JRC TECHNICAL REPORT

Does decentralization of governance promote urban diversity? Evidence from Spain

JRC Working Papers on Territorial Modelling and Analysis
No 07/2019

Authors:
Díaz-Lanchas, J.,
Florax, R.,
Mulder, P.

2019

Joint Research Centre

Contact information
Name: Simone Salotti
Address: Edificio Expo, C/Inca Garcilaso 3, 41092 Sevilla (Spain)
Email: simone.salotti@ec.europa.eu
Tel.: +34 954488250

EU Science Hub
https://ec.europa.eu/jrc

JRC117397

Seville: European Commission, 2019

How to cite this report: Díaz-Lanchas, J., Florax, R., and Mulder, P. (2019). Does decentralization of governance promote urban diversity? Evidence from Spain. JRC Working Papers on Territorial Modelling and Analysis No. 07/2019, European Commission, Seville, JRC117397.

The **JRC Working Papers on Territorial Modelling and Analysis** are published under the supervision of Simone Salotti and Andrea Conte of JRC Seville, European Commission. This series mainly addresses the economic analysis related to the regional and territorial policies carried out in the European Union. The Working Papers of the series are mainly targeted to policy analysts and to the academic community and are to be considered as early-stage scientific papers containing relevant policy implications. They are meant to communicate to a broad audience preliminary research findings and to generate a debate and attract feedback for further improvements.

Acknowledgments

The authors dedicate this article to the memory of Dr. Raymond J.G.M. Florax (Department of Agricultural Economics, Purdue University, and Department of Spatial Economics, Vrije Universiteit Amsterdam), who passed away unexpectedly on March 1, 2017. Raymond was instrumental in the conceptualization of this project and contributed greatly to early versions of this manuscript.

This project started while the first author visited the Department of Spatial Economics at the Vrije Universiteit Amsterdam in summer 2015 for his PhD research project which was awarded in 2016 with "The national award for the best thesis in economics" by the Royal Spanish National Academy of PhDs. The authors gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness (MINECO ECO2013-46980-P, co-financed with FEDER) and the FPU-2010 Scholarship for Pre-Doctoral Education Training by the Spanish Ministry of Education.

Abstract

The worldwide trend to decentralize the responsibilities and budgets of governments impacts local firm dynamics in various ways. We use the example of Spain to test empirically whether the decentralization of governance has been conducive to increased diversification and a more even city-size distribution in the Spanish urban system during a period of continuous reductions in transport costs. To this end we develop a bivariate probit regression framework to assess the probability that cities diversify or specialize over time, using a sample of 69 urban areas in Spain during the period 1995-2007. We exploit unique firm-level and time-varying transport-cost data to control for the role of a city's market potential, city size, transport costs, labor-force skill composition, product standardization and historical patterns of specialization. We find a high probability that a city will diversify if it is the capital of a regional government or located in a relatively decentralized region, while the opposite is true for the probability that a city will specialize. Also, we find that a city's status as a regional capital reinforces the positive effect of a high (low) internal market potential on the probability of diversification (specialization). A high (low) external market potential only increases the probability that a city will become specialized (diversified) if it is a regional capital. We argue that the link between decentralization and specialization patterns in the urban system deserves more attention in the empirical literature on decentralization's impact on economic growth, income inequality and regional disparities.

Does decentralization of governance promote urban diversity?
Evidence from Spain.

Jorge Díaz-Lanchas[a 1], Raymond J.G.M. Florax[b] and Peter Mulder[c]

[a] Joint Research Centre - European Commission.
[b] Dept. of Agricultural Economics, Purdue University & Department of Spatial Economics, Vrije Universiteit Amsterdam.
[c] Department of Spatial Economics, Vrije Universiteit Amsterdam.

JEL: R12, R58, C35
Keywords: Decentralization, urban diversity, specialization, generalized transport costs, Spain.

1. **Introduction**

When do cities specialize in production and when do they diversify? There is no shortage of theories to explain urban diversity in a system of cities. Most of these theories are quantitative models in the spirit of the seminal works by Christaller (1933), Muth (1969), Henderson (1974) and Fujita et al. (1999). The different theoretical approaches that have emerged in the economic-geography and urban-economic literature attribute growth and transformation patterns across cities to changes in trading costs and interactions with other cities; technological change, product life-cycles and learning processes – starting from different assumptions about (the micro foundations of) agglomeration and dispersion forces.[2] It is, however, remarkable that in this broad literature the topic of government decentralization has received little attention.[3] Over the last decades, decentralization – i.e. a 'deconcentration' of institutional capacities from a central to a more local level – has been at the forefront of institutional and policy transformations all around the world (Bardhan 2002). What is the impact of this wave of decentralization on diversification versus specialization patterns across cities?

An important economic rationale for decentralization is that it makes governments more responsive and efficient in the provision of public goods and services, thanks to their supposed information advantages and flexibility in adapting to citizens' diverse preferences (Tiebout 1956, Oates 1972, Breton 1996, Martinez-Vasquez and McNab 2003). However, the decentralization of governance also impacts local firm dynamics in more than the provision of public goods and services. "There is indeed plenty of anecdotal evidence that different policies implemented by governments at local and regional levels are influencing

1 Corresponding author. Email: Jorge.DIAZ-LANCHAS@ec.europa.eu.
Disclaimer: "The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission."
2 We refer to Duranton and Puga (2000, 2013) and Redding and Rossi-Hansberg (2016) for in-depth reviews.
3 An exception is Henderson and Abdel-Rahman (1991); see Section 2.

local and regional performance" (Ezcurra and Rodriguez-Pose 2013:398). Subnational governments with enough fiscal and political authority attract qualified people, thus contributing to agglomeration economies in cities that host regional government seats (Bardhan 2002). Also, decentralization gives subnational governments the opportunity to actively pursue economic development policies which fit the strengths and weaknesses of their regions better than central government policies do (Lessman 2012). Since fiscal decentralization often involves fiscal competition, subnational governments have incentives to foster local business development in order to increase their tax base, which materializes in, for example, tax privileges, more flexible labor markets or other forms of assistance (Martinez-Vasque and McNab 2003). This could help less-developed regions compete with richer ones (Qian and Weingast 1997) and increase the efficiency of the urban system (Henderson and Abdel-Rahman 1991). But fiscal competition may also come at the expense of poorer regions, which (initially) can offer only a relatively low quality of public goods and services (Prud'homme 1995), while subnational governments can also be influenced by the rent-seeking behavior and vested interests of old and inefficient firms (Bardhan 2002).

In this paper we use the example of Spain to test empirically whether the decentralization of governance is a potentially important determinant of the sectoral composition of cities in an urban system. To this end we examine structural economic dynamics across 69 urban areas in Spain during the period 1995-2007. We develop a bivariate probit regression framework to assess the probability that cities develop a certain typology over time. In our analysis, we consider the possibility that cities, in addition to the polar cases of complete diversity and specialization, also may combine both typologies through co-agglomeration economies (cf. Helsley and Strange, 2014) or feature none of them. We exploit unique firm-level and time-varying transport-cost data to control our analysis for various key mechanisms and assumptions in the theoretical literature on city formation, including the role of a city's market potential, city size, transport costs, labor-force skill composition, product standardization and historical patterns of specialization. To measure decentralization, we use the regional authority index developed by Hooghe et al. (2008); this index allows us to quantify the degree of government autonomy over time at a sub-national level.

Spain is a particularly interesting country to analyze for several reasons. First, Spain became one of the most decentralized countries in the world in just over three decades, departing from a highly centralized institutional framework (Lago-Peñas et al. 2017; Rodriguez-Pose and Ezcurra 2011). With the end of the dictatorship and the new constitution of 1978, Spain transitioned from a unitary to a more federal arrangement with the gradual creation of 17 autonomous communities (ACs) (İrepoğlu Carreras 2016).[4] This territorial administrative decentralization established a set of new regional capital cities (15 out of a total of 69 cities), hosting regional government headquarters that developed a complete list of competencies in regional and urban policies related to economic regulations, managerial tasks and fiscal issues, allowing these cities to attract firms and workers to meet the new requirements and necessities of the regional governments. Second, following a peaceful transition to democracy (1975–79), and accession to the European Economic Community (1986), Spain underwent deep and far-reaching social, urban and economic transformations under the influence of market reforms, European economic integration and falling transport costs (cf. Moreno 2002). Infrastructure growth development between 1980 and 2007 has been especially spectacular. After decades of limited infrastructure improvements, the total highway kilometers in Spain increased from only 335 km in 1980 to 9,557 km in 2007; this led to an average reduction of inter-city transport costs of more than 16%.[5] Third, as we show in this paper, during the period 1995-2007 it was more the rule than the exception that cities fundamentally changed their economic structure; hence the Spanish urban landscape changed substantially, with an increasing number of cities diversifying their production structure.

In this paper, we hypothesize that in an era of falling transport costs, fiscal and political decentralization in Spain helped smaller cities to expand and diversify their economic structure. We find an increasing probability that a city will diversify if it is the capital of a regional government or located in a relatively decentralized region, while the opposite is true for the probability that a city will specialize. Also, we find that a city's being the capital of a regional government reinforces the positive effect of a high (low) internal market potential on its probability of diversifying (specializing). A high (low) external market potential only increases the probability that a city will specialize (diversify) if that city is the capital of a regional government. We argue that the link between decentralization and specialization patterns in the urban system deserves more attention in the empirical literature on decentralization's impact on economic growth, income inequality and regional disparities.

The article proceeds as follows. In Section 2 we briefly review the literature's explanations for the potential impact of decentralization and varied quality of local governance on urban diversity. In Section 3 we describe our dataset and the calculation of key variables and present a brief descriptive analysis of (changes in) the urban system of Spain. In Section 4 we develop our bivariate probit econometric approach and present our

[4] Carrion-i-Silvestre et al. (2008) show that in 1980 89.5% of public expenditures were concentrated in the central government, whereas local government accounted for only 10.5% of total expenditures and no funds were attached to regional governments. By 2001 the share of expenditures in the central government was reduced to 60.5%, but the fiscal capacity of subnational levels increased up to 26.4% for regions and 13.1% for local governments. Indeed, Lago-Peñas et al. (2017) argue that Spain's fiscal decentralization proceeded even further between 2001 and 2009 as a result of tax-sharing mechanisms introduced in 2001.

[5] Own calculation from Zofío et al. (2014).

main regression results. Section 5 concludes.

2. Urban diversity and governance decentralization

The decentralization of governance encompasses fiscal, political and administrative decentralization. They refer, respectively, to the ability of subnational governments to raise revenues, obtain decision-making authority, and deliver public goods and services. The economic (geography) literature on decentralization has thus far focused predominantly on cross-country analyses of the impact of fiscal and political decentralization on public-service delivery, economic growth, poverty reduction, income inequality and regional disparities.[6] As mentioned before, in the theoretical literature on the determinants of diversification versus the specialization patterns in a system of cities, decentralization has not yet received much attention. A major issue in this literature, however, is the efficiency of the urban system (Duranton and Puga 2000). Existing theories are divided on this point. Models in the spirit of Henderson (1974) generally imply that efficiency in the size and number of specialized versus diverse cities can be achieved if some mechanism enabling the creation and development of new cities is present.[7] The two usual mechanisms are autonomous local governments and the market with land developers. An interesting example of such a model is Henderson and Abdel-Rahman (1991), who develop a theoretical model to analyze the role of fiscal decentralization in solving inefficiency problems of unregulated monopolistic competition. It shows that when product diversity corresponds to urban diversity – i.e. each city specializes in the export production of one differentiated product – fiscal decentralization in the form of granting local government the autonomy to subsidize local production may result in equilibrium employment levels within cities but will not solve for the sub-optimal degree of diversity across cities.

An important assumption underlying the models in this tradition is that interactions between cities are costless. More recently developed quantitative models of economic geography, however, can accommodate many asymmetric locations that may differ by geography, productivity or amenities, and that are systematically linked through distance-dependent trade, commuting and migration flows.[8] For example, Tabuchi and Thisse (2011) develop a new economic geography model of central places, to analyze the size and location of cities (the urban aspect) as well as the spatial distribution of each industry across cities (the industrial aspect). They show that increasing economic integration under the influence of falling trade costs favors the emergence of large and diversified cities, which can then coexist with small and specialized cities. In other words, weak spatial

---

[6] See, for example, Ezcurra and Rodríguez-Pose (2013), Lessmann (2012), Martinez-Vazquez and McNab (2003), Rodríguez-Pose and Ezcurra (2010, 2011), Rodríguez-Pose and Gill (2004). For a recent overview see Martinez-Vazquez et al. (2017).

[7] See, for example, Abdel-Rahman (1990, 1996), Abdel-Rahman and Fujita (1993), Henderson (1974, 2003).

[8] For an overview, see Duranton and Turner (2012), Desmet and Rossi-Hansberg (2014), Redding and Rossi-Hansberg (2017).

frictions tend to concentrate economic activity in space.[9]

In the case of Spain, however, falling trade costs and increasing economic integration since the 1980s went together with the development of a more even city-size distribution and the emergence of a range of relatively small diversified cities, as we will show in the next section (see also González-Val et al. 2014, 2015). The theoretical model developed by Anas and Xiong (2005) may help to explain this empirical observation. Their model treats the internal urban structure and questions of efficiency as being in the Henderson tradition but assumes that inter-city trade is costly. Starting from one diversified city that manufactures a product with a variety of services as inputs, they show that a specialized city will self-organize if land developers do not act just in time to set up diversified cities. In this setting, low inter-city trading costs will increase the size of the specialized city. They conclude that in an era of falling inter-city transport and communication costs, the idea that land developers will set up cities at efficient times may be largely anachronistic. In this paper, however, we argue that in the era of falling transport costs since the 1980s, fiscal and political decentralization in Spain has allowed Spanish cities to diversify their economic structure by virtue of hosting regional government headquarters. The creation of 17 autonomous local governments, rather than the actions of developers, may well have been the mechanism that established 'new' diversified cities through lumpy adjustments at the optimal time, confirming the theoretical predictions of Henderson and Becker (2000).

According to a different strand of the literature, factors such as urban infrastructure, institutional capacity and industrial composition may be more conducive to (urban) economic growth than city size and agglomeration economies (Meijers et al. 2016, Camagni et al. 2015, Frick and Rodriguez-Pose 2018). The arguments therein, amongst others, observe that in Europe second-tier cities often outperform first-tier cities in economic growth rates; this leads to a re-appraisal of connectivity in urban networks as a potential substitute for agglomeration benefits. In this paper, we hypothesize that it may be the fragmentation of central authority and the introduction of more intergovernmental competition that generate growth, and thereby cause second-tier cities to often outperform first-tier cities in economic growth rates.

3. Data and Descriptive Analysis

3.1 Data

Urban Areas

We consider 69 urban areas: that is, functional urban areas as defined by the OECD (Brezzi

[9] This argument is empirically supported by Lin (2017), who shows for China that industries with a higher reliance on nonroutine cognitive skills benefit relatively more from High Speed Railway-induced market access.

et al. 2012).[10] Together, these areas comprise, on average for the period 1995-2007, 46% of Spain's total population and 93% of its urban population. Spain has 2 large metropolitan areas (with a population of at least 1.5 million): Barcelona and Madrid; 6 metropolitan areas (with a population between 500,000 and 1.5 million); 22 medium-size urban areas (with a population between 200,000 and 500,000); and 39 small urban areas (with a population below 200,000). The large metropolitan areas, and most of the medium-size cities, are located around province (NUTS-3) capitals. Data on urban-area population are obtained from the Spanish Statistical Institute (INE) Census.

Diversification and specialization

We define the diversity and specialization of cities in terms of relative employment shares, following Duranton and Puga (2000). To measure the degree of specialization in city *i* at time *t* we use the Relative Specialization Index ($\mathrm{RZI_{it}}$), defined as:

$$RZI_{it} = max_{it}\left(\frac{s_{ikt}}{s_{kt}}\right) \tag{3.1}$$

with s_{ikt} being the share of employment *x* of sector *k* in city *i* in time *t*: $s_{ikt} = \frac{x_{it}^{k}}{\sum_k x_{it}}$; and s_{kt} the share of each sector *k* at the national level: $s_{kt} = \frac{x_t^k}{\sum_k x_t}$. Accordingly, the degree of diversification of city *i* at time *t* is measured by using the Relative Diversification Index ($\mathrm{RDI_{it}}$), defined as:

$$RDI_{it} = \frac{1}{\sum_k |s_{ikt} - s_{kt}|} \tag{3.2}$$

As RZI_{it} and RDI_{it} are continuous variables, and still following Duranton and Puga (2001), we consider the median per year for each index to categorize cities by their specialization and diversity patterns respectively. As a result, we create two discrete variables, S_{it} (specialization) and D_{it} (diversification), where S_{it} takes the value 1 for a city *i* with RZI values above the median in year *t*, and 0 otherwise; similarly, D_{it} defines diverse cities on the basis of RDI values above the median in year *t*. Our empirical strategy exploits the combination of these two discrete variables. We thus obtain four typologies of cities (see Table 1): diversified (0,1); specialized (1,0); co-agglomerated (1,1) and non-typified (0,0). The last refers to cities that cannot be specified in terms of specialization or diversification.

[10] For Spain, the OECD identifies 76 urban areas, but we limit our analysis to urban areas for which there exist employment data and transport costs both at the municipality level according to the SABI and in the transport costs databases (see Sections 3.2 and 3.3). Urban areas located outside the Iberian Peninsula (Spanish islands and autonomous cities of Ceuta and Melilla) are not considered.

Table 1: Categorization of Cities.

Type of City	Diversified ($D_{it} = 1$)	Non-Diversified ($D_{it} = 0$)
Specialized ($S_{it} = 1$)	(1,1)	(1,0)
Non-Specialized ($S_{it} = 0$)	(0,1)	(0,0)

To calculate the diversification and specialization of cities, we rely on the so-called SABI database, a unique firm-level database for services and manufacturing sectors at the municipality level (NUTS-5). The dataset is produced by Bureau van Dijk on the basis of information registered in the Spanish Registry of Commerce (Registro Mercantil). This is similar to databases for measuring Marshallian economies (Henderson 2003). SABI identifies the municipality in which each firm is located, the sector in which it operates and its number of workers. For our period of analysis (1995-2007) it includes economic and financial information for about 1.3 million firms. We use the NACE classification to aggregate employment information for up to 38 different sectors, including agriculture, manufacturing, banking, services and public-sector activities.[11]

Decentralization

The decentralization of governance encompasses fiscal, political and administrative decentralization. These refer, respectively, to the ability of subnational governments to raise revenues, obtain decision-making authority, and deliver public goods and services. We use two variables to measure decentralization in our regression analysis. First, we include a dummy (Reg_Gov_i) to qualify cities as regional capitals, defined as cities that host institutional and regional government headquarters. In our sample this applies to 15 of our 69 urban areas.[12] Second, we use the frequently used Regional Authority Index (RAI) developed by Hooghe et al. (2016). This cross-country index aims to capture the degree of regional government authority over the period 1950-2010 and is perhaps the most comprehensive attempt to measure the political dimension of decentralization and sub-national autonomy (Ezcurra and Rodriguez-Pose 2013). Its regional authority concept accounts for the following dimensions: institutional depth, policy scope, fiscal autonomy, borrowing autonomy, representation, law-making, executive control, fiscal control, borrowing control and constitutional reform. To the best of our knowledge, it is also the only decentralization index that goes beyond aggregated country-level information to provide disaggregated information for regions. We use the RAI data at the NUTS-3 (provinces) level for Spain between 1995-2007, plus the year 1980, and impute the same index value to all the cities in each province, assuming a similar regional decentralization in

[11] See Tables B.2, B.3 and B.4 in suplementary Appendix B for details.

[12] This definition was set as part of the decentralization. Most of these cities are officially recognized as regional capital cities. Exceptions include Vitoria and Valladolid, which are considered "institutional cities" only.

all cities of that province.[13]

Accessibility

Obviously, the relative accessibility of urban areas plays a key role in determining the specialization patterns of an urban system. In our analysis we operationalize this by calculating the Relative Market Potential (RMP) of each urban area, as a function of population size and transport costs. Following Head and Mayer (2010), we define the RMP for city *i* as:

$$RMP_{it} = \underbrace{\frac{Urban\ Population_{it}}{GTC_{iit}}}_{Internal\ RMP} + \underbrace{\sum\nolimits_{j\neq 1}^{N} \frac{Urban\ Population_{jt}}{GTC_{ijt}}}_{External\ RMP} \tag{3.2}$$

with an urban-area population defined as before and GTC standing for Generalized Transport Costs; more on this below. Equation 3.2 implies that our measure of RMP identifies each urban area's degree of both internal and external accessibility. The internal RMP accounts for a city *i*'s home-market effect, and the external RMP accounts for its accessibility to other cities. Using annual data for population size and GTC, we calculate annual time-varying indices for both internal and external RMP. This allows us to accurately decompose the change in a city's total RMP into changing within-city population or transport costs dynamics and changes of the city's relative position within the system of cities – which we think is the most appropriate way to capture urban dynamics in a system of cities (see, for example, Anas and Xiong 2005 and Tabuchi and Thisse 2011).

Transport costs are usually bilateral time-invariant measures (physical units in kilometers or travel time in minutes) between an origin and a destination. Clearly, more precise and time-variant measures of transport costs are to be preferred over time-invariant proxies such as physical distance or travel time (Combes and Lafourcade 2005, Díaz-Lanchas et al. 2019). We therefore resort to the Generalized Transport Costs (GTC) measure created by Zofío et al. (2014) for Spain. Those authors use a digitalized road map and GIS software (Arc/GIS) to calculate the GTC as the least-cost itinerary between an origin *i* and a destination *j*, as follows:

$$GTC_{ij} = min\left(Distance\ Cost_{ijt} + Time\ Cost_{ijt}\right) \tag{3.3}$$

The GTC measure thus differentiates the economic costs related to both distance and time,

[13] The RAI includes time-variant identical scores for all Spanish provinces except those located in regions with more autonomy. These are: the Basque Country, Galicia, Navarra and Catalonia. The RAI is calculated for regions (NUTS-2) and provinces (NUTS-3) with similar scores. We take those at the province level as the closest spatial measure to cities.

and accounts for their time-varying components. The distance economic cost (euros per kilometer) includes: fuel (fuel price); tolls (unit cost per km, multiplied by the length of the road); accommodations and allowance; tires; and vehicle maintenance, repair and operation. Separately, the time economic cost (euros per hour) includes: labor (gross salaries), financial costs associated with amortization, insurance, taxes, the financing costs of the vehicle, and indirect costs associated with other operating expenses, including administration and commercial costs. Zofío et al. (2014) combine all these economic components of the GTC into an annual time-varying bilateral GTC measure for 678 transport zones in Spain, distinguishing intra-zone transport costs (GTC_{iit}) from inter-zone transport costs (GTC_{ijt}).

We aggregate these inter and intra GTCs to the level of our 69 urban areas. To this end, we define an average Generalized Transport Cost, $\overline{\mathrm{GTC}}$, as the average minimum economic transport cost from city *i* to all destinations *j* across the system, including the original bilateral GTCs in a system of N cities, as follows:

$$\overline{GTC}_{it} = \frac{\sum_{i=1}^{j} GTC_{ijt}}{N} \tag{3.4}$$

Note that Equation (3.4) includes city *i* as a possible destination to control for intra-city transport costs. The digitalized road networks to calculate bilateral GTCs are available every five years from 1980 to 2005 plus the year 2007. Given the annual availability of data in the SABI database for the period 1995-2007, we use the road network from 1995 onwards, and linearly interpolate for the remaining years. Next, we calculate the average $\overline{\mathrm{GTC}}$ for each city *i* and year, to obtain our annual time-varying transport cost measure.

Agglomeration Economies

In our analysis we control for different dimensions of agglomeration economies, including education level, sectoral composition and the degree of product standardization. Education level – measured as the share of highly educated people in city *i* – is included to show that larger cities tend to attract more highly skilled workers (Glaeser et al. 2014), but also to account for possible key specialization drivers in diversified and co-agglomeration economies (Ellison et al. 2010). Data on education levels are obtained from the Spanish Statistical Institute (INE) Census. We measure for each city *i* its sector composition as the ratio of manufacturing workers to service workers ($Ratio\_MS_{it}$). Finally, we control for the fact that diversified and specialized cities also differ in their degree of product standardization. Duranton and Puga (2001) argue that firms move to specialized cities once their internal economies of scale afford them the efficiency gains to produce standard products. Clark and Stanley (1999) propose a measure of standardization based on plant-level scale economies and minimum efficient scales (MES). Accordingly, there exists a positive correlation between the MES and product standardization: i.e. standardized products are the result of plant-scale economies whose costs decline as plant size increases. We take this industry-level definition (Clark and Stanley 1999, Cilasun and Günalp 2012) and apply it at the city level. We define MES (product standardization) as the average sales per firm (*p*) corresponding to the first *P* largest firms out of the total number of firms *F* located in city *i* such that they account for at least 50% of the city's total sales:

$$MES_{it} = \frac{\sum_{p=1}^{P} sales\_\text{per}\_firm_i^p}{P_i} \mid \left(\frac{\sum_{p=1}^{P} sales\_\text{per}\_firm_i^p}{\sum_{f=1}^{F} sales\_\text{per}\_firm_i^f} \right) \geq 50\% \tag{10}$$

The data on sales and sectoral employment needed for $Ratio\_MS_{it}$ and MES_{it} originate in the SABI database.

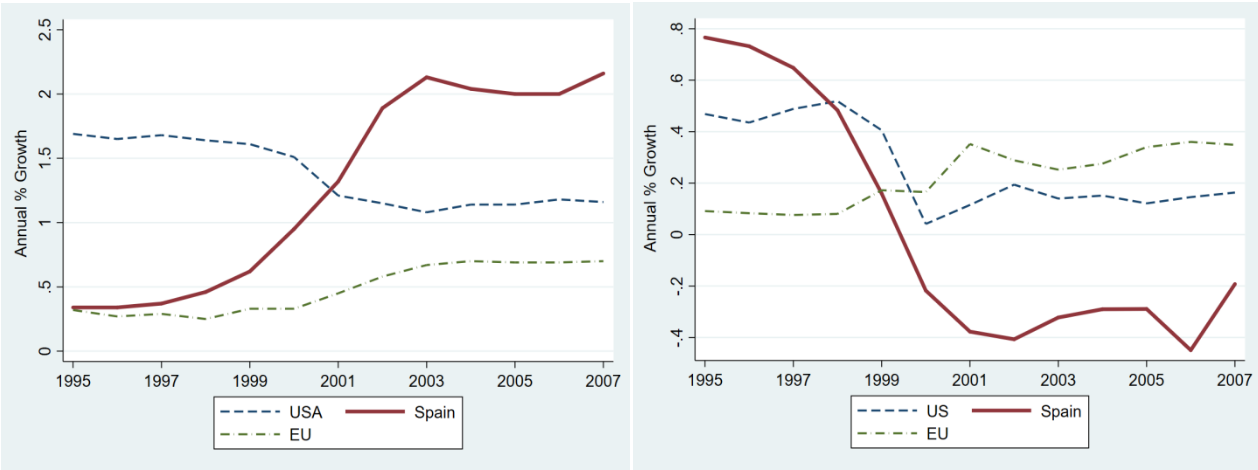
3.2 Descriptive Analysis

As noted, the urban system in Spain has changed considerably over the past decades. The urban dynamics encompass at least three dimensions. First, recent urban population growth in Spain has been relatively high, as illustrated by the left side of Figure 1: from 1995 to 2007, annual urban growth rates converged to around 1% in the US and the EU, whereas in Spain the rate increased from less than 0.5% per year in 1995 to over 2% per year after 2002. Second, Spain's accelerating urbanization manifested itself through the apparent relatively strong growth of medium-size cities, as illustrated on the right side of Figure 1: the population share living in the largest cities (more than a million people) declined sharply. We can confirm the trend towards a more even city-size distribution by calculating two well-known measures: the Gini population index and Zipf's coefficient.[14] These findings contrast with facts for the US, where the distribution of relative city sizes and individual city specializations are quite stable over time (Black and Henderson 1998).

[14] Zipf's coefficient is obtained as β coefficient from an OLS regressions of the (log) rank of the city on its (log) population: $lnRank_{it} = \alpha_i + \beta_i lnPop_{it} + \varepsilon_{it}$, where t = 1980, 1995 and 2007. The Zipf coefficient increased from 0.901 in 1980 to 1.030 in 2007. The Gini population index decreased from 0.644 in 1980 to 0.590 in 2007.

Figure 1: Urban population patterns between 1995 and 2007.

(a) Total urban population growth (Growth rate) | (b) % population in cities >1 million (Growth rate)



Source: Own elaboration from World Bank and Census databases.

Third, during the period 1995-2007 it was more the rule than the exception that Spanish cities fundamentally changed their economic structure, with an increasing number of them diversifying their production structure. Table 2 summarizes these dynamics by presenting a transition matrix for our typology of cities (using the 1995 threshold to define a cities' typology in both years). First, the diagonal of Table 2 indicates that purely diversified cities are the most prominent and stable cities (main diagonal). This stability (path-dependence) motivates us to control our econometric analysis for levels of specialization and diversification in 1995. Second, and more importantly, Table 2 shows that only 28 of the 69 cities considered (i.e. 40%) maintain their original typology over time. Hence, cities that change their typology over time are more the rule than the exception. More specifically, cities that were non-typified and specialized in 1995 tended to diversify their economic structure by becoming either diverse cities or co-agglomeration economies by 2007. Finally, co-agglomerated cities maintained their situation or lost their specialization in favor of diversified cities.

Table 2: Transition matrix for each type of city (1995-2007)

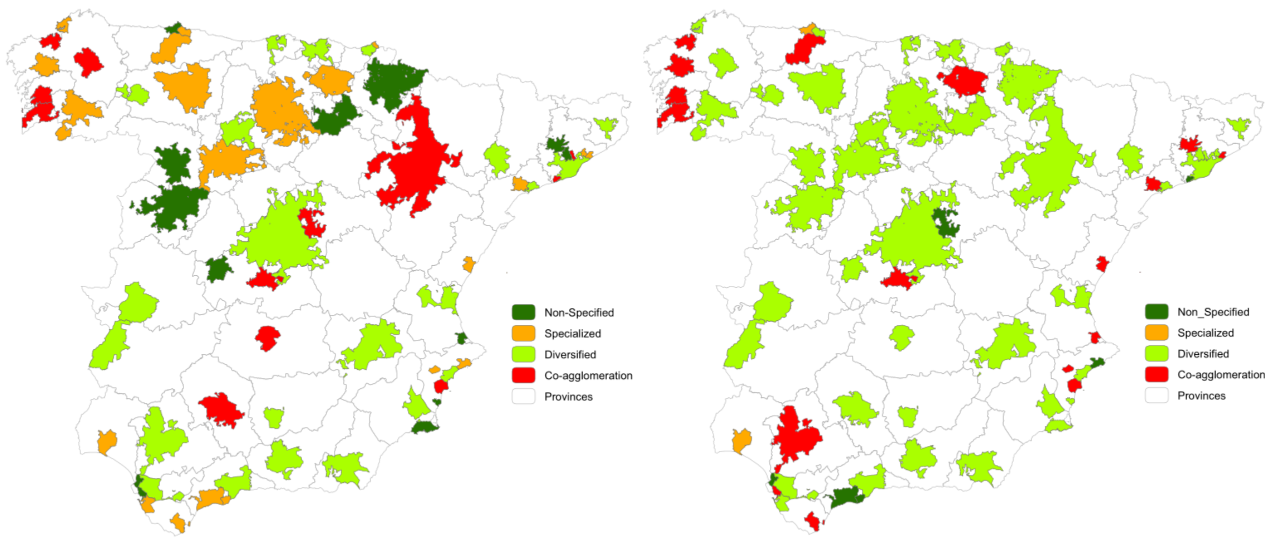
	City (2007)				
City (1995)	**Non-typified**	**Specialized**	**Diversified**	**Co-agglomerated**	**Total**
# Non-typified	1	1	8	3	13
Share (%)	*7.69*	*7.69*	*61.54*	*23.08*	*18.84*
# Specialized	3	1	9	9	22
Share (%)	*13.64*	*4.55*	*40.91*	*40.91*	*31.84*
# Diversified	0	0	21	1	22
Share (%)	*0.00*	*0.00*	*95.45*	*4.55*	*31.84*
# Co-agglomerated	2	0	5	5	12
Share (%)	*16.67*	*0.00*	*41.67*	*41.67*	*17.39*
# Total	6	2	43	18	69
Share (%)	*8.70*	*2.90*	*62.32*	*26.09*	*100*

Source: Own elaboration from SABI Database. Note: Type of city in 1995 and 2007 is defined using cities' definition in 1995.

Using our typology of cities, we visualize in Figure 2 the urban diversity dynamics across space by presenting maps of the urban system in 1995 (left) and in 2007 (right). The left side of Figure 2 shows that in 1995 diversified cities (0,1) were either mainly province capitals (Sevilla, Albacete, Alicante and Badajoz, among others) or the richest cities in Spain (Madrid, Barcelona, Bizkaia and Valencia). These cities were predominantly located in the southeast and center of the country. Co-agglomerated cities (1,1) were either located farther from the center (Lugo, Pontevedra, Ciudad Real) or consisted of very small cities surrounding larger diverse cities (Toledo, Guadalajara), except for Zaragoza, which is one of the most populated cities in Spain. Specialized cities (1,0) were mainly located in the northwest and close to cities with a certain diversity. Finally, non-typified cities (0,0) formed a dispersed pattern, being often located near diverse cities. The right side of Figure 2 shows that in 2007 many cities had diversified their economic structure. Cities in the northwest that used to be non-typified or purely specialized became either diversified (e.g. Valladolid, Salamanca, León) or co-agglomerated (e.g. Oviedo, La Coruña, Pontevedra, Álava). At the same time, several large diverse cities, such as Seville, transformed into co-agglomerated cites. In contrast, some co-agglomerated cities, like Zaragoza, lost their specialization pattern and developed into diversified cities. Various smaller cities near the biggest cities became either purely specialized (e.g. Guadalajara) or diversified (e.g. Pamplona, Cádiz and Manresa).

Figure 2: Four type of cities in Spain, 1995 and 2007.

(a) 1995 (b) 2007



Source: Own elaboration from SABI database. Note: Type of city in 1995 and 2007 is defined using cities' 1995 definition.

We conclude this section by presenting in Tables 3 and 4 a summary of facts and growth dynamics about our key variables, which we believe are the driving forces of the specialization and diversification patterns described above. There we distinguish four types of cities by specialization (see Table 1) and by the presence or absence of regional government headquarters. Table 3 shows that, on average for 1995-2007, non-typified cities are mainly of small and medium size in population. They feature a moderate population of highly educated residents, moderate transport costs and a preponderance of relatively small firms. Specialized cities face, on average, the highest transport costs (low accessibility) and have the smallest populations and smallest shares of highly educated people. In contrast, diversified cities are the most populated, host the largest share of highly educated people and have the lowest transport costs (high accessibility). Co-agglomerated cities, for their part, tend to be medium-size and big cities with a large share of highly educated people and relatively high transport costs. For the relative market potential, diversified cities and co-agglomeration economies have the largest internal market, whereas non-typified cities present the smallest one. The opposite holds for external market potential. That is, non-typified and specialized cities are surrounded by cities with huge market potential, whereas diversified and co-agglomerated cities mostly depend on their internal markets. The values of the Regional Authority Index (RAI) are identical for specialized and diversified cities; these two opposing city types are thus randomly distributed over regions with different degrees of political decentralization. The ratio of manufacturing to service workers is, on average, highest in specialized and non-typified cities and lowest in diverse ones. Product standardization is higher in cities with some sort of specialized structure (specialized and co-agglomerated cities) or even in non-typified cities. Cities with regional government headquarters are relatively large and diversified, featuring a relatively high internal and low external market potential, a relatively high share of educated people and, interestingly, also a relatively high ratio of manufacturing to service workers and high product standardization. RAI values are identical for cities with and without regional government headquarters. Evidently, regional capitals can be found in regions with different degrees of political decentralization; this confirms that our two measures of decentralization do not coincide.

The summary of growth dynamics in Table 4 shows that diversification increased most in co-agglomerated cities, followed by diversified cities. Specialization increased most in specialized cities and declined in non-typified cities. Population growth has been highest in non-typified cities for the period 1980-2007, followed by co-agglomerated and diversified cities in the period 1995-2007. Transport costs declined for all type of cities but declined the most for diversified cities up to 1995 and for specialized cities after 1995. Internal market potential increased the most in non-typified cities and the least in specialized cities. External market potential increased the most in specialized cities. Cities with regional government headquarters featured relatively high population growth and an increase in market potential, especially internal market potential. Their increase in diversification was remarkably lower than that of cities without regional government headquarters. By contrast, their decline in specialization and transport costs was not significantly different.

Table 3: Descriptive statistics (average values for 1995-2007)

Type of City	Stats	RDI Index	RZI Index	Population (1000s)	$\overline{GTC}$ (Euros)	Internal RMP (Logs)	External RMP (Logs)	RAI (Logs)	High Education (% pop.)	Ratio M/S	MES (Logs)
Non-typified	Mean	1.65	9.48	144.14	661.16	7.48	11.00	3.13	8.41	0.60	14.93
	Median	1.63	5.60	115.01	626.88	7.51	11.01	3.15	8.05	0.47	14.78
	Std	0.35	14.19	75.45	156.46	0.56	0.18	0.03	3.39	0.43	0.61
Specialized	Mean	1.66	21.38	181.26	739.75	7.76	10.88	3.09	8.22	0.53	14.83
	Median	1.61	9.78	168.22	774.98	7.78	10.88	3.11	8.06	0.42	14.77
	Std	0.36	31.52	91.39	162.48	0.71	0.24	0.08	2.95	0.44	0.49
Diversified	Mean	2.33	3.76	774.69	645.6	8.43	10.89	3.09	10.02	0.29	14.88
	Median	2.18	3.25	255.46	638.89	8.17	10.88	3.11	9.63	0.27	14.77
	Std	0.72	2.22	1,285.34	146.43	1.21	0.19	0.08	2.66	0.13	0.49
Co-agglomerated	Mean	1.94	17.83	245.622	681.72	7.98	10.97	3.12	9.60	0.45	14.93
	Median	1.89	7.65	177.95	673.80	7.81	10.98	3.11	9.26	0.41	14.81
	Std	0.44	25.06	191.75	218.65	0.86	0.34	0.03	3.04	0.24	0.52
Regional Capitals	Mean	2.39	8.23	1,009.06	612.04	8.71	10.88	3.11	11.19	0.53	15.20
	Median	2.27	4.64	311.00	586.00	8.39	10.89	3.11	10.80	0.43	15.17
	Std	0.78	9.73	1,486.77	161.04	1.21	0.25	0.08	2.81	0.33	0.49
No Regional Capitals	Mean	1.79	14.20	198.452	705.05	7.75	10.93	3.11	8.48	0.43	14.79
	Median	1.72	5.59	150.163	728.93	7.63	10.95	3.11	8.29	0.32	14.69
	Std	0.45	25.11	163.829	169.53	0.77	0.24	0.07	2.87	0.36	0.49
Total	Mean	1.92	12.90	374.67	684.83	7.96	10.92	3.11	9.07	0.45	14.88
	Median	1.89	5.38	176.84	670.32	7.78	10.94	3.11	8.83	0.35	14.78
	Std	0.59	22.80	782.00	172.16	0.97	0.24	0.07	3.07	0.35	0.52

Source: Own elaboration from SABI, GTC and Urban Areas Databases.

Table 4: Population, GTC and RMP patterns by type of city. Growth rates.

	RZI	RDI	Population		GTC		Internal RMP		External RMP	
Type of city	**95-07**	**95-07**	**80-95**	**95- 07**	**80-95**	**95-07**	**80-95**	**95-07**	**80-95**	**95-07**
Non-typified	-34.4%	-3.3%	27.3%	34.6%	-12.2%	-7.4%	36.4%	37.6%	19.7%	19.9%
Specialized	76.6%	-20.9%	4.8%	3.7%	-13.5%	-10.2%	14.2%	9.6%	20.0%	21.0%
Diversified	-55.4%	25.1%	9.0%	13.7%	-14.1%	-8.4%	18.7%	17.5%	18.3%	20.4%
Co-agglomerated	20.6%	29.4%	9.9%	15.4%	-12.4%	-9.8%	18.9%	21.2%	17.1%	21.6%
Regional Capitals	-30.2%	20.7%	12.0%	16.3%	-13.6%	-8.9%	21.0%	20.8%	18.4%	21.0%
No Regional Capitals	-29.0%	28.8%	6.5%	13.2%	-13.1%	-8.2%	15.9%	17.0%	17.5%	19.8%

Source: Own elaboration. Note: Growth rates are calculated with cities' definition in 2007.

4. **Econometric Analysis**

4.1 **Econometric Specification**

In this section we develop a bivariate probit econometric approach to identify the role of key mechanisms driving the observed patterns of diversification and specialization in Spanish cities. More specifically, we estimate the probability of a city *i* to become either specialized (S_i) or diversified (D_i). We do so by combining two independent probit models for city *i*, where 1 (specialization) and 2 (diversification) identify each equation (t subscripts have been removed):

$$S_{i1}^* = X_{i1}\beta_1 + \varepsilon_{i1}, \quad S_{i1} = 1 \quad if \ S_{i1}^* > 0, 0 \ otherwise \tag{4.1a}$$

$$D_{i2}^* = X_{i2}\beta_2 + \varepsilon_{i2}, \quad D_{i2} = 1 \quad if \ D_{i2}^* > 0, 0 \ otherwise \tag{4.1b}$$

Where $\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \Big| X_1, X_2 \Big) \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$

S_{i1}^* and D_{i2}^* indicate each type of city. The key point in this setting is ρ, the *tetrachoric correlation* between ε_{i1} and ε_{i2}. If $\rho = 0$, both expressions (4.1a) and (4.1b) are independent, so we could simply estimate them as two independent probit models. This would imply that we estimate the single probability of getting a diversified or a specialized city. By contrast, if $\rho \neq 0$, and ε_{i1} and ε_{i2} are correlated, we must estimate expressions (4.1) and (4.2) simultaneously as a bivariate probit model (Greene, 2012). In a more general setting, the log-likelihood function would take the form:

$$logL = \sum_{i=1}^{2} log\Phi_2 \begin{bmatrix} (2S_{i1} - 1) \\ (2D_{i2} - 1) \\ (2S_{i1} - 1)(2D_{i2} - 1)\rho \end{bmatrix} = \sum_{i=1}^{2} log\Phi_2[q_{i1}\beta_1 X_{i1}, q_{i2}\beta_2 X_{i2}, q_{i1}q_{i2}\rho] \tag{4.2}$$

where $logL$ refers to the log-likelihood function, $q_{i1} = (2S_{i1} - 1) = -1$ if $S_{i1}^* = 0$ and $q_{i1} = 1$ if $S_{i1}^* = 1$; $q_{i2} = (2D_{i2} - 1) = -1$ if $D_{i2}^* = 0$ and $q_{i2} = 1$ if $D_{i2}^* = 1$. Note that Equation (4.2) is dependent on ρ. Now let $\omega_{i1} = \rho_{i1}X_{i1}\beta_1$ and $\omega_{i2} = \rho_{i2}X_{i2}\beta_2$. Thus, the probabilities entering Equation (4.2) are:

$$Prob(S_{i1}^* = S_{i1}, D_{i2}^* = D_{i2}|X_1, X_2) = \Phi_2(\omega_{i1}, \omega_{i2}, q_{i1}q_{i2}\rho) \tag{4.3}$$

The probabilities change as long as $\rho \neq 0$. To the extent that expressions (4.1a) and (4.1b) are dependent, Equation (4.3) would estimate the joint probability for a city to be diversified and specialized at the same time. These joint probabilities exactly correspond to the four city typologies we distinguished in Table 1, implying a 2x2 probability matrix as presented as Table 5.

Table 5: Probability of each city type

Probability of city type	Diversified ($D_{it} = 1$)	Non-Diversified ($D_{it} = 0$)
Specialized ($S_{it} = 1$)	P(1,1)	P(1,0)
Non-Specialized ($S_{it} = 0$)	P(0,1)	P(0,0)

In order to estimate these probabilities, we take logarithms in (4.1a) and (4.1b) and include time (γ_t) and spatial fixed effects (μ_i),[15] to obtain the expressions (now with *t* subscripts):

$$S^*_{i1t} = \alpha_1 + lnX_{i1t}\beta_1 + \gamma_{1t} + \varepsilon_{i1t} \quad (4.4a)$$

$$D^*_{i2t} = \alpha_2 + lnX_{i2t}\beta_2 + \gamma_{2t} + \varepsilon_{i2t} \quad (4.4b)$$

Vectors X_{i1} and X_{i2} are two X-vectors that include, as regressors, the range of variables presented in Section 3, plus a dummy variable indicating whether the city is a regional government capital (Reg_Gov_{it}):

$$X_{i1t} = (lnPop_{it}, ln\overline{GTC}_{it}, lnRMP\_Internal_{it}, lnRMP\_External_{it}, RegGov_{it}, lnRAI_{it}, lnSh\_HighEduc_{it}, lnRatio\_MS_{it}, lnMES_{it}, lnRZI_{1995}) \quad (4.5a)$$

$$X_{i2t} = (lnPop_{it}, ln\overline{GTC}_{it}, lnRMP\_Internal_{it}, lnRMP\_External_{it}, Reg\_Gov_{it}, lnRAI_{it}, lnSh\_HighEduc_{it}, lnRatio\_MS_{it}, lnMES_{it}, lnRDI_{1995}) \quad (4.5b)$$

We hypothesize that our governance-decentralization variables (Reg_Gov and RAI) positively impact the probability that cities diversify and negatively impact the probability that cities specialize. Falling inter-city trading cost and relatively low levels of external RMP may advance the emergence of either diversified cities (cf. Tabuchi and Thisse 2011) or specialized cities, as high transport costs cause economic activity – and thus population – to concentrate (inefficiently) in diversified cities (cf. Anas and Xiong 2005). We expect higher levels of internal RMP – i.e. relatively large cities that have good internal accessibility – to favor diversification. The remaining covariates control for agglomeration economies. We hypothesize that the greater the share of highly educated people, the higher the probability that a city diversifies (Glaeser, et al. 2014, Viladecans-Marsal 2004). By contrast, we assume that the ratio of manufacturing to service workers and product standardization increases the probability that a

[15] The RAI is collinear with the spatial effects (μ_i). To assess the impact of this index, in the baseline specifications we consider only time-fixed effects. In Table A.1 in Annex A we include the probit estimations with spatial effects and without adding the RAI. These spatial effects are at the NUTS-2 level (Autonomous Communities). The use of NUTS-2 spatial effects makes sense in the way administrative levels are organized in Spain. They comprise NUTS-3 and urban levels, with NUTS-2 representing fiscally autonomous regions with remarkable differences across them in terms of infrastructure and regional and tax policy.

city is specialized (Henderson 1974, Duranton and Puga 2000). Note that population and GTC cannot be included jointly with the RMP variables, as the latter are composed by the former. As argued in Section 3, RMP variables are a better measure of city dynamics, as they simultaneously incorporate changes in both population and transport costs and permit a distinction between intra- and inter-city components. Therefore, in the main text we show regression results for RMP variables only; regression results for population and GTC are presented in the suplementary Appendix to this paper (Tables A.3 and A.4).

4.2 **Results**

Below we summarize our main regression results. In Section 5 we discuss our findings in more detail, in relation to the existing literature to date. Tables 6 and 7 shows the estimated coefficients for the two probit models for the specifications in (4.5a) and (4.5b). The left and right sides of Table 6 indicate, respectively, the probability that a city is specialized or diversified. Models (1) and (4) include the regression results for our most elementary specification; models (2) and (5) add the decentralization variables, Reg_Gov_i and RAI; models (3) and (6) interact the relative market potential (RMP) variables with decentralization variables and constitute our complete baseline specifications.[16] In Table 7, we present the regression results for our two probit models with lagged (1980) RMP and RAI variables, to account for potential endogeneity problems between these variables and the typology of cities.

All regression models in Tables 6 and 7 yield ρ values that are statistically significant and different from zero, implying that the two probit models are indeed interdependent and thus should be estimated simultaneously in a bivariate regression framework. Moreover, the negatives values for ρ indicate a negative correlation between both types of probabilities. In addition, the regression results for all models in Tables 6 and 7 show that, as expected, the probability that cities are specialized increases as the share of highly educated people decreases, the ratio of manufacturing to service workers increases and the degree of product standardization increases. The opposite is true for the probability that cities are diversified. Furthermore, our results provide robust evidence of path dependency: the probability that cities are specialized (diversified) is influenced by historical (1995) levels of specialization (diversification).

As for the role of decentralization – our key issue of interest – we find that a city's being a regional government capital (Reg_Gov) positively affects the probability that it will diversify, while a city's degree of decentralization (RAI) positively (negatively) affects the probability that it will diversify (specialize). Also, we find no statistically significant effect for the interaction term between the regional government capital dummy and the degree of decentralization (RAI); this is reassuring, as it confirms that a city’s status as regional government capital does not relate to its region's degree of decentralization (cf Table 3).

[16] In models (3) and (6), Reg_Gov_i is included only in interaction but not independently to avoid collinear problems between the former and the latter.

The regression results of models (1)-(2) and (4)-(5) in Table 6 show that the coefficient for internal RMP – the home-market effect – is negative (positive) for models in which the dependent variable is the probability of a city to become specialized (diversified); the negative coefficients in models (1)-(2), however, are statistically insignificant. The coefficients for external RMP – the inter-city accessibility effect – are negative and statistically significant in all four models. However, when we interact our respective decentralization and RMP variables in models (3) and (6) we find that a city's being a regional government capital qualifies the effect of RMP on the probability that it will develop either typology. The coefficient of the interaction term between the regional government capital dummy and internal RMP is statistically significant and negative (positive) if the dependent variable is the probability that cities are specialized (diversified). The opposite holds for the coefficient of the interaction term between the regional government capital dummy and external RMP. Together, these results first imply that a higher internal (external) RMP unconditionally increases (decreases) the probability that a city is diversified, especially if a city hosts a regional government headquarters. Second, and by contrast, our results imply that only when a city is a regional capital does a higher internal RMP decrease the probability that the city is specialized. A city's being a regional capital strongly counters the negative impact of a high external RMP on the probability of its specialization. If we replace the RMP and RAI variables by their lagged values (in Table 7), we find very similar results. It is true that the direct effect of internal RMP on the probability that a city will diversify becomes statistically insignificant, but the interaction between internal RMP and the regional government capital dummy in this case is again statistically significant and positive, and more so than before.

In Table 8 we present for each type of city the implied joint marginal effects from the bivariate estimation results. The provided average marginal effects have the same interpretation as in an independent probit model but take into account the joint determination of our two endogenous variables (S_{i1}^* and D_{i2}^*). The marginal effects presented on the left of Table 8 are based on models (2) and (4) from Table 6; the marginal effects presented on the right of Table 8 are based on models (3) and (6) from Table 6. The results shown in Table 8 clearly reinforce our finding that a regional government capital or a city located in a relatively decentralized region has an increased (decreased) probability of diversifying (specializing). Also, it reinforces and particularizes our findings on the interaction between decentralization and the role of relative market potential. A city's being a regional government capital reinforces the positive effect of a high (low) internal RMP – the home-market effect – on the probability of diversification (specialization). Only for a regional capital does a high (low) external RMP – the inter-city accessibility effect – increase the probability of specialization (diversification).

The various marginal effects for non-typified and co-agglomerated cities lead to a mixed result between these two opposite types of cities. A city's probability of becoming non-typified city (i.e. of developing no clear typology) decreases if it is a regional government capital but increases if it is located in a relatively decentralized region. The opposite holds for a city's probability of becoming co-agglomerated. Finally, the average marginal effects for the various agglomeration economy variables as well as the historical levels of diversification and specialization have the expected (significant) effects, thus confirming the previous findings.

We conclude this section by presenting in Table 9 the average marginal effects obtained from the regression results presented in Table 7, thus using lagged (1980) values for our RAI and RMP variables. The table shows that most of our results are robust to this control for potential endogeneity problems. It is to be noted, however, that the marginal effects of an independent effect of internal RMP in model (1) now become statistically insignificant. On the other hand, the marginal effects of the interaction term between the lagged internal RMP variable and the regional government capital dummy remain statistically significant and are stronger than before. The latter also holds for the interaction term between the lagged external RMP variable and the regional government capital dummy. Also, note that in Table 9 the marginal effects for the impact of the lagged (1980) degree of decentralization (RAI_80) are weaker than they would be for current values (Table 8). This is evident, as decentralization in Spain started only in 1978 and the degree of decentralization increased (substantially) over time.

Table 6: Probit estimations – Baseline models

	$Pr(S_i=1\|X)$			$Pr(D_i=1\|X)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$ln(MP\_internal_i)$	-0.081	-0.088	0.145*	0.240***	0.207**	0.124
	(0.056)	(0.059)	(0.080)	(0.077)	(0.082)	(0.088)
$ln(MP\_External_i)$	-0.749***	-0.628***	-0.786***	-0.660***	-0.710***	-0.597**
	(0.214)	(0.215)	(0.250)	(0.221)	(0.223)	(0.236)
Reg_Gov_i	---	-0.055	---	---	0.352**	---
		(0.135)			(0.137)	
$Ln(RAI_i)$	---	-3.351***	-2.618***	---	1.778***	1.263*
		(0.719)	(0.781)		(0.623)	(0.702)
$Reg\_Gov*ln(MP\_int_i)$	---	---	-0.741***	---	---	0.392*
			(0.177)			(0.204)
$Reg\_Gov*ln(MP\_ext_i)$	---	---	0.616**	---	---	-0.566**
			(0.304)			(0.28)
$Reg\_Gov*ln(RAI_i)$	---	---	-0.207	---	---	1.067
			(1.055)			(1.038)
$ln(sh\_HighEduc_i)$	-1.357***	-1.410***	-1.627***	2.145***	1.991***	2.031***
	(0.210)	(0.227)	(0.248)	(0.228)	(0.242)	(0.245)
$Ln(ratio\_MS_i)$	0.358***	0.354***	0.365***	-0.059	-0.070	-0.064
	(0.072)	(0.073)	(0.074)	(0.065)	(0.065)	(0.064)
$Ln(MES_i)$	0.362***	0.312**	0.305**	-0.608***	-0.609***	-0.611***
	(0.127)	(0.125)	(0.127)	(0.126)	(0.128)	(0.128)
$Ln(RZI1995_i)$	0.862***	0.902***	0.819***	---	---	---
	(0.094)	(0.099)	(0.099)			
$Ln(RDI1995_i)$	---	---	---	1.308***	1.375***	1.337***
				(0.224)	(0.227)	(0.227)
ρ	-0.375***	-0.357***	-0.334***	-0.375***	-0.357***	-0.334***
	(0.068)	(0.068)	(0.069)	(0.068)	(0.068)	(0.069)
N	897	897	897	897	897	897
Year Fixed-Effects	YES	YES	YES	YES	YES	YES
Spatial Fixed-Effects	NO	NO	NO	NO	NO	NO

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01,

Table 7: Probit estimations – 1980 models

	Pr(S_i=1|X)		Pr(D_i=1|X)	
	(7)	(8)	(9)	(10)
ln(MP_Internal80$_i$)	-0.017	0.203***	0.010	-0.075
	(0.053)	(0.071)	(0.064)	(0.070)
ln(MP_External80$_i$)	-0.323*	-0.457**	-0.573***	-0.492***
	(0.183)	(0.211)	(0.190)	(0.191)
Reg_Gov$_i$	-0.214	---	0.593***	---
	(0.140)		(0.136)	
Ln(RAI80$_i$)	0.002	0.863	1.247***	0.619
	(0.485)	(0.555)	(0.451)	(0.498)
Reg_Gov*ln(MP_int80$_i$)	---	-0.917***	---	0.475**
		(0.213)		(0.193)
Reg_Gov*ln(MP_ext80$_i$)	---	0.677***	---	-0.713***
		(0.194)		(0.253)
Reg_Gov*ln(RAI80$_i$)	---	-0.478	---	2.206*
		(0.994)		(1.319)
ln(sh_HighEduc$_i$)	-1.278***	-1.597***	1.937***	2.065***
	(0.221)	(0.250)	(0.238)	(0.246)
Ln(ratio_MS$_i$)	0.370***	0.316***	-0.114	-0.058
	(0.074)	(0.076)	(0.070)	(0.069)
Ln(MES$_i$)	0.397***	0.426***	-0.544***	-0.520***
	(0.130)	(0.135)	(0.123)	(0.123)
Ln(RZI1995$_i$)	0.876***	0.760***	---	---
	(0.099)	(0.095)		
Ln(RDI1995$_i$)	---	---	1.476***	1.434***
			(0.221)	(0.225)
ρ	-0.351***	-0.314***	-0.351***	-0.314***
	(0.067)	(0.069)	(0.067)	(0.069)
N	897	897	897	897
Year Fixed-Effects	YES	YES	YES	YES
Spatial Fixed-Effects	NO	NO	NO	NO

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i* refers to city *i* and 80 indicates variables in 1980. Columns (7) and (8) indicate the probability for a city to become specialized, whereas columns (9) and (10) do the same for the probability of diversification.

Table 8: Average marginal effects – Baseline Models.

	Baseline Model 1				Baseline Model 2			
	Non-typified	Specialized	Diversified	Co-agglomerated	Non-typified	Specialized	Diversified	Co-agglomerated
	P(0,0)	P(1,0)	P(0,1)	P(1,1)	P(0,0)	P(1,0)	P(0,1)	P(1,1)
$\ln(MP\_internal_i)$	-0.021	-0.061***	0.056***	0.026	-0.052**	0.002	-0.006	0.055**
	(0.018)	(0.022)	(0.021)	(0.019)	(0.021)	(0.026)	(0.026)	(0.022)
$\ln(MP\_External_i)$	0.255***	0.028	-0.005	-0.278***	0.266***	-0.028	0.047	-0.285***
	(0.056)	(0.068)	(0.065)	(0.060)	(0.064)	(0.073)	(0.073)	(0.067)
Reg_Gov_i	-0.054**	-0.085**	0.076*	0.063*	---	---	---	---
	(0.027)	(0.042)	(0.045)	(0.035)				
$Ln(RAI_i)$	0.324**	-1.033***	1.009***	-0.299*	0.264	-0.767***	0.780***	-0.276
	(0.153)	(0.223)	(0.216)	(0.161)	(0.176)	(0.235)	(0.237)	(0.187)
$Reg\_Gov*\ln(MP\_int_i)$	---	---	---	---	0.068	-0.224***	0.228***	-0.071
					(0.051)	(0.055)	(0.059)	(0.051)
$Reg\_Gov*\ln(MP\_ext_i)$	---	---	---	---	-0.011	0.236***	-0.235**	0.009
					(0.071)	(0.091)	(0.092)	(0.076)
$Reg\_Gov*\ln(RAI_i)$	---	---	---	---	-0.164	-0.261	0.247	0.179
					(0.238)	(0.338)	(0.335)	(0.256)
$\ln(sh\_HighEduc_i)$	-0.095	-0.699***	0.656***	0.138**	-0.075	-0.735***	0.723***	0.087
	(0.062)	(0.075)	(0.069)	(0.065)	(0.066)	(0.077)	(0.076)	(0.069)
$Ln(ratio\_MS_i)$	-0.056***	0.084***	-0.085***	0.057***	-0.058***	0.084***	-0.087***	0.062***
	(0.018)	(0.021)	(0.020)	(0.019)	(0.018)	(0.020)	(0.021)	(0.020)
$Ln(MES_i)$	0.053*	0.190***	-0.177***	-0.066**	0.058*	0.186***	-0.180***	-0.064*
	(0.030)	(0.041)	(0.039)	(0.032)	(0.031)	(0.041)	(0.040)	(0.033)
$Ln(RZI1995_i)$	-0.176***	0.176***	-0.182***	0.182***	-0.158***	0.158***	-0.168***	0.168***
	(0.020)	(0.020)	(0.022)	(0.022)	(0.020)	(0.020)	(0.022)	(0.022)
$Ln(RDI1995_i)$	-0.256***	-0.292***	0.256***	0.292***	-0.256***	-0.277***	0.256***	0.277***
	(0.043)	(0.051)	(0.043)	(0.051)	(0.044)	(0.051)	(0.044)	(0.051)
N	897	897	897	897	897	897	897	897
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Spatial FE	NO	NO	NO	NO	NO	NO	NO	NO

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i*

Table 9: Average marginal effects – Variables from 1980.

	Baseline Model 1				Baseline Model 2			
	Non-typified	Specialized	Diversified	Co-agglomerated	Non-typified	Specialized	Diversified	Co-agglomerated
	P(0,0)	P(1,0)	P(0,1)	P(1,1)	P(0,0)	P(1,0)	P(0,1)	P(1,1)
ln(MP_internal80i)	0.001	-0.005	0.005	-0.001	-0.024	0.054***	-0.057***	0.027
	(0.015)	(0.018)	(0.017)	(0.016)	(0.018)	(0.021)	(0.021)	(0.019)
ln(MP_External80i)	0.171***	0.057	-0.042	-0.186***	0.183***	0.014	-0.000	-0.196***
	(0.050)	(0.055)	(0.053)	(0.054)	(0.054)	(0.057)	(0.060)	(0.058)
Reg_Govi	-0.071***	-0.159***	0.157***	0.074**	---	---	---	---
	(0.027)	(0.039)	(0.048)	(0.036)				
Ln(RAI80i)	-0.234*	-0.264*	0.233*	0.264**	-0.285**	0.038	-0.059	0.306**
	(0.123)	(0.143)	(0.132)	(0.131)	(0.143)	(0.148)	(0.151)	(0.156)
Reg_Gov*ln(MP_int80i)	---	---	---	---	0.084	-0.273***	0.282***	-0.093
					(0.056)	(0.056)	(0.062)	(0.058)
Reg_Gov*ln(MP_ext80i)	---	---	---	---	0.008	0.276***	-0.278***	-0.006
					(0.056)	(0.069)	(0.071)	(0.058)
Reg_Gov*ln(RAI80i)	---	---	---	---	-0.335	-0.544	0.526	0.354
					(0.285)	(0.367)	(0.361)	(0.298)
ln(sh_HighEduci)	-0.110*	-0.662***	0.619***	0.154**	-0.094	-0.730***	0.731***	0.093
	(0.062)	(0.072)	(0.066)	(0.065)	(0.068)	(0.076)	(0.075)	(0.070)
Ln(ratio_MSi)	-0.052***	0.097***	-0.095***	0.050**	-0.049**	0.072***	-0.077***	0.053***
	(0.019)	(0.022)	(0.021)	(0.020)	(0.019)	(0.021)	(0.021)	(0.021)
Ln(MESi)	0.024	0.193***	-0.181***	-0.036	0.019	0.188***	-0.189***	-0.018
	(0.031)	(0.041)	(0.038)	(0.033)	(0.032)	(0.040)	(0.040)	(0.034)
Ln(RZI1995i)	-0.173***	0.173***	-0.176***	0.176***	-0.145***	0.145***	-0.158***	0.158***
	(0.020)	(0.020)	(0.021)	(0.021)	(0.019)	(0.019)	(0.021)	(0.021)
Ln(RDI1995i)	-0.276***	-0.313***	0.276***	0.313***	-0.277***	-0.294***	0.277***	0.294***
	(0.042)	(0.050)	(0.042)	(0.050)	(0.044)	(0.050)	(0.044)	(0.050)
N	897	897	897	897	897	897	897	897
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Spatial FE	NO	NO	NO	NO	NO	NO	NO	NO

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i* refers to city *i*.

5. **Conclusions**

The worldwide trend to decentralize the responsibilities and budgets of governments impacts local firm dynamics in various ways. In this paper we have developed a bivariate probit regression framework to test empirically whether the decentralization of governance is an important determinant of the sectoral composition of cities in an urban system. We have used a sample of 69 urban areas in Spain during the period 1995-2007, and exploited unique firm-level and time-varying transport-cost data to control our analysis for the role of a city's market potential, city size, transport costs, labor-force skill composition, product standardization and historical patterns of specialization. In around 1980 Spain undertook an intense decentralization of its government, and it has since become one of the most decentralized countries in the world. It has established a set of new regional capitals (15 of out 69 cities), hosting regional government headquarters that developed a complete list of competencies in regional and urban policies.

We have shown that Spain's decentralization was concurrent with a rapid decline in transport costs, relatively high rates of urban growth and an equalization of city-size distribution. In addition, we have observed that many cities changed their economic structure; specialized cities in particular tended to develop diversified or co-agglomerated production structures, whereas diverse cities tended to maintain their economic structure. During the period 1980-1995, specialized and non-typified cities experienced the highest increase in population, and during the period 1995-2007 diversified and co-agglomerated cities clearly caught up. These findings contrast with facts for the US, where the distribution of relative city sizes and individual city specializations is quite stable over time (Black and Henderson 1998).

Our regression results support the hypothesis that governance decentralization and the establishment of regional government headquarters in specific cities have been conducive to an increasing diversification and a more even city-size distribution in the Spanish urban system during a period of continuous reductions in transport costs. More specifically, we have found that the probability of diversification increases for a regional government capital or a city located in a relatively decentralized region, while the opposite holds for a city that specializes. The results for non-typified and co-agglomerated cities, however, lie between the results for the two opposite types of cities. A city's probability of becoming non-typified (i.e. of developing no clear typology) decreases if it is a regional government capital but increases if it is located in a relatively decentralized region. The opposite holds for a city's probability of becoming co-agglomerated. Furthermore, we have found that a city's being a regional government capital reinforces the positive effect of a high (low) internal Relative Market Potential – a home-market effect – on the probability of diversification (specialization). Our results also show that only for a regional capital city does a high (low) external Relative Market Potential – i.e. degree of accessibility within the system of cities – increase the probability of specialization (diversification).

In line with most of the existing literature on growth and transformation patterns across cities (e.g. Abdel-Rahman and Fujita 1993; Duranton and Puga 2000, 2001; Henderson 1974, 2003; Anas and Xiong 2005; Ellison et al. 2010; Helsley and Strange 2014), we have found a higher probability of specialization for cities with relatively low-skilled labor forces, with production structures biased towards manufacturing and standardized products, and of relatively small size. The opposite characteristics increase the probability that a city will diversify, whereas a mix of these results helps explain co-agglomeration economies. Finally, we have found that historical patterns of specialization and diversification help explain current patterns.

We are inclined to interpret Spain's decentralization as a political version of the lumpy adjustment from developers that helps cities – in this case regional government capitals – to diversify in the spirit of the theoretical urban systems models of Henderson and Becker (2000) and Anas and Xiong (2005). The transfer of political power from a central administration to regions and localities enabled a series of relatively small cities to diversify their economic structure and thereby counterbalance the effects of the continuous fall in transport costs that began in the 1980s. Consequently, Spain's urban system has become more egalitarian over time. Our findings also lend support to the idea that decentralization may help explain why second-tier cities often outperform first-tier cities (as in Meijers et al. 2016, McCann and Acs 2011). We have shown that governance decentralization causes local institutional capacity to interact with transport costs and city size in driving industry composition. Evidently, industry composition shapes the role of cities as engines of economic growth and drivers of spatial inequality. The economic (geography) literature about the impact of fiscal and political decentralization on economic growth, income inequality and regional disparities may therefore benefit from greater emphasis on decentralization's role in shaping specialization patterns in an urban system.

References

[1] Abdel-Rahman, H., (1990). Agglomeration economies, types and sizes of cities. *Journal of Urban Economics*, 27(1), 25–45.

[2] Abdel-Rahman, H. and Fujita, M., (1993). Specialization and diversification in a system of cities. *Journal of Urban Economics*, 33(2), 189–222.

[3] Abdel-Rahman, H., (1996). When do cities specialize in production? *Regional Science and Urban Economics*, 26(1):1–22.

[4] Anas, A. and Xiong, K., (2005). The formation and growth of specialized cities: efficiency without developers or malthusian traps. *Regional Science and Urban Economics*, 35(4):445– 470.

[5] Bardhan, P., (2002). Decentralization of Governance and Development. *Journal of Economic Perspectives*, 16(4), 185-205.

[6] Black, D. and Henderson, V., (2003). Urban evolution in the USA. *Journal of Economic Geography*, 3(4), 343-372.

[7] Breton, A. (1996). *Competitive governments: An economic theory of politics and public finance*. Cambridge MA: Cambridge University Press.

[8] Brezzi, M., Piacentini, M., Rosina, K., and Sanchez-Serra, D., (2012). Redefining urban areas in OECD countries. Redefining" urban": a new way to measure metropolitan areas, *OECD*, Paris, 19-58.

[9] Camagni, R., Capello, R., Caragliu, A. (2015) The rise of second-rank cities: What role for agglomeration economies? *European Planning Studies*, 23, 1069–1089

[10] Carrion-i-Silvestre, J. L., Espasa, M. and Mora, T., (2008). Fiscal decentralization and economic growth in Spain. *Public Finance Review*, 36(2), 194-218.

[11] Christaller, W., (1933). Central places in southern Germany. Translated Edition. *Prentice Hall*.

[12] Cilasun, S.M., and Günalp, B., (2012). An empirical analysis of the average plant start-up Size in Turkish Manufacturing Industries. *Review of Industrial Organization*, 40(4), 273-289.

[13] Combes, P.P. and Lafourcade, M., (2005). Transport costs: measures, determinants, and regional policy implications for France. *Journal of Economic Geography*, 5(3), 319-349.

[14] Desmet, K. and Rossi-Hansberg, E., (2014). Spatial development. *American Economic Review*, 104(4), 1211-1243.

[15] Díaz-Lanchas, J., Llano-Verduras, C. and Zofío, J.L., (2019). A trade hierarchy of cities based on transport cost thresholds. *JRC Working Papers on Territorial Modelling and Analysis*, Joint Research Centre, European Commission. 2019/02.

[16] Duranton, G., and Puga, D., (2000). Diversity and specialization in cities: why, where and when does it matter? *Urban studies*, 37(3), 533–555.

[17] Duranton, G. and Puga, D., (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91(5), 1454–1477.

[18] Duranton, G. and Puga, D. (2014). The growth of cities. In *Handbook of economic growth*, Vol. 2, Elsevier, pp. 781-853.

[19] Duranton, G. and Turner, M.A., (2012). Urban growth and transportation. *Review of*

Economic Studies, 79(4), 1407-1440.

[20] Ellison, G. and Glaeser, E.L., and Kerr, W.R., (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100, 1195–1213.

[21] Ezcurra, R. and Rodríguez-Pose, A. (2013) Political Decentralization, Economic Growth and Regional Disparities in the OECD. *Regional Studies*, 47(3), 388-401.

[22] Frick, S.A. and Rodríguez-Pose, A. (2018). Big or Small Cities? On city size and economic growth. *Growth and Change*, 49, 4–32.

[23] Fujita, M., Krugman, P. and Venables, A., (1999). The spatial economy: Cities, regions, and international trade. *MIT Press*.

[31] Glaeser, E.L., Ponzetto, G. and Tobio, K., (2014). Cities, skills and regional change. *Regional Studies*, 48(1), 7–43.

[32] González-Val, R., Lanaspa, L. and Sanz-Gracia, F., (2014). New evidence on Gibrat's law for cities. *Urban Studies*, 51(1), 93-115.

[33] González-Val, R., Ramos, A., Sanz-Gracia, F. and Vera-Cabello, M., (2015). Size distributions for all cities: Which one is best? *Papers in Regional Science*, 94(1), 177-196.

[34] Greene, W., (2012). Econometric Analysis. *Pearson Education*, seventh edition.

[35] Head, K. and Mayer, T., (2010). Gravity, market potential and economic development. *Journal of Economic Geography*, 11(2): 281-294.

[36] Helsley, R.W. and Strange, W., (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, 122(5), 1064–1093.

[37] Henderson, V., (1974). The sizes and types of cities. *American Economic Review*, 64(4), 640–656.

[38] Henderson, V., (2003). Marshallian scale economies. *Journal of Urban Economics*, 53:1–28.

[39] Henderson, V. and Abdel-Rahman, H., (1991). Urban diversity and fiscal decentralization. *Regional Science and Urban Economics*, 21, 491-509.

[40] Henderson, V. and Becker, R., (2000). Political economy of city sizes and formation. *Journal of Urban Economics*, 48(3), 453-484.

[41] Hooghe, L., Marks, G., Schakel, A., Chapman-Osterkatz, S., Niedzwiecki, S., and Shair-Rosenfield, S., (2016), Measuring regional authority. *Volume I: A postfunctionalist theory of governance*. Oxford University Press.

[42] İrepoğlu Carreras, Y., (2016). Fiscal decentralization and inequality: the case of Spain. *Regional Studies, Regional Science*, 3(1), 295-302.

[43] Lago-Peñas, S., Fernández-Leiceaga, X. and Vaquero-García, A., (2017). Spanish fiscal decentralization: A successful (but still unfinished) process. *Environment and Planning C: Politics and Space*, 35(8), 1509-1525.

[44] Lessmann, C., (2012) Regional inequality and decentralization: an empirical analysis. *Environment and Planning A*, 44, 1363-1388

[45] Lin, Y., (2017). Travel costs and urban specialization patterns: Evidence from China's high speed railway system. *Journal of Urban Economics*, 98, 98-123.

[46] McCann, P., and Acs, Z., (2011). Globalization: countries, cities and multinationals.

Regional Studies, 45(1), 17-32.

[47] Martinez-Vazquez, J., Lago-Peñas, S. and Sacchi, A., (2017). The Impact of Fiscal Decentralization: A Survey. *Journal of Economic Surveys*, 31(4), 1095-1129.

[48] Meijers, E.J. Burger, M.J., Hoogerbrugge, M.M. (2016). Borrowing size in networks of cities: City size, network connectivity and metropolitan functions in Europe. *Papers in Regional Science*, 95, 181-198.

[49] Moreno, L., (2002). Decentralization in Spain, *Regional Studies*, 36(4), 399-408.

[50] Muth, R.F., (1969). Cities and housing: The spatial pattern of urban residential land use. *The University of Chicago Press*.

[51] Oates, W. (1972). *Fiscal Federalism*. New York: Harcourt Brace Jovanovich.

[52] Prud'homme, R. (1995). The dangers of decentralization. *The World Bank Research Observer*, 10, 201-220.

[53] Qian, Y. and Weingast, B.R. (1997). Federalism as a commitment to preserving market incentives. *Journal of Economic Perspectives*, 11, 83-92.

[54] Redding, S., and Rossi-Hansberg, E., (2017). Quantitative spatial economics. *Annual Review of Economics*, 9(1).

[55] Rodríguez-Pose, A. and Ezcurra, R. (2010). Does decentralization matter for regional disparities? A cross-country analysis. *Journal of Economic Geography*, 10, 619–644.

[56] Rodríguez-Pose, A. and Ezcurra, R. (2011). Is fiscal decentralization harmful for economic growth? Evidence from the OECD countries. *Journal of Economic Geography*, 11, 619-643.

[57] Rodríguez-Pose, A. and Gill, N. (2004). Is there a global link between regional disparities and devolution? *Environment and Planning A* ,36, 2097-2117.

[58] Tabuchi, T., and Thisse, J.-F., (2011). A new economic geography model of central places. *Journal of Urban Economics*, 6(4):240–252.

[59] Tiebout, C.M. (1956). A pure theory of local expenditures. *Journal of Political Economy*, 64, 416-24.

[60] Viladecans-Marsal, E., (2004). Agglomeration economies and industrial location: city-level evidence. *Journal of Economic Geography*, 4(5):565–582.

[61] Zofío, J.L., Condeço-Melhorado, A.M., Maroto-Sánchez, A., and Gutiérrez, J. (2014). Generalized transport costs and index numbers: A geographical analysis of economic and infrastructure fundamentals. *Transportation Research Part A: Policy and Practice*, 67, 141-157.

Supplemental Appendix

Does decentralization of governance promote urban diversity? Evidence from Spain.

Abstract

This appendix contains materials supplemental to the manuscript. This document contains the results of a Robustness Analysis (Part A) and additional technical details regarding the Data Representativeness (Part B).

Part A: Robustness Analysis.

Table A.1: Probit estimations – All models with spatial fixed effects

	$Pr(S_i=1\|X)$			$Pr(D_i=1\|X)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$ln(MP\_internal_i)$	-0.137*	-0.139*	-0.014	0.153*	0.076	0.079
	(0.072)	(0.075)	(0.086)	(0.088)	(0.092)	(0.095)
$ln(MP\_External_i)$	0.464	0.475	-0.093	-2.132***	-2.267***	-2.324***
	(0.404)	(0.401)	(0.446)	(0.489)	(0.502)	(0.504)
Reg_Gov_i	---	0.011	---	---	0.641***	---
		0.011			0.641***	
$Reg\_Gov*ln(MP\_int_i)$	---	---	-0.623***	---	---	-0.130
			(0.192)			(0.185)
$Reg\_Gov*ln(MP\_ext_i)$	---	---	0.497***	---	---	0.160
			(0.149)			(0.141)
$ln(sh\_HighEduc_i)$	-1.183***	-1.184***	-1.301***	3.193***	2.990***	2.997***
	(0.253)	(0.258)	(0.268)	(0.373)	(0.380)	(0.380)
$Ln(ratio\_MS_i)$	0.225**	0.224**	0.254***	0.218**	0.199**	0.198**
	(0.089)	(0.089)	(0.090)	(0.090)	(0.089)	(0.089)
$Ln(MES_i)$	0.298**	0.296**	0.311**	-0.545***	-0.601***	-0.603***
	(0.133)	(0.133)	(0.134)	(0.137)	(0.138)	(0.138)
$Ln(RZI1995_i)$	0.929***	0.925***	0.837***	---	---	---
	(0.101)	(0.103)	(0.104)			
$Ln(RDI1995_i)$	---	---	---	1.182***	1.296***	1.313***
				(0.243)	(0.247)	(0.249)
ρ	-0.258***	-0.270***	-0.277***	-0.258***	-0.270***	-0.277***
	(0.074)	(0.075)	(0.076)	(0.074)	(0.075)	(0.076)
N	897	897	897	897	897	897
Year Fixed-Effects	YES	YES	YES	YES	YES	YES
Spatial Fixed-Effects	YES	YES	YES	YES	YES	YES

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i* refers to city *i* and 80 indicates variables in 1980. Columns (1)-(3) indicate the probability for a city to become specialized, whereas columns (4)-(6) do the same for the probability of diversification.

Table A.2: Probit estimations – All models with spatial fixed effects; 1980 models

	$Pr(S_i=1\|X)$		$Pr(D_i=1\|X)$	
	(7)	(8)	(9)	(10)
$ln(MP\_Internal80_i)$	-0.086	0.028	-0.073	-0.089
	(0.062)	(0.072)	(0.074)	(0.077)
$ln(MP\_External80_i)$	0.583*	0.073	-1.646***	-1.580***
	(0.348)	(0.377)	(0.402)	(0.411)
Reg_Gov_i	-0.033	---	0.678***	---
	(0.168)		(0.174)	
$Reg\_Gov*ln(MP\_int80_i)$		-0.760***		0.148
	---	(0.201)	---	(0.17)
$Reg\_Gov*ln(MP\_ext80_i)$		0.507***		-0.045
	---	(0.13)	---	(0.108)
$ln(sh\_HighEduc_i)$	-1.212***	-1.342***	3.003***	3.009***
	(0.256)	(0.269)	(0.366)	(0.366)
$Ln(ratio\_MS_i)$	0.218**	0.251***	0.220**	0.216**
	(0.090)	(0.092)	(0.089)	(0.089)
$Ln(MES_i)$	0.282**	0.304**	-0.570***	-0.573***
	(0.131)	(0.133)	(0.135)	(0.135)
$Ln(RZI1995_i)$	0.932***	0.832***		
	(0.102)	(0.101)	---	---
$Ln(RDI1995_i)$			1.411***	1.390***
	---	---	(0.244)	(0.244)
ρ	-0.270***	-0.264***	-0.270***	-0.264***
	(0.075)	(0.076)	(0.075)	(0.076)
N	897	897	897	897
Year Fixed-Effects	YES	YES	YES	YES
Spatial Fixed-Effects	YES	YES	YES	YES

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i* refers to city *i* and 80 indicates variables in 1980. Columns (7) and (8) indicate the probability for a city to become specialized, whereas columns (9) and (10) do the same for the probability of diversification.

Table A.3: Probit estimations – Baseline Models with Population and average GTC

	$Pr(S_i=1\|X)$			$Pr(D_i=1\|X)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(Pop_i)$	-0.173***	-0.205***	-0.027	0.611***	0.582***	0.622***
	(0.066)	(0.072)	(0.091)	(0.088)	(0.093)	(0.104)
$\ln(GTC_i)$	0.944***	0.889***	0.807***	0.548***	0.638***	0.594***
	(0.196)	(0.205)	(0.211)	(0.192)	(0.198)	(0.198)
Reg_Gov_i	---	0.128	---	---	0.247*	---
		(0.143)			(0.142)	
$Ln(RAI_i)$	---	-3.418***	-3.417***	---	1.902***	1.748**
		(0.737)	(0.683)		(0.683)	(0.718)
$Reg\_Gov*\ln(Pop_i)$	---	---	-0.462***	---	---	-0.192
			(0.143)			(0.154)
$Reg\_Gov*\ln(RAI_i)$	---	---	1.909***	---	---	0.854
			(0.581)			(0.622)
$\ln(sh\_HighEduc_i)$	-0.868***	-1.018***	-1.300***	2.013***	1.960***	1.883***
	(0.228)	(0.241)	(0.261)	(0.259)	(0.275)	(0.278)
$Ln(ratio\_MS_i)$	0.285***	0.285***	0.302***	-0.090	-0.099	-0.103*
	(0.071)	(0.071)	(0.072)	(0.062)	(0.062)	(0.062)
$Ln(MES_i)$	0.419***	0.337***	0.344***	-0.739***	-0.724***	-0.718***
	(0.127)	(0.123)	(0.124)	(0.137)	(0.139)	(0.140)
$Ln(RZI1995_i)$	0.837***	0.866***	0.834***	---	---	---
	(0.088)	(0.092)	(0.093)			
$Ln(RDI1995_i)$	---	---	---	1.121***	1.190***	1.174***
				(0.216)	(0.221)	(0.220)
ρ	-0.384***	-0.375***	-0.379***	-0.384***	-0.375***	-0.379***
	(0.070)	(0.071)	(0.071)	(0.070)	(0.071)	(0.071)
N	897	897	897	897	897	897
Year Fixed-Effects	YES	YES	YES	YES	YES	YES
Spatial Fixed-Effects	NO	NO	NO	NO	NO	NO

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i* refers to city *i*. Columns (1)-(3) indicate the probability for a city to become specialized, whereas columns (4)-(6) do the same for the probability of diversification.

Table A.4: Probit estimations – Baseline Models with Population and average GTC, 1980 models

	$Pr(S_i=1\|X)$		$Pr(D_i=1\|X)$	
	(7)	(8)	(9)	(10)
$ln(Pop80_i)$	-0.175**	-0.090	0.554***	0.607***
	(0.068)	(0.079)	(0.090)	(0.097)
$ln(GTC80_i)$	1.191***	1.181***	0.429*	0.465*
	(0.233)	(0.243)	(0.254)	(0.250)
Reg_Gov_i	-0.064	---	0.414***	---
	(0.147)		(0.144)	
$Ln(RAI80_i)$	-1.166**	-1.414***	0.879	0.508
	(0.494)	(0.503)	(0.553)	(0.561)
$Reg\_Gov*ln(Pop80_i)$	---	-0.307**	---	-0.279**
		(0.138)		(0.133)
$Reg\_Gov*ln(RAI80_i)$	---	1.376**	---	1.415**
		(0.632)		(0.601)
$ln(sh\_HighEduc_i)$	-0.770***	-0.917***	1.748***	1.693***
	(0.242)	(0.255)	(0.272)	(0.270)
$Ln(ratio\_MS_i)$	0.362***	0.375***	-0.181***	-0.170**
	(0.072)	(0.072)	(0.067)	(0.066)
$Ln(MES_i)$	0.357***	0.395***	-0.747***	-0.713***
	(0.128)	(0.132)	(0.143)	(0.143)
$Ln(RZI1995_i)$	0.846***	0.804***	---	---
	(0.096)	(0.096)		
$Ln(RDI1995_i)$	---	---	1.089***	1.084***
			(0.218)	(0.218)
ρ	-0.376***	-0.393***	-0.376***	-0.393***
	(0.070)	(0.072)	(0.070)	(0.072)
N	897	897	897	897
Year Fixed-Effects	YES	YES	YES	YES
Spatial Fixed-Effects	NO	NO	NO	NO

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i* refers to city *i* and 80 indicates variables in 1980. Columns (7) and (8) indicate the probability for a city to become specialized, whereas columns (9) and (10) do the same for the probability of diversification.

Table A.5: Average marginal effects –
Baseline Models with Population and average GTC

	Baseline Model 1				Baseline Model 2			
	Non-typified	Specialized	Diversified	Co-agglomerated	Non-typified	Specialized	Diversified	Co-agglomerated
	P(0,0)	P(1,0)	P(0,1)	P(1,1)	P(0,0)	P(1,0)	P(0,1)	P(1,1)
$\ln(Pop_i)$	-0.071***	-0.161***	0.153***	0.079***	-0.116***	-0.132***	0.127***	0.121***
	(0.021)	(0.025)	(0.026)	(0.023)	(0.026)	(0.030)	(0.031)	(0.025)
$\ln(GTC_i)$	-0.298***	0.043	-0.056	0.311***	-0.279***	0.042	-0.043	0.280***
	(0.052)	(0.064)	(0.062)	(0.053)	(0.053)	(0.065)	(0.062)	(0.054)
Reg_Gov_i	-0.068**	-0.030	0.017	0.081**	---	---	---	---
	(0.028)	(0.046)	(0.046)	(0.037)				
$Ln(RAI_i)$	0.311*	-1.070***	1.051***	-0.292*	0.346**	-1.043***	1.016***	-0.319*
	(0.159)	(0.238)	(0.232)	(0.164)	(0.165)	(0.231)	(0.226)	(0.164)
$Reg\_Gov*\ln(Pop_i)$	---	---	---	---	0.131***	-0.054	0.053	-0.130***
					(0.041)	(0.045)	(0.044)	(0.039)
$Reg\_Gov*\ln(RAI_i)$	---	---	---	---	-0.552***	0.212	-0.209	0.550***
					(0.166)	(0.183)	(0.178)	(0.156)
$\ln(sh\_HighEduc_i)$	-0.174**	-0.608***	0.580***	0.202***	-0.107	-0.644***	0.625***	0.126*
	(0.069)	(0.082)	(0.075)	(0.071)	(0.074)	(0.084)	(0.078)	(0.074)
$Ln(ratio\_MS_i)$	-0.037**	0.077***	-0.076***	0.037*	-0.041**	0.082***	-0.080***	0.039**
	(0.018)	(0.020)	(0.019)	(0.019)	(0.018)	(0.020)	(0.019)	(0.018)
$Ln(MES_i)$	0.072**	0.217***	-0.206***	-0.082**	0.071**	0.215***	-0.209***	-0.078**
	(0.032)	(0.043)	(0.040)	(0.033)	(0.033)	(0.043)	(0.040)	(0.033)
$Ln(RZI1995_i)$	-0.171***	0.171***	-0.174***	0.174***	-0.168***	0.168***	-0.164***	0.164***
	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
$Ln(RDI1995_i)$	-0.228***	-0.247***	0.228***	0.247***	-0.230***	-0.238***	0.230***	0.238***
	(0.042)	(0.048)	(0.042)	(0.048)	(0.043)	(0.048)	(0.043)	(0.048)
N	897	897	897	897	897	897	897	897
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Spatial FE	NO	NO	NO	NO	NO	NO	NO	NO

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1. Note: *i* refers to city *i* and 80 indicates variables in 1980.

Part B: Data Representativeness

The SABI database is compared with two alternative databases from the Spanish Statistical Institute and the Spanish Ministry of Labor. These two official databases do not provide information at the urban level, but at a higher regional level (NUTS-3 level). Because of this restriction we use only the official databases to assess SABI's reliability and check its representativeness, aggregating its data by the following dimensions: the national level by year and size of the firm; the sectorial level (38 NACE Classification), and the regional level (Provinces, NUTS-3 level). As can be observed, its representativeness increases as it includes 73% of the official labor force by 2007. It very accurately covers employment at medium and big firms, as official registration is compulsory for these firms but not for small firms (less than 50 employees), and employment in the main sectors of the Spanish economy. Also, the main provinces are well captured even with a slight bias in favor of the richest Spanish provinces (Madrid, Barcelona and Bizkaia), as firms' headquarters are located mainly in these regions.

Table B.1: Ratios used for representativeness of SABI Database

Dimension	Simple Ratio	Ratio of Ratios
Size of Firm (s)	$Ratio_t^s = \frac{\sum_i x_t^s SABI}{\sum_i x_t^s Official}$	$(Ratio - Ratio)_t^s = \frac{\frac{\sum_i x_t^s SABI}{\sum_i x_t SABI}}{\frac{\sum_i x_t^s Official}{\sum_i x_t Official}}$
Industry (k)	$Ratio_k = \frac{\sum_t x_t^k SABI}{\sum_t x_t^k Official}$	$(Ratio - Ratio)_k = \frac{\frac{\sum_t x_t^k SABI}{\sum_t x_t SABI}}{\frac{\sum_t x_t^k Official}{\sum_t x_t Official}}$
Provinces (r)	$Ratio_r = \frac{\sum_t x_t^r SABI}{\sum_t x_t^r Official}$	$(Ratio - Ratio)_r = \frac{\frac{\sum_t x_t^r SABI}{\sum_t x_t SABI}}{\frac{\sum_t x_t^r Official}{\sum_t x_t Official}}$

Table B.2: SABI and Official database at the firm level (by year)

Year	Workers SABI	Workers Official Data	Ratio SABI/Official Database (1) Total	(2) Small	(3) Medium	(4) Big	Ratio of Ratios (5) Small	(6) Medium	(7) Big
1995	1,357,964	6,006,100	.22	.17	.26	.30	.74	1.14	1.33
1996	3,660,271	7,886,500	.46	.28	.59	.88	.59	1.25	1.87
1997	3,860,721	8,364,800	.46	.30	.6	.75	.65	1.29	1.61
1998	4,271,373	8,988,200	.47	.30	.57	.82	.64	1.19	1.74
1999	4,401,396	9,659,100	.45	.29	.54	.80	.63	1.18	1.75
2000	4,193,074	10,234,600	.41	.23	.50	.76	.56	1.24	1.85
2001	4,367,842	10,652,300	.41	.22	.49	.79	.53	1.19	1.93
2002	5,448,871	10,977,300	.50	.30	.58	.88	.61	1.16	1.78
2003	6,929,670	11,278,000	.61	.45	.66	.95	.74	1.09	1.54
2004	8,081,766	11,772,100	.69	.56	.68	.99	.81	.99	1.44
2005	9,786,354	12,499,500	.78	.63	.69	1.23	.80	.88	1.57
2006	9,507,272	13,100,300	.73	.64	.68	.96	.88	.94	1.33
2007	9,928,389	13,557,300	.73	.67	.70	.88	.92	.96	1.2

Columns (1)-(4): Ratio of the number of workers on SABI over the Official Database by type of firm. Columns (5)-(7): Ratio of Ratio of SABI over the Official Database by type of firm. Small Firm: less than 50 employees; Medium Firm: 50-250 employees; Big Firm: more than 250 employees.
Source: SABI Database and Spanish Ministry of Employment.

Table B.3: SABI and Official database at the industry level (all years, 1995-2007)

NACE	SABI/Official	Ratio of Ratios
01-03	.18	.42
05-09	.18	.43
10-12	.04	.09
13-15	.70	1.78
16-18	.42	.96
19	3.12	7.15
20	.42	.96
21	.97	2.23
22-23	.30	.70
24-25	.47	1.10
26	1.76	4.10
27	.59	1.35
28	1.24	2.85
29-30	.36	.85
31-33	.85	1.96
35	.64	1.46
36-39	.75	1.72
41-43	.03	.07
45-47	.32	.73
49-53	1.97	4.52
55-56	.32	.75
58-60	1.5	3.46
61	.11	.25
62-63	.85	1.96
64-66	.32	.74
69-71	.50	1.14
72	2.75	6.33
73-75	4.47	10.27
77-82	.07	.15
85	.09	.20
90-93	.60	1.36
94-96	.00	.00
97-98	.00	.00

SABI/Official: Ratio of the number of workers on SABI over the Official Database by Industry.
Ratio of Ratio: Ratio of Ratio of SABI over the Official Database by Industry.
Source: SABI Database and Spanish National Statistical Institute.

Table B.4: SABI and Official database at the province level (total all years, 1995-2007)

Province (r)	Ratio SABI/Official	Ratio of Ratios	Province (r)	Ratio SABI/Official	Ratio of Ratios
Álava	.32	1	—	—	—
Albacete	.21	.66	León	.15	.46
Alicante	.25	.79	Lleida	.21	.65
Almería	.18	.57	La Rioja	.23	.71
Ávila	.10	.32	Lugo	.14	.44
Badajoz	.14	.43	Madrid	.58	1.83
Barcelona	.41	1.3	Málaga	.19	.59
Burgos	.25	.77	Murcia	.27	.86
Cáceres	.13	.42	Navarra	.3	.95
Cádiz	.16	.49	Ourense	.19	.60
Castellón	.35	1.1	Asturias	.26	.80
Ciudad Real	.17	.52	Palencia	.14	.44
Córdoba	.18	.57	Pontevedra	.28	.89
La Coruña	.26	.81	Salamanca	.15	.47
Cuenca	.12	.37	Cantabria	.18	.57
Girona	.29	.90	Segovia	.14	.43
Granada	.15	.48	Sevilla	.21	.66
Guadalajara	.15	.49	Soria	.15	.46
Guipuzkoa	.27	.86	Tarragona	.23	.73
Huelva	.18	.55	Teruel	.17	.54
Huesca	.17	.52	Toledo	.19	.60
Jaén	.12	.52	Valencia	.3	.96
Valladolid	.3	.93	Bizkaia	.39	1.21
Zamora	.12	.37	Zaragoza	.31	.98

SABI/Official: Ratio of the number of workers on SABI over the Official Database by Province.
Ratio of Ratio: Ratio of Ratio of SABI over the Official Database by Industry.
Source: SABI Database and Spanish National Statistical Institute.

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.