability, I estimate all models without time effects. Model 6 includes both EMP_{it} and $EuMP_t$. I use R's 'plm' package (Croissant and Millo 2008) to estimate all these panel data models.

Moreover, EMP_{it} can be seen as a spatially lagged endogenous variable of the wage equation and thus also captures global spillovers (Mion 2004; Bruna et al. 2016). This interpretation is useful to test the robustness of results to the estimation of spatial econometrics panel data models, using R's 'splm' package (Millo and Piras 2012). I define a spatial weights matrix (W) as a row-standardized binary matrix that includes the four nearest neighbors (see online Appendix for additional details). With this W, the spatial lag of a variable y_i for region i is the mean of y_j for its four nearest j regions.

Starting from a generic representation of the fixed effects models, $y_t = \beta x_t + u + u_t$, the Spatial Error Model (SEM) assumes that possible ignored explanatory factors can correlate spatially, as captured by the following spatial model of the unexplained cross-sectional variation of the dependent variable: $u_t = \lambda W u_t + \epsilon_t$.[8] The Spatial Autoregressive Model (SAR), however, assumes spillovers from income per capita of neighboring regions, adding the spatial lag of the dependent variable (Wy_t) as an explanatory variable: $y_t = \beta x_t + \rho W y_t + u + u_t$. Since y_t and x_t are the logs of GVApc and External Market Potential, respectively, the SAR model introduces some duplicate information about external markets: $Wy_t = \sum_{j \neq i}^{4} {w_{ij}}^{-1} \log GVApc_{jt}$ partially overlaps with $x_t = \log EMP_{it} = \log \sum_{j \neq i}^{R-1} {d_{ij}}^{-1} GVA_{jt}$, where ${w_{ij}}^{-1}$ is the corresponding non-diagonal element of my W matrix (displayed here in a format comparable to the weighting scheme in EMP).[9]

Additionally, estimation residuals of my fixed effects estimations tend to display not only spatial autocorrelation but also serial dependence. Further, I analyze the robustness of the panel results to the inclusion of the time lag of the dependent variable, y_{t-1}. The SEM and SAR models to be estimated are thus as follows:

---

[8] Unlike the notation in the 'splm' package, I use λ for the spatial parameter of the SEM model.

[9] A key difference is that ${d_{ij}}^{-1}$ is measured as inverse absolute distances, while standardized W matrices ignore sample geography to focus on local issues. See Bruna et al. (2016).

$$y_t = \beta x_t + \gamma y_{t-1} + \lambda \mathrm{W} u_t + u + \epsilon_t \qquad \text{Model 7}$$

$$y_t = \beta x_t + \gamma y_{t-1} + \rho \mathrm{W} y_t + u + u_t \qquad \text{Model 8}$$

The calculation of time lags for Models 3, 7, and 8 use data for 1995. To facilitate comparability, however, I estimate all panel data models with the same number of observations for 27 years (1996-2022).

I do not estimate Models 1-8 with methods that correct for biases due to potential endogeneity of the explanatory variables, as is common in causal analysis in the NEG empirical literature. My story is not about causality but about the statistical properties of the variables transformed to short-term variations. I argue that these properties do not capture what they are supposed to within NEG framework.

To illustrate the intuitions about the statistical features behind the black box estimations, I compare data aggregated for the entire European sample with data of four regions, selected based on their different access to the main European markets (see online Appendix for details): Mittelfranken, Algarve, Kriti, and Nord-Norge. Figure 1 shows their location and the time series of their log GVApc, as well as the latter variable for the whole sample (solid line).[10] For the period 1996-2022, the Pearson correlations[11] of the times series of Algarve and Kriti with the European time series are 0.24 and -0.08, respectively. This correlation is lower than 0.4 for 15% of the regions but higher than 0.8 for 77% of the regions, indicating that regional European income per capita evolves quite synchronically (Giannone and Reichlin 2006; Kunroo 2023).

[10] Panel unit root tests reveal that the time series of lGVApc are generally stationary (see online Supplementary Appendix). In any case, my argument involves not potential spurious results due to nonstationarity but the misinterpretation of the empirical results of the fixed effects wage equation.

[11] To detect covariation, or synchronicity, Pearson's (linear) correlation is more stringent than Kendall or Spearman correlations, which are rank-based coefficients.

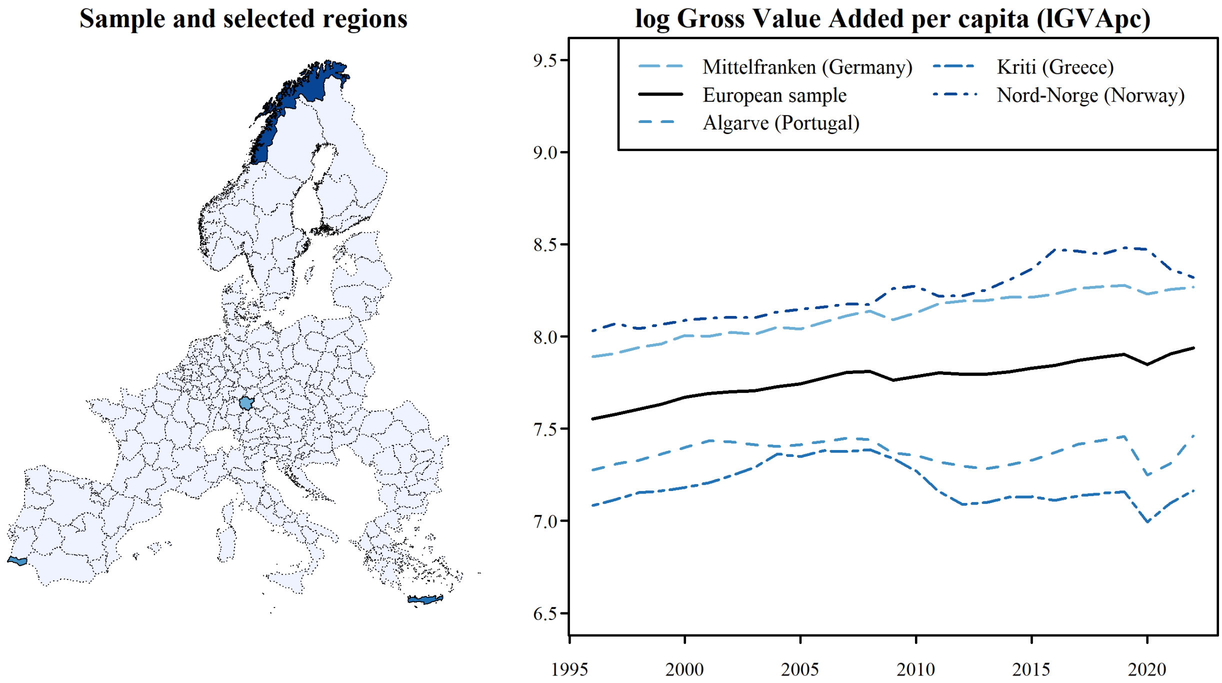

Figure 1. Location of selected regions and evolution of their income per capita.

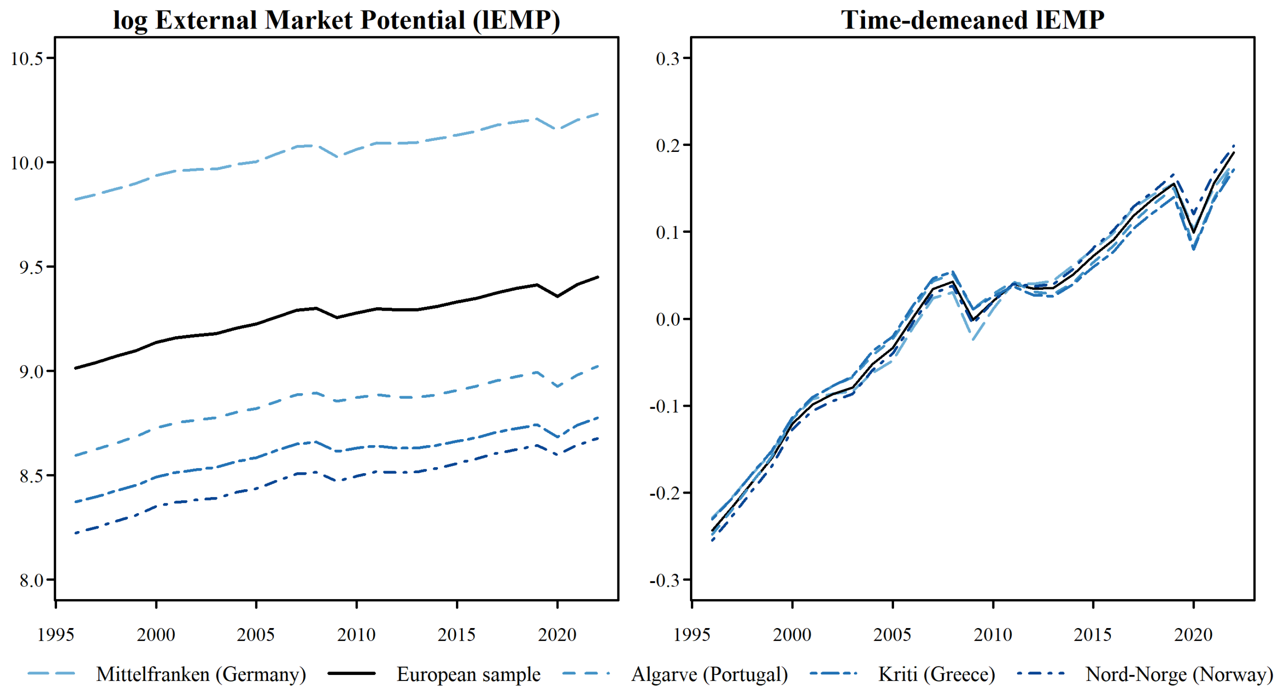

Figure 2. Market Potential for selected regions and its time-demeaned version.

5 Results

Figure 2 represents the time series of lEMP and its time-demeaned version for the four regions selected and the European sample. The latter is measured by the artificial indicator $EuMP_t$. The plot on the left shows that the region of Norway has low Market Potential due to both its peripherality and its larger size (Bruna 2024a).[12] Geographical centrality and small rich neighborhoods give the German region high Market Potential. For two reasons, all these series evolve similarly, following the evolution of the sum of GVA for the whole sample (solid line). Firstly, the general synchronicity of the time series of regional GVA (see Figure 1 for GVApc) tends to produce similar movements in data aggregated for several regions. Secondly, the smoothing effects of the sum for $R-1$ regions in Harris's accessibility index tend to dissipate local differences of the temporal changes in GVA. Given that the regional indicators of Market Potential tend to move together, the time-demeaned series in the right-hand plot are very similar. In other words, in a fixed effects estimation, Market Potential does not really capture relevant changes in the accessible market size for firms in each region. Rather, it captures the European business cycle.

To test this explanation, Table 1 compares estimation of the eight models described in Section 4. As expected, $\hat{\beta}_c$ estimates of log External Market Potential are similar in Columns (1) and (2). They are also similar in Columns (3) and (4), which refer to $\hat{\beta}_d$ and $\hat{\beta}_f$, respectively, and take values of around 1. This means that a 1% increase in the growth rate of EMP_{it} generates a 1% increase in the growth rate of $GVApc_{it}$.[13] Under a NEG interpretation of the results, these estimates could be considered biased due to the endogeneity of Market Potential. I argue, however, that this high estimate reflects only the translation of the European economic cycle in GVA to the common movements in the regional times series of GVA per capita. As I discuss below, this high unitary elasticity collapses to 0.06 if the lagged dependent variable is added to the equation in Column (4) (see online Appendix), suggesting that the results of Market Potential are driven by the business cycle.

[12] Norwegian regions are an exception to the rough core-periphery pattern of GVApc around the so-called European 'blue (or hot) banana'. They are relatively rich (Figure 1) but peripheral and so have low EMP (Figure 2). See: https://en.wikipedia.org/wiki/Blue_Banana.

[13] For the four regions chosen, the mean correlation between the time series of EMP in first difference and in discrete growth rates is 0.988.

Column (5) of Table 1 shows the paper's main contribution. When the regional values of External MP (EMP_{it}) are replaced by the artificial variable of European MP ($EuMP_t$), adjusted R^2 and the estimate are very similar to those in Column (4). The regional differences of time-demeaned External MP thus make little difference compared to the aggregate evolution of European GVA, as shown in Figure 2. This result has nothing to do with NEG theory, as the explanation of the fixed effects results must be confined to the purely statistical domain. The key issue explaining the synchronic evolution of regional GVApc is the aggregate evolution of European GVA.

Table 1 Cross-sectional and panel data models of the effects of two variables of Market Potential (MP) on gross value added per capita ($lGVApc$)

	Cross-section	Pooled panel	First differences	Within group panel (regional fixed effects)				
				Nonspatial			Spatial	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$log\ External\ MP$	0.608^{***}	0.671^{***}	1.034^{***}	1.014^{***}		3.858^{***}		
	(0.174)	(0.083)	(0.030)	(0.048)		(0.479)		
$log\ European\ MP$					0.964^{***}	-2.893^{***}	0.051^{***}	-0.023^{***}
					(0.051)	(0.467)	(0.010)	(0.007)
$log\ GVApc\ (t-1)$							0.926^{***}	1.003^{***}
							(0.005)	(0.012)
Wu [λ parameter]							0.670^{***}	
							(0.010)	
$Wlog\ GVAp$ [ρ]								0.189^{***}
								(0.007)
Adjusted R^2	0.209	0. 217	0.345	0.567	0.499	0.650		
Log likelihood							-5,236	12,236

Note: Cross-sectional data (233 observations) refer to 2019 and panel data (6,291 obs.) to 1996-2022. The intercept in Columns (1) and (2) is not reported. Standard errors (in brackets) are clustered by country in Column (1) and by region and year, and robust to persistent common shocks, in Columns (2) to (6). Coefficients of variables in Column (8) are total impact estimates after 300 simulations to compute the impact distribution. *p<0.1; **p<0.05; ***p<0.01.

Column (6) presents the results when both variables of Market Potential are included in the fixed effects estimation. Their estimates are statistically significant but with opposing signs, due to multicollinearity: The average correlation of all regional series of EMP_{it} with the artificial variable $EuMP_t$ is 0.997.

Tests reveal the presence of temporal and spatial dependence in the residuals of the previous models.[14] Columns (7) and (8) shows estimations of the model in Column (5)[15] after considering temporal and spatial dependence. The inclusion of the time lag of the dependent variable captures inertia in the regional business cycle. To correct for spatial autocorrelation, Columns (7) and (8) show result of a SEM and SAR models, respectively.[16] Both spatial parameters are statistically significant.

SEM estimation shows a way of capturing departures from the European cycle due to spatially autocorrelated omitted variables. Controlling for that, Column (7) shows that the regional effects of the contemporary European cycle, as captured by European Market Potential, are still significant, with a parameter of 0.05.[17]

The SAR model in Column (8) includes the spatial lag of the dependent variable, thus capturing spillovers from the evolution of the dependent variable of nearest neighbors. Once the local business cycle in each area of the map is incorporated, the effects of the aggregate cycle captured by European MP become negative. The reason, again, is multicollinearity. The correlation between the times series of the spatial lag of the dependent variable and the log of European MP is lower than 0.4 for 12% of the regions but higher than 0.8 for 85% of the regions. This result means that the spatial lag of the dependent variable also follows the European cycle but is more precise in capturing local departures from it to explain the evolution of regional income per capita.

6 Conclusions

Panel data literature emphasizes the need to control for individual heterogeneity to avoid biases caused by time-invariant regional factors. It should not be forgotten, however, that

[14] I conducted Breusch-Godfrey tests for serial dependence and Pesaran's tests for cross-sectional dependence in the residuals of the models in Columns (4) and (5). See online Supplementary Appendix.

[15] Conclusions derived from Columns (7) and (8) are very similar if these models are based on the equation for External Market Potential in Column (4). See Appendix.

[16] Panel Lagrange Multiplier tests for spatial dependence in the equation in column (5) reveal a weak preference for the SAR over the SEM correction for spatial autocorrelation. See Appendix.

[17] The estimate of European MP in Column (5) becomes 0.050 when the time lag of the dependent variable is added to the equation (see Appendix). The similarity of this number to the SEM 0.051 estimate in Column (7) indicates correct specification of the model including $EuMP_t$ and $log\ GVApc_{it-1}$.

controlling for unobserved regional fixed effects forces use of transformed data on temporal changes. The resulting within-region estimates are not comparable to the between-region estimates obtained using cross-sectional data. For data in log form defined at intervals of one or a few years, panel data estimates approximate the effects of the short-term growth rates of variables and are thus not a good way of testing long-term theoretical predictions.

This study shows that the evolution of regional income per capita is highly synchronic in Europe, following the European cycle. The smoothing effects of summation when constructing the variable Market Potential intensify this variable's synchronicity: the temporal changes in Market Potential are very similar for all regions. Changes in Market Potential thus capture changes in the European business cycle, which explain the average changes in regional income per capita well. Short-term growth rates in Market Potential capture short-term growth rates of aggregate European economic activity, which are also similar to the average short-term growth rates of income in neighboring regions. The panel data estimates capture correlations or spillovers from the aggregate cycle but not the mechanisms studied by the NEG. When using these data, therefore, a variable of Market Potential that is statistically significant in a fixed effects estimation does not confirm the NEG.

This research should be extended to study other geographical samples. For long-run theories, panel data methods are more appropriate when using long historical data sets for time intervals of decades. The 'pooled mean group estimator' (Pesaran et al. 1999) of dynamic panels with a large number of time and spatial units takes into account possible common long-run relationships across spatial units. An alternative approach comes from the literature on spatial and temporal unit roots and cointegration in panel data (Banerjee 1999; Baltagi and Shu 2024).

Ultimately, researchers should choose statistical methods based on the time frame of the underlying theory. They also should be careful with the temporal and spatial properties of the (transformed) data and use graphical and statistical tools before black box estimations.

References

Acemoglu D, Johnson S, Robinson JA, Yared P (2008) Income and Democracy. American Economic Review 98:808–42

Amaral PV, Lemos M, Simões R, Chein F (2010) Regional Imbalances and Market Potential in Brazil. Spatial Economic Analysis 5:463–482. https://doi.org/10.1080/17421772.2010.516441

Baldwin RE, Forslid R, Martin P, et al (2003) Economic Geography and Public Policy. Princeton University Press, Princeton, N.J.

Baltagi BH, Fingleton B, Pirotte A (2014) Estimating and Forecasting with a Dynamic Spatial Panel Data Model. Oxford Bulletin of Economics and Statistics 76:112–138. https://doi.org/10.1111/obes.12011

Baltagi BH, Griffin JM (1984) Short and Long Run Effects in Pooled Models. International Economic Review 25:631–45. https://doi.org/10.2307/2526223

Baltagi BH, Shu J (2024) A Survey of Spatial Unit Roots. Mathematics 12:1052. https://doi.org/10.3390/math12071052

Banerjee A (1999) Panel Data Unit Roots and Cointegration: An Overview. Oxford Bulletin of Economics and Statistics 61:607–629. https://doi.org/10.1111/1468-0084.0610s1607

Boulhol H, de Serres A (2010) Have developed countries escaped the curse of distance? J Econ Geogr 10:113–139. https://doi.org/10.1093/jeg/lbp015

Breinlich H (2006) The spatial income structure in the European Union—what role for Economic Geography? J Econ Geogr 6:593–617. https://doi.org/10.1093/jeg/lbl018

Bruna F (2024a) Market Potential, spatial theories, and spatial trends. Spat Econ Anal. https://doi.org/10.1080/17421772.2024.2325517

Bruna F (2015) A generalized NEG wage-type equation. In: Díaz-Roldán C, Perote J (eds) Advances on International Economics. Cambridge Scholars Publishing, Newcastle, pp 63–82

Bruna F (2024b) Market potential: the measurement of domestic market size. Lett Spat Resour Sci 17:1–13. https://doi.org/10.1007/s12076-024-00378-8

Bruna F, Lopez-Rodriguez J, Faíña A (2016) Market Potential, Spatial Dependences and Spillovers in European Regions. Reg Stud 50:1551–1563. https://doi.org/10.1080/00343404.2015.1048796

Combes P-P, Mayer T, Thisse J-F (2008) Economic geography: the integration of regions and nations. Princeton University Press, Princeton, N.J.

Croissant Y, Millo G (2008) Panel Data Econometrics in R: The plm Package. Journal of Statistical Software 27:1–43 de Sousa J, Poncet S (2011) How are wages set in Beijing? Regional Science and Urban Economics 41:9–19. https://doi.org/10.1016/j.regsciurbeco.2010.07.004

Duranton G, Puga D (2004) Micro-foundations of urban agglomeration economies. In: Henderson JV, Thisse J-F (eds) Handbook of Regional and Urban Economics 4. North Holland, Amsterdam, pp 2063–2117

Elhorst JP (2024) Raising the bar in spatial economic analysis: two laws of spatial economic modelling. Spatial Economic Analysis 19:115–132. https://doi.org/10.1080/17421772.2024.2334845

Fingleton B (2008) Competing models of global dynamics: evidence from panel models with spatially correlated error components. Economic Modelling 25:542–558

Fingleton B (2009) Testing the NEG model: Further evidence from panel data. Région et Développement 30:141–158

Fingleton B, Fischer MM (2010) Neoclassical theory versus new economic geography: competing explanations of cross-regional variation in economic development. The Annals of Regional Science 44:467–491. https://doi.org/10.1007/s00168-008-0278-z

Fingleton B, Longhi S (2013) The Effects of Agglomeration on Wages: Evidence from the Micro-Level. Journal of Regional Science 53:443–463. https://doi.org/10.1111/jors.12020

Fingleton B, Palombi S (2013) The wage curve reconsidered: is it truly an "empirical law of economics"? Région et Développement 38:49–92

Giannone D, Reichlin L (2006) Trends and cycles in the euro area: how much heterogeneity and should we worry about it? ECB Working Paper

Gómez-Antonio M, Fingleton B (2012) Analyzing the Impact of Public Capital Stock Using the Neg Wage Equation: A Spatial Panel Data Approach. Journal of Regional Science 52:486–502. https://doi.org/10.1111/j.1467-9787.2011.00725.x

Hanson GH (2005) Market potential, increasing returns and geographic concentration. J Int Econ 67:1–24. https://doi.org/10.1016/j.jinteco.2004.09.008

Harris CD (1954) The Market as a Factor in the Localization of Industry in the United States. Ann Assoc Am Geogr 44:315–348. https://doi.org/10.1080/00045605409352140

Head K, Mayer T (2011) Gravity, market potential and economic development. J Econ Geogr 11:281–294

Head K, Mayer T (2004) The empirics of agglomeration and trade. In: Henderson JV, Thisse J-F (eds) Handbook of Regional and Urban Economics 4. North Holland, Amsterdam, pp 2609–2669

Head K, Mayer T (2006) Regional wage and employment responses to market potential in the EU. Reg Sci Urban Econ 36:573–594. https://doi.org/10.1016/j.regsciurbeco.2006.06.002

Head K, Mayer T (2014) Gravity equations: workhorse, toolkit, and cookbook. In: Gopinath G, Helpman E, Rogoff K (eds) Handbook of International Economics 4. Elsevier, Amsterdam, pp 131–195

Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99:483–499. https://doi.org/10.1086/261763

Krugman P (1993) First Nature, Second Nature, and Metropolitan Location. Journal of Regional Science 33:129–144. https://doi.org/10.1111/j.1467-9787.1993.tb00217.x

Krugman P (2011) The New Economic Geography, Now Middle-aged. Reg Stud 45:1–7. https://doi.org/10.1080/00343404.2011.537127

Kunroo MH (2023) Business cycle synchronization and its determinants in the OECD countries: panel data evidence. Applied Economics 0:1–16. https://doi.org/10.1080/00036846.2023.2288049

Millo G, Piras G (2012) splm: Spatial Panel Data Models in R. Journal of Statistical Software 47:1–38

Mion G (2004) Spatial externalities and empirical analysis: the case of Italy. Journal of Urban Economics 56:97–118. https://doi.org/10.1016/j.jue.2004.03.004

Ottaviano GIP, Pinelli D (2006) Market potential and productivity: Evidence from Finnish regions. Regional Science and Urban Economics 36:636–657. https://doi.org/10.1016/j.regsciurbeco.2006.06.005

Pesaran MH, Shin Y, Smith RP (1999) Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. Journal of the American Statistical Association 94:621–634. https://doi.org/10.1080/01621459.1999.10474156

Redding SJ (2011) Economic Geography: a Review of the Theoretical and Empirical Literature. In: Bernhofen D, Falvey R, Greenaway D, Kreickemeier U (eds) Palgrave Handbook of International Trade. Palgrave Macmillan, London, pp 497–531

Redding SJ, Venables AJ (2004) Economic geography and international inequality. J Int Econ 62:53–82. https://doi.org/10.1016/j.jinteco.2003.07.001

Rokicki B, Cieślik A (2023) Rethinking regional wage determinants: regional market potential versus trade partners' potential. Spat Econ Anal 18:126–142. https://doi.org/10.1080/17421772.2022.2070657

Wang C-Y, Haining R (2017) Testing the new economic geography’s wage equation: a case study of Japan using a spatial panel model. Ann Reg Sci 58:417–440. https://doi.org/10.1007/s00168-016-0804-3