GEN - Governance and Economics research Network

GEN Working Paper B 2014 – 2
http://webs.uvigo.es/infogen

THE CAUSES OF URBAN SPRAWL IN SPANISH URBAN AREAS: A SPATIAL APPROACH

Miguel Gómez-Antonio*
Miriam Hortas-Rico**
Linna Li***

** Complutense University of Madrid*
***Complutense University of Madrid and GEN*
****California State University, Long Beach*

Grupo de investigación GEN
Universidade de Vigo

The causes of urban sprawl in Spanish urban areas: a spatial approach*

Miguel Gómez-Antonio (Complutense University of Madrid)
Miriam Hortas-Rico (Complutense University of Madrid & Governance and Economic Research Network)
Linna Li (California State University, Long Beach)

Abstract. This paper explores the role of interjurisdictional competition among local governments in fostering urban sprawl. The structure of local public finance along with housing and land-use policies make land a valuable commodity the supply of which is monopolised by local governments. This situation creates economic incentives for local governments (in terms of higher income and tax revenues) to influence development patterns and even engage in strategic interaction with neighbouring jurisdictions to compete for the creation of new residential areas. We empirically assess the presence of local spatial interaction in urban sprawl in Spanish urban areas. Thanks to the recent availability of a novel data set based on remotely sensed data from aerial photography and satellite imaging, we have been able to study urban development patterns across the country with unprecedented detail. We make use of GIS techniques to calculate a sprawl measure as the dependent variable and compile a database of independent variables on land use and topographic features, complemented with additional information on weather conditions, social, demographic, political and economic variables, which are then used in a spatial regression model. The results confirm our main hypothesis: there exists spatial interaction in the levels of sprawl between neighbouring municipalities, suggesting that local governments do indeed compete for the creation of new suburban settlement developments, hence promoting excessive urban sprawl.

JEL codes: C21, H7, R14.
Keywords: urban sprawl, spatial econometrics, strategic interaction, local governments

* We would like to thank Jan Brueckner and Frank Davenport for valuable comments on the paper. We also thank several participants at the 11th PEARL Seminar for helpful comments and suggestions.
Corresponding author: Miriam Hortas-Rico (miriamhortas@ccee.ucm.es)

1. Introduction

In recent years, urban sprawl has become a matter of concern in the process of urban development in European countries (EEA, 2006, 2010). Scattered, leapfrogging and low-density outward expansion of cities, associated more with US cities at the beginning of the 20^{th} century, has now become part of the European landscape. During the last 20 years, the rates of conversion to residential land use have far exceeded population growth rates in the continent (20% versus 6%). As a result, the amount of urbanised land consumed per person during this period has more than doubled, leading to the formation of both new edge cities around traditional urban centres and scattered residential developments located on the urban fringe. Lower urban densities, high losses of non-urban land covers, depopulation of the metropolitan inner core, increasing importance of single housing and the expansion of transportation infrastructures confirm the generalisation of the dispersed urban model (Catalán et al, 2008)[1].

Proponents of sprawl consider it an efficient outcome of the development process as it fulfils residents' preferences for large and affordable housing located near open space (Gordon and Richardson, 2000), while helping reduce the impact of higher densities of compact cities, especially congestion. Opponents of sprawl, on the other hand, suggest that it results from several market failures leading to the excessive spatial expansion of cities and creating many of the challenges that cities face today (Brueckner, 2001). Urban sprawl encourages excessive use of cars, increasing commuting time and traffic congestion while contributing to air pollution, global warming and loss of farmland and open space (Sierra Club, 1999; Downs, 1999; Brueckner 2001; Glaeser and Khan, 2004). Claims are also made that sprawl reduces social interaction and causes apathy, antisocial behaviour and obesity. At the same time, sprawl weakens agglomeration economies and economies of scale in public service provision (Carruthers and Ulfarsson, 2003; Hortas-Rico and Solé-Ollé, 2010), contributing to socioeconomic segregation, income inequality and polarisation, fostering urban decay in core areas (Mills and Price, 1984; Mieskowski and Mills, 1993; Brueckner and Helsley, 2012). Changes in urban forms and development patterns are essential for understanding the role of cities as engines of growth. From an economic point of view, efficient land-use patterns trade off social, economic and environmental costs against the benefits of urban sprawl. If the benefits of sprawl are offset by its costs and consumers face a welfare loss, then policies must be implemented to curb sprawl and mitigate its negative consequences. Nonetheless, to formulate efficient environmental and land-use policies and to

[1] Since the mid-1950s, European cities have expanded on average by 78% whereas the population has grown by only 33%. A major consequence of this trend is that European cities have become much less compact. More than 90% of all residential areas built after the mid-1950s were low density areas, with less than 80% of the land covered by buildings, roads and other structures (EEA, 2006).

evaluate their effectiveness, we must gain a better understanding of the drivers of this urban development pattern (Brueckner, 2001; Wassmer, 2008; Wu, 2006).

What are the driving forces of this process of suburbanisation? Urban economic theory tells us that the spatial growth of cities is a result of three fundamental forces, namely population growth, rising incomes and lower transportation costs derived from important investments in transportation infrastructures. Individual preferences and the Tiebout sorting, market and public sector failures and certain political determinants are also considered as causes of urban sprawl. The pioneering work by Brueckner and Fansler (1982) and more recently McGrath (2005), Wassmer (2008) and Burchfield et al (2006), among others, provides empirical evidence in this regard. Among the aforementioned causes, local government behaviour is perhaps of particular interest as it might heavily distort land-use decision-making, fostering urban sprawl. In this setting, the structure of public finance and local revenue reliance becomes crucial (Slack, 2002). Urban sprawl has been traditionally blamed for increasing levels of expenditure, as it may raise the provision costs of certain local public goods and requires greater investment in extending basic infrastructures for new urban development located on the urban fringe (Carruthers and Ulfarsson, 2003; Hortas-Rico and Solé-Ollé, 2010). Despite its costs, short-sighted local governments may see sprawl as a potential source of finance, in terms of immediately relevant building-associated revenues (Hortas-Rico, 2013). A system to finance municipalities that relies heavily on revenues related to development could encourage municipalities to inefficiently use land use conversion as a funding tool, as land-based financing of infrastructure investment has the biggest payoff where there is rapid urban growth. The design of the local finance system might provide incentives to local governments to pursue excessive development, as they can raise revenues by selling land and promoting urban growth, as well as to compete with neighbouring jurisdictions to attract residents to their administrative boundaries, especially in a setting where local governments have a limited set of policy instruments to raise revenues. This strategic interaction among neighbouring jurisdictions can, however, generate an inefficient allocation of resources, leading to a non-optimal level of urban sprawl. Heubeck (2009) develops a theoretical model that shows how competition for the creation of new residential areas can generate early development and may lead to inefficient growth of cities. In imperfectly competitive urban land markets, local policy decisions to develop land at a particular location may be affected by the decisions of other nearby local governments. These policy choice interdependencies create feedback effects across space that are endogenous in a cross-sectional model of decision-making that must be considered in characterising equilibrium (Brueckner, 1998). Previous literature on the causes of urban sprawl has, however, considered municipalities as isolated entities. Nonetheless, strategic interaction among local governments is a key element in explaining the urban expansion of cities and, as such, these spatial effects must be taken into account. After analysing the relationship between business tax competition and commercial land use, Buettner and Ebertz (2012) conclude that their results suggest that other margins of land-use might be affected by tax competition in another fiscal environment. In particular, it seems that tax competition for mobile residents provides an incentive to enhance residential land use. In this paper we focus on the role played by institutions and their strategic behaviour regarding land-use development patterns and competition for mobile middle- and upper-income residents as an additional cause of urban sprawl.

The paper makes several contributions to the literature. First, building on the theoretical grounds of traditional urban economics theory we expand previous research on the causes of urban sprawl, so as to account for the spatial interaction in the levels of suburban development. We do so by relying on spatial econometrics techniques, which allow us to account for not only the strategic interaction among local governments but also other spatial-based interactions[2]. In particular, we estimate a spatial lag model where the dependent variable is a sprawl measure. Note, however, that this approach is different from the usual application where the spatial lag framework involves a policy variable[3]. In our particular case, both the policy variable (land-use conversion) and the outcome variable (urban sprawl) are equivalent due to the following facts. First, local authorities are responsible for land-use decision-making. Second, most of the Spanish construction activity during the nineties took the form of dispersed suburban development[4].

Secondly, the paper develops an accurate measure of urban sprawl as the dependent variable in our spatial regression. Despite existing research on the causes of urban sprawl, the spatial dimension of urban development (i.e. whether it is compact or scattered) and the factors influencing the variation of this urban development pattern across space have not been subject to much research to date. We follow Burchfield et al (2006), and define sprawl as the percentage of open space in the square kilometre surrounding an average residential development. Thanks to the recent availability of a novel data set based on remotely sensed data from aerial photography and satellite imaging, we have been able to study urban development patterns across Spanish urban areas with unprecedented detail. We make use of GIS techniques to calculate the urban sprawl variable and compile a database of variables on land use and topographic features, complemented with additional information on weather conditions, social, demographic, political and economic variables and used these variables to estimate an econometric land use model that takes into account spatial dependence among municipality-level urban sprawl.

[2] Spatial dependence can arise from interactions among spatial units, either because nearby jurisdictions directly affect each other (spatial lag dependence), because they are affected by the same unobserved factors (spatial error dependence) or because some of the variables used in the empirical models might be measured with error, as the scale at which they are measured may not match the scale of the underlying spatial process (spatial heterogeneity).

[3] Existing studies on land-use markets have focused their attention on strategic interactions among local governments regarding property tax or choice of urban growth control measures. See Wilson (1999) for a survey of the literature on tax competition, and Brueckner (2003) for a literature review on strategic interaction among local governments.

[4] More than 80% of new urban fabric was discontinuous (OSE, 2006).

Finally, this research represents a novel application to a European country and thus enables comparison with previous studies of U.S. cities. Although sprawling development patterns in the U.S. and Europe share some common trends they both have their own peculiarities. Sprawl is at least partially the result of government policies that encouraged it by promoting infrastructure improvements that considerably extended highway networks, opening up suburban real estate for development. As noted by Nechyba and Walsh (2004), Europe is far less car-dependent than the U.S., partly because of greater investments in public transportation. Compared to most American urban cores, their European counterparts have traditionally been more compact and, thus, reasonably healthy. In addition, the urbanisation process in many European countries is characterised by its duality, combining a compact city model with a more dispersed one. The complexity and diversity of urban form throughout Europe (that sometimes does not comply with national borders and even inside each country) must be accounted for when designing urban policies (see, e.g., Schwarz, 2010). In spite of everything, U.S. urban land use and development patterns have inspired a lot of research in the last few years, whereas research in European countries has been quite modest. This lack of empirical studies is obviously not due to the absence of sprawl in Europe but more probably due to data availability constraints. Recent availability of U.S. comparable satellite imaging allows us to overcome this limitation. Thus, the paper contributes to the existing literature on land-use development patterns by providing empirical evidence of the causes of sprawl for a European country.

The rest of the paper is organised as follows. In the next section we define the concept of urban sprawl and provide a measure of the phenomenon. A literature review on the causes of urban sprawl is presented in the third section. The fourth section discusses the econometric strategy, while the definitions of the variables, the unit of analysis and data sources are presented in the fifth section. In the sixth section we report the main results. Finally, we present our conclusions in the last section.

2. Defining and measuring urban sprawl

The definition of urban sprawl. Although urban sprawl has become a matter of great concern, there is no universally accepted definition of the phenomenon (see Torrens, 2008 for a review). There are, however, several common characteristics pervading the literature that can help us understand and even measure its occurrence (Brody, 2013). In this regard, urban sprawl can be considered to be a low-density, scattered, discontinuous auto-dependent urban development pattern, taking place on the edges of urban centres, often leapfrogging away from current denser city centres, to transform open, undeveloped land, into single-family residential subdivisions.

Sprawl is the leading edge of urban growth without systematic large-scale or regional public land-use planning and implies little planning control of land subdivision[5].

How do we measure urban sprawl? A review of the literature also shows that there is no consensus on which is the most suitable variable to capture urban sprawl[6]. So far, density has been the most widely used indicator of sprawl because of its simplicity and the difficulty of obtaining data about alternative measures. Density is conceptually and practically straightforward, but provides a simplified assessment, as it is only part of the picture and it sometimes provides an ambiguous image of the urban form, without telling us anything about how residential uses are distributed (Carruthers and Ulfarsson, 2003). To avoid this oversimplification problem, scholars have explored another dimension of sprawl measurement: its internal distribution pattern. As noted by Jaeger et al (2010), "every meaningful method to measure the degree of urban sprawl needs to be based on a clear definition of urban sprawl disentangling causes and consequences of urban sprawl from the phenomenon of urban sprawl itself, as urban sprawl has differing causes and consequences in different regions and regulatory contexts"[7]. In this sense, the increasing availability of spatially disaggregated data on land use and land cover from aerial photography and satellite imagery has also been instrumental. As a result, a growing body of literature on geography and urban planning is focusing on accurate and more refined measures of urban sprawl, providing considerable information about urban form beyond density and suburban population growth (e.g. Europe: Schwartz, 2010, Arribas-Bel et al, 2011; Israel: Frenkel and Ashkenazi, 2008; U.S.: Galster et al, 2001, Song and Knaap, 2004, Tsai, 2005, Wolman et al, 2005, Torrens, 2008; India: Battha et al, 2010; Switzerland: Jaeger et al, 2010). This has not been the case, however, in (regional and urban) economic research, where simple sprawl measures continue to be used. A notable exception is Burchfield et al (2006), who use satellite photographs to define a measure of sprawl that explicitly accounts for the scatteredness of urban development[8].

The urban sprawl variable. As noted above, there are many different ways of measuring urban sprawl. Here we propose one index to characterize this urban development pattern, which is based on the spatial distribution of different land uses. This index is similar to the index adopted by Burchfield et al. (2006) and it is based on the data provided by the European Commission and European Environment Agency CORINE (Coordination of Information on the Environment)

---

[5] For further details of the definition of urban sprawl, see EEA (2006), Hayden and Wark (2006), and Bruegmann (2005).

[6] See Battha et al (2010) for a review on the measures of urban sprawl.

[7] See also Galster et al (2001).

[8] Specifically, they have constructed a new land database by merging high-altitude photographs with satellite images. Then the U.S. map they obtain distinguishes all different types of land use. By dividing the map into 30x30 metre cells, they calculate an index of urban sprawl which considers the percentage of non-urbanised land inside the square kilometre surrounding each residential development.

Land Cover (CLC) Programme. This project is a satellite-based GIS source intended to provide consistent localised geographical information on land cover in 44 (1990) or 85 (2000) classes under five categories (artificial surfaces, agricultural areas, forest and seminatural areas, wetlands and water bodies), at a scale of 1:100,000, a minimum mapping area of 25 hectares and a minimum width of linear elements set to 100 metres. This land use database is operationally available for most areas of Europe, and has been recorded on three dates: 1990, 2000 and 2006[9]. For this study, raster data with a resolution of 30 metres for land cover of Spain in 2000 were obtained from the Spanish National Geographic Institute (responsible for the CLC data in Spain).

Our sprawl index calculates the percentage of undeveloped land (i.e. total land minus urban fabric, industrial and commercial units) around residential land (defined as discontinuous and continuous urban fabric[10]) within the immediate neighbourhood. To determine the size of the neighbourhood, nearest neighbour distance was computed on discontinuous urban fabrics. According to the cumulative distribution function of the pair distances between all discontinuous urban fabrics, 75% of the residential development (i.e. discontinuous urban fabric) is less than 1.8 kilometres from other residential development. Therefore, the immediate neighbourhood is defined as 1.8x1.8 square kilometres[11]. First, the vector dataset of land use was converted to a raster dataset with 30-by-30-meter cells. For each cell of residential land, the percentage of undeveloped land in the immediate neighbourhood was computed. Finally, the average percentage of all residential cells in a spatial unit (e.g., municipality or urban fabric) was calculated. This index characterises the percentage of open space (i.e. undeveloped land) surrounding an average residential development that could be developed. We calculate this index for all municipalities within urban areas and then examine the reasons why sprawl differs across space.

3. The causes of urban sprawl

3.1. Strategic interaction among local governments: a spatial econometrics approach.

---

[9] Some problems with digitisation and image interpretation of land covers in the most recent dataset (year 2006) prevented us from using it and the 2000 data was used instead. Indeed, an exploratory analysis of the data showed that it had many missing values (i.e. the urban developments of some municipalities were not included in 2006), which made the use of the variable unfeasible.

[10] Residential areas have been classified into two main classes, continuous and discontinuous. The main difference between the classes is the intensity of land use: in the continuous class, buildings and related structures cover more than 80% of the total surface and, in the discontinuous class, coverage ranges from 10 to 80%.

[11]Spanish municipalities are very heterogeneous in terms of extension and population (there is no correlation between size (extension) and population), so discontinuous urban fabric leapfrogs more than in the U.S. and it is difficult to obtain a sensible distance that captures 97% of the discontinuous urban fabric (as in the U.S. case, see Burchfield et al, 2006). In fact, the distance that captures 97% of the discontinuous urban fabric would be too large (10 km), meaningless for the phenomenon being measured. Thus, 1,8km is taken as the distance for the dependent variable, as it captures 75% of the discontinuous urban fabric.

Urban sprawl is the outcome of many agents' decisions involved in land-use development. Land developers, on the one hand, act as profit-maximising agents in construction activity. Residents/homeowners, on the other hand, act as utility maximisers who want to fulfil their individual housing preferences for larger, single-family detached housing, greater proximity to open spaces, and segregation from some of the problems suffered by the inner city. Local governments' activity is also prominent, and local government finance is a major factor that determines their contribution to the process of urban development. While urban sprawl increases the provision costs of certain local public goods and requires greater investment in extending basic infrastructure for new urban development located at the urban fringe, it also results in higher incomes and tax revenues. In addition, poorly defined land-use regulations together with the absence of control and intergovernmental coordination on matters relating to urban planning, have given municipalities plenty of room to manoeuvre in their urban growth decision-making. Local authorities have become land developers, making the construction of new residential developments a competitive enterprise and their main funding tool, in particular, in an environment with interjurisdictional competition for mobile residents[12]. Indeed, they compete for limited tax bases, with an aim to attract the middle and upper income taxpayer to their jurisdictions, as it translates to more revenues linked to construction activity (such as planning permissions, construction taxes or taxes on land value improvements) as well as the impact on property tax, the main tax revenue source on a local scale. This competition, however, causes development to occur too quickly compared to welfare maximising development, leading to a non-optimal level of urban development, hence fostering urban sprawl (Haubeck, 2009).

Strategic interaction among local governments is, therefore, a key element in explaining the urban expansion of cities and, as such, these spatial effects must be taken into account. In imperfectly competitive urban land markets, local policy decisions to develop land at a particular location are thus interdependent, as they may be affected by the decisions of other nearby local governments. Unlike previous studies on land use modelling, where jurisdictions are considered to be spatially uncorrelated, this paper analyses the causes of sprawl by explicitly taking spatial dependence into account. We aim to provide evidence that local governments' strategic behaviour is an additional cause of urban sprawl, as their land-use decisions interact with each other across space. We therefore incorporate spatial econometrics into our empirical specification. A given city is likely to be interacting with many competing cities in the housing market, and the challenge is to allow for such interaction in the empirical specification. Hence, the urban sprawl of a given municipality will depend on its own city characteristics and on a variable measuring the amount

[12] In Spain, land-use regulatory responsibilities are shared by different levels of government. The central government establishes the land-use regulation benchmark (as regards the protection of areas designated 'non-developable'), while local governments are responsible for passing municipal land-use plans. Local authorities enjoy considerable freedom in determining a municipality's urban planning and, in practice, they control the supply of urban land for real estate development.

of urban sprawl in competing neighbouring cities. In the spatial econometrics literature, we capture these effects by means of a Spatial Autoregressive or Spatial Lag model, which will be formally introduced in Section 4. Note that the spatial lag econometric model is appropriate when there is a theoretical model of the structural interaction among local governments determining the levels of urban sprawl, and one is interested in measuring the strength of that interactive relationship.

3.2. The Determinants of Urban Spatial Sizes: What Does the Literature Say?

As noted above, it is desirable to include a large number of city characteristics in the empirical model to fully capture the causes of urban sprawl. This strategy also helps to partially eliminate spatial error dependence, which arises when spatially dependent variables are omitted from the model. Sprawl is the result of a complex set of interrelated demographic, socioeconomic, political and geographical forces and, as such, characteristic variables can be summarised as follows[13].

The monocentric city model and its generalisations: Tiebout sorting, individual preferences and local amenities. The urban economics literature focuses on the Alonso-Muth-Mills monocentric city model to explain the basic determinants of urban sprawl. In this setting, three fundamental forces (i.e. population growth, rising real incomes and falling commuting costs) are responsible for the increasing demand of land in the suburbs and, therefore, for the spatial growth of cities. Brueckner and Fansler (1989), as well as the extensions made in McGrath (2005) and Song and Zenou (2006), confirm the robustness of the Alonso-Muth-Mills model through their regression findings. Nonetheless, urban development due to these three fundamental forces cannot be faulted as inefficient, unless certain market failures distort their operation (see Brueckner, 2000). In that situation, the invisible hand fails to allocate resources in a socially desirable manner, so as to maximise aggregate economic welfare.

As noted in Wassmer (2008), however, the Alonso-Muth-Mills model does not account for other household characteristics due to the assumption that, with the exception of income, households are identical in the characteristics that influence their land use preferences. As a result, the monocentric city model leads to the identification of the primary historical cause of urban sprawl, but empirical evidence strongly suggests that factors other than population growth, transport and commuting costs or income are more likely to be driving the process today. Mieszkowski and Mills (1993), for instance, explain urban sprawl in terms of Tiebout sorting. Residents *vote with their feet* and choose their location within an urban area depending not only on their income and transport costs, but also according to their preferences. In this context, fiscal

[13] See also Nechyba and Walsh (2004) and Ewing et al (2014) for a literature review on the causes and consequences of urban sprawl.

and social problems of central cities lead middle-class residents to move to the suburbs, so that they form separate homogeneous communities of individuals of like income, education or race.

Burchfield et al (2006), on the other hand, consider that the monocentric city model cannot explain leapfrog development where parcels of land are left undeveloped while others farther away are built up. According to the authors, one explanation has to do with the amenity value to public open space: individuals may be willing to incur the additional commuting costs associated with locating further away from the city centre in order to have open space near their home. As a result, those locations endowed with desirable amenities that make public open space more attractive (i.e. a pleasant temperate climate, forests, dryness – an inverse of average precipitation- or proximity to the coast) will experience more sprawl (Glaeser, Kolko, Saiz, 2001; Burchfield et al, 2006)[14].

In addition, if moving is costly, the willingness to trade-off commuting costs against access to public open space will depend on expectations of how long that space will stay undeveloped. In areas where the population is growing fast, a rational agent anticipates that nearby vacant land will be developed sooner and, consequently, is not willing to incur large additional commuting costs to gain access to this open space. Developers may expect that cities that have been growing relatively fast in the past will continue to do so in the near future (Burchfield et al, 2006).

Physical geography. Despite technological progress, the physical environment continues to play an important role in shaping cities (Burchfield et al, 2006; Saiz, 2010). Nature can either promote or contain sprawl through physical barriers hindering urban expansion. For instance, the presence of aquifers (inland waters) can facilitate suburban development, as it lowers the cost of obtaining household water, whereas land is undevelopable whenever water bodies (such as wetlands and oceans) are present and, therefore, sprawl is contained. The presence of mountains also limits urban expansion, as they make development more costly. In contrast, small-terrain irregularities have the opposite effect, as hillsides where development is more costly alternate with flat portions where development is less costly.

The role of politics: public spending in roads and highways. Both European and American sprawl have arisen at least partially from government policy. There has been considerable investment in public transport and infrastructure by public authorities over the last twenty years. As a result, a growing body of the literature has focused on the influence of transportation system improvement and availability of roads on urban growth (see, for example, Baum-Snow, 2007; García-López, 2012; Duranton and Turner, 2011).

The structure of public finance and local revenue reliance. Another theory of urban sprawl is related to the structure of local public finance and revenue reliance (e.g. Slack, 2002). Firstly,

[14] Several other studies reveal the link between local amenities and residential location (see, for example, Brueckner, Thisse and Zenou, 2000; Cullen and Levitt, 1999; Hortas-Rico, 2013; Wu, 2006).

a fiscal effect arising from local property taxation may also contribute to urban sprawl. Property tax reduces the intensity of land development, lowering population density and, in turn, causing cities to excessively spread out (Brueckner and Kim, 2003; Song and Zenou, 2006). Secondly, many local governments face fiscal viability problems and use grants to balance their budgets and this apparent softening of budget constraints could distort local policy decisions (Hortas-Rico, 2013). The over-reliance of municipalities on grants to make adjustments to their budgets also highlights a potential moral-hazard problem. Additional infrastructure requirements associated with spatially expansive growth are funded in the main by upper tiers of government, encouraging municipalities to promote urban expansion without necessarily considering the full fiscal consequences of such policies (Hortas-Rico, 2013). In addition, there are inefficiencies attributed to grant financing of new urban developments on the urban fringe, as new developers fail to internalise the full costs that they generate, leaving the local government to pay for them. As a result, sprawl does not pay for itself but rather becomes a burden on all taxpayers (Slack, 2002).

4. Econometric strategy

A satisfactory strategy to find the spatial econometric model that best describes the data must be based on theoretical grounds, and focus on the parameter of interest, whereas searching for a source of exogenous variation that can plausibly be used to identify this parameter of interest is also crucial (Gibbons and Overman, 2012). The modelling strategy in the spatial econometric literature is under revision and two different approaches can be identified. The standard approach in most empirical work is to start with a non-spatial linear regression model and then test whether or not the model needs to be extended with spatial interaction effects (*specific-to-general*)[15]. Alternatively, according to Elhorst (2010), we should no longer be limited to the Spatial Lag or Error model but begin with the largest possible specification, subsequently attempting to simplify it (*general-to-specific*). There are many pros and cons for each approach and the evidence seems to favour *general-to-specific,* although not to the extent suggested by Hoover and Perez (2004). Mur and Angulo (2009), in a simulation exercise, conclude that under all standard assumptions both strategies produce hardly distinguishable results. However the *general-to-specific* approach produces better results when distortions, such as non-normality in the errors or heteroskedasticity with a spatial pattern, are introduced into the data generation process (DGP). On the other hand, the impact of endogeneity on the explanatory variables seems to be more acute in the *general-to-specific* approach. We rely on theory and assume the Spatial Lag model is the preferred specification, as mentioned in the previous section, but will compare the results from both approaches due to violation of some of the assumptions in the DGP.

[15] In applied econometrics, an implicitly or explicitly *specific-to-general* approach is predominant.

We first follow the *specific-to-general* approach, estimating the non-spatial model by Ordinary Least Squares (OLS). Then we test whether the Spatial Lag model or the Spatial Error model is more appropriate describing the data. For this purpose, we use the classic Lagrange Multipliers tests on estimated residuals (LM-tests) and their robust versions. These tests reinforce the theoretical assumption designating that the preferred specification is the Spatial Lag model presented in equation (1)[16].

$$\boldsymbol{y} = \alpha\boldsymbol{\tau_N} + \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{y}$ represents the N-dimensional vector consisting of one observation on the dependent variable for every unit in the sample $(i = 1, \dots, N)$, $\boldsymbol{\tau_N}$ is an $Nx1$ vector of ones associated with the constant term parameter α, $\boldsymbol{X}$ denotes an NxK matrix of explanatory variables, with associated parameters $\boldsymbol{\beta}$ contained in a $Kx1$ vector, and $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_N)'$ is a vector of independently and identically distributed (i.i.d.) disturbance terms with zero mean and variance σ^2. $\boldsymbol{W}$ denotes the NxN spatial weight matrix that describes the structure of dependence between spatial units in the sample[17]. It has zero diagonal elements (as no spatial unit can be viewed as its own neighbour) and a representative off-diagonal element is w_{ij}. The values of the w_{ij}'s are specified arbitrarily to reflect prior expectations regarding the spatial patterns of interaction. The variable Wy denotes the endogenous interaction effects among the dependent variables (i.e. the spatial lag), and the coefficient $\boldsymbol{\rho}$ measures the intensity of interaction between location pairs.

We first estimate the model through Maximum Likelihood (ML) but specification diagnostics (Kiefer-Salmon test) determine that residuals are not normal, so we need to implement a robust method such as instrumental variables (IV) techniques (in particular, Generalized Spatial Two Stage Least Squares, GS2SLS hereinafter). As spatial units have different sizes, another assumption that does not hold is homoscedasticity, as shown by the Koenker-Basset heteroskedasticity test (KB=746.28, p-value<2.2e-16).

Alternatively, the *general-to- specific* approach starts with the most general model, i.e. the model that includes a spatially lagged dependent variable, spatially lagged independent variables, and a spatially autocorrelated error term simultaneously:

$$\begin{aligned} \boldsymbol{y} &= \alpha\boldsymbol{\tau_N} + \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{WX\theta} + \boldsymbol{u} \\ \boldsymbol{u} &= \boldsymbol{\lambda W u} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma_i^2) \end{aligned} \tag{2}$$

[16] Robust LM lag = 170.5312, p-value < 2.2e-16 vs Robust LM error = 18.8145, p-value = 1.441e-05.
[17] Queen contiguity weight matrix was used. For a robustness check, other spatial weight matrices were used, leading to similar regression results (the results are available upon request to the authors).

The notation is the same as in equation (1) and the additional terms are the exogenous interaction effects among the independent variables, $\boldsymbol{WX}$, and the interaction effects among the disturbance terms of the different spatial units, $\boldsymbol{Wu}$, where u follows an autorregresive process and $\boldsymbol{\epsilon}$ is a white noise. $\boldsymbol{\lambda}$ is the spatial autocorrelation coefficient, and $\boldsymbol{\theta}$ is a $Kx1$vector that captures the fixed but unknown parameters in case of spatial dependence on the explanatory variables. However, as Manski (1993) notes, at least one of the interaction effects must be excluded, because otherwise the parameters are unidentified. At this point we face two alternatives, starting with the Spatial Durbin model (SDM), which excludes the autocorrelated error term $\boldsymbol{u}$ (i.e. imposing the restriction $\boldsymbol{\lambda} = 0$), or starting with the SAC or Kelejian-Prucha model, excluding the spatially lagged independent variables $\boldsymbol{WX}$ (i.e. imposing the restriction $\boldsymbol{\theta} = 0$)[18]. Note that we obviously do not exclude the spatial interaction in the dependent variable as it is the variable of interest in the theory we want to test. LeSage & Pace (2009) argues that one should start with the SDM:

$$\boldsymbol{y} = \alpha\boldsymbol{\tau_N} + \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{WX\gamma} + \boldsymbol{\epsilon}\,; \qquad \boldsymbol{\epsilon} \sim N(0, \sigma_i^2) \tag{3}$$

and then check whether the estimated parameters produce a Spatial Lag or a Spatial Error model. Nonetheless, we follow the Elhorst (2010) test procedure to find out which model is the most likely candidate to explain the data. First, we estimate the OLS model and use LM-tests to test whether it should be expanded to either a Spatial Lag or a Spatial Error model. As the OLS model is rejected in favour of the Spatial Lag and Spatial Error model, we then estimate the SDM. When these models are estimated by ML, a likelihood ratio (LR) test can subsequently be used to examine whether the SDM can be simplified to either the Spatial Lag or the Spatial Error model. The Spatial Lag model does not best describe the data, even the (robust) LM pointed to the Spatial Lag model, as we can reject the null hypothesis in the LR test as LR=143.1 (p<2.2e-16). Similarly, when we implement the LR test to check whether the Spatial Error model should be estimated, we also reject the null (LR=216.71; p<2.2e-16). As the (robust) LM tests point to a different model than the LR tests, then the SDM should be adopted, as this model generalizes both the Spatial Lag and the Spatial Error model.

This procedure hinges crucially on the ML estimation method testing hypothesis. However, disturbances are not normal, invalidating this estimation method. A second potential problem is the endogeneity of the regressors included in X (i.e. feedback effects between city characteristics and the level of urban sprawl), which could bias the estimated coefficients present in equation (2). Single equation ML estimators cannot handle endogenous explanatory covariates.

[18] The best solution excludes the spatially autocorrelated error term, because the cost of ignoring spatial dependence in the disturbances only determines efficiency lost. By contrast, the cost of ignoring spatial dependence in the dependent or independent variables is relatively high (omitted variable bias).

Fingleton and LeGallo (2008) show that IV/GMM estimators are extremely useful in those cases where linear spatial dependence models contain one or more endogenous explanatory variables that have to be instrumented[19]. Nevertheless, the structuring of the GS2SLS (and equivalent heteroskedastic version) makes it effectively impossible to fit a SDM. Even if one tries (using higher lags by hand), the results are typically numerically unstable. Pace et al. (2013), using Monte Carlo experiments, show that the performance of IV techniques, especially when estimating SDM models, can be sensitive to spatial correlation in the regressors even when using thousands of observations. In other words, although the identification problem seems to be solved, the accuracy of the estimator becomes a problem.

At this point the violation of the assumptions prevents us from using the ML estimation method and, as we cannot implement IV techniques for estimating the SDM, we are left with the second alternative, starting with the SAC general model[20]. Piras (2010) and Kelejian and Prucha (1998) argue that the spatial patterns involved in a SAC model are richer than those involved in either the Spatial Error or the Spatial Lag model. As previously mentioned, our empirical application requires the use of spatial heteroskedasticity and autocorrelation consistent (HAC) estimators[21]. Piras (2010) allows for the estimation of two different models, one assuming no specific structure in the disturbance process and, as a special case, a second alternative that assumes an autorregresive structure in the residuals. First, we estimate the SAC model through the GS2SLS estimation method assuming that the disturbance vector is generated by a very general process where $\boldsymbol{\varphi}$ is a vector of innovations and $\mathbf{R}$ is a NxN non-stochastic matrix the elements of which are not known[22]:

$$\boldsymbol{y} = \alpha\boldsymbol{\tau_N} + \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon}\,; \quad \boldsymbol{\varepsilon} = \boldsymbol{R\varphi} \tag{4}$$

[19] The GMM estimator has significant advantages over the ML version of the model. Though ML provides consistent and efficient estimates when the model is correctly specified and the errors truly follow the assumed normal distribution, there is no reason to expect in practice that the errors are actually normally distributed or homoscedastic; GMM estimation does not require any distribution-related assumptions and it often turns out to be less sensitive to model misspecification.

[20] The general model (SAC) faces identification issues; it is, however, used in some analyses, often in GM estimators, so the ML version is useful for comparison, but does need care in selecting starting values for numerical optimisation.

[21] The spatial HAC estimator is robust against possible misspecification of the disturbances and allows for (unknown) forms of heteroskedasticity and correlation across spatial units. Nonetheless, even if we assume such a general specification for the disturbance process we still have to be concerned about possible misspecifications (e.g., due to an incorrect specification of the weights matrices).

[22] Note that this specification of the error term covers SARMA(p,q) processes as special cases. We test the robustness of the model specification to different Kernel functions, and also compare the coefficients significance to the results obtained with the robust estimator to those obtained under the non-robust standard errors.

Subsequently, we assume that the disturbance process ϵ is known to follow a first order spatial autoregressive process:

$$\boldsymbol{y} = \alpha\boldsymbol{\tau_N} + \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{\epsilon}\,; \qquad \boldsymbol{\epsilon} = \boldsymbol{\lambda W \epsilon} + \boldsymbol{\varphi} \tag{5}$$

where innovations $\boldsymbol{\varphi_1}, .., \boldsymbol{\varphi_n}$ are assumed independent with zero mean and non-constant variance σ_i^2. The suggested estimation procedure consists of two steps alternating Generalized Moments (GM) and IV estimators (see Piras, 2010). Kelejian and Prucha (2010) give results concerning the joint asymptotic distribution of IV and GM estimators for the SAC Model. Their results enable testing the (joint) hypothesis of no spatial spillovers originated from the endogenous variables or disturbances. As a result, we observe that the residual autorregresive coefficients turn out not to be statistically different from zero, pointing to the spatial Lag model as in the *specific-to-general approach*.

To recapitulate, the econometric strategy produces the same results under both approaches. First when we implement the *specific-to-general* approach the preferred specification is the Spatial Lag model estimated through the correct IV method. When the *general-to-specific* approach is adopted we cannot estimate the SDM under an IV method so we are forced to start from the SAC model. The results determine that the Spatial Lag model is the preferred specification as the autoregressive parameter in the disturbance was not significantly different from zero.

5. Regression model and data

Following the urban economic theory and the literature outlined above, the determinants of urban sprawl can be examined by estimating the regression given by expressions (4) and (5), where y represents the vector consisting of one observation on the urban sprawl variable for every municipality in the sample for the year 2000. As noted above, the variable Wy denotes the endogenous interaction effects among the dependent variables, and the coefficient on this "competing sprawl" variable,ρ, measures the strength of the dependence between municipality pairs. This autoregressive parameter indicates how a given city responds to the level of sprawl in nearby jurisdictions, giving the slope of its reaction function. A non-zero coefficient indicates that sprawl choices are interdependent across cities, and strategic interaction occurs, whereas a zero coefficient means that strategic interaction is not present. In such situations, one city's urban sprawl choice is unaffected by the position of the other city's reaction function, which depends on that city's characteristics and objectives. X denotes a matrix of observed municipality's characteristics in the initial year (1990) expected to influence differences in the residential development patterns of urban areas (see definitions and data sources in Table 1), with associated parameters β. The covariates that fill out the matrix X can be grouped into four categories: (i) A basic set of control variables to account for *the monocentric city model and its generalisations*: population, vehicles per household, certain demographic and political characteristics that capture residents' preferences for development (i.e. % Aged 25-45, % Aged >65, number children, % Graduated, left government dummy) and employment structure (% Manufacturing, % Retail, % Other services). Note that, as population might be endogenous, we replace it by the mean decennial population growth; (ii) A set of *amenity factors* deemed important for location decisions (i.e. mean maximum temperature, mean minimum temperature, average rainfall, % open space (forest and agricultural land) and % land devoted to leisure and sports facilities), and two variables that proxy *urban blight* (i.e. poverty and % immigrants); (iii) Two budget variables that measure the *structure of public finance and local revenue reliance* (i.e. % property tax, % transfers); (iv) An additional set of variables to account for *physical geography*. Here we include elevation range, the terrain ruggedness index and percentage of wetlands and inland water bodies; (v) Finally, two variables that account for the amount of *public spending on roads and highways* (i.e. km of roads, distance to the nearest road), and a last variable that measures the distance to the central city, so as to capture whether urban sprawl increases or diminishes as distance from the city centre grows[23].

The sample. According to the European Environmental Agency (EEA, 2006), approximately 75% of the European population live in urban areas, a proportion that will have risen to 80% (or even 90% in certain countries) by 2020. The demand for land in and around cities is therefore in a constant state of growth, occurring in a scattered way throughout Europe's countryside, and reshaping landscapes everywhere. This spread urban expansion is regarded as one of the major common challenges facing urban Europe today, and European cities are recognised as key players in promoting effective land use and development policies (Schwartz, 2010).

The definition of urban areas is not, however, straightforward. Urban or metropolitan areas do not fit well with administrative boundaries, as they change over space and time reflecting the evolution of the economy and society. In addition, comparison between the metropolitan units identified in different countries is difficult as countries use different methodologies. There is, therefore, a need for divisions that adequately reflect urban reality, at least more accurately than the administrative divisions (e.g. NUTS) used by the European Union. Boix et al. (2012) proposed a general methodology to identify functional metropolitan areas for comparative purposes[24],

[23] We measure proximity (Euclidean distance) from each municipality centroid to the urban area's central city centroid.

[24] It is an adaptation of the U.S. Federal Register's methodology (Office of Management and Budget, 1990) and represents an improvement with respect to the Functional Urban Regions (FURs). Similar to the FUR, the metropolitan area comprises a central core (which must have at least 50,000 inhabitants) and a hinterland (determined according to the commuting threshold of the neighboring municipalities where resident employees commutes to the central core). The main difference is that the initial relative threshold of

according to which one can identify 67 metropolitan areas in Spain, which clearly diverge from the administrative boundaries (regions or provinces). They account for 49% of Spanish municipalities (around 4,000), 76% of the population (31 million) and 77% of employment (16.3 million jobs). In addition, almost half of the metropolitan population and employment is found in the country's largest metropolitan areas, with more than one million inhabitants. In particular, there are five large metropolitan areas in Spain (Madrid, Barcelona, Valencia, Seville and Bilbao) which account for about 35% of the national population and 38% of employment. Nonetheless, for computational reasons and given the spatial treatment of the data, we exclude the urban areas of Ceuta, Melilla and those located in the Balearic and Canary Islands. Hence, we end up with 62 urban areas (see Map 1).

commuting for the formation of the core and the hinterland is more exigent although it is iterated to take advantage of the trend of labour markets to be self-contained. For further details, see Boix et al (2012).

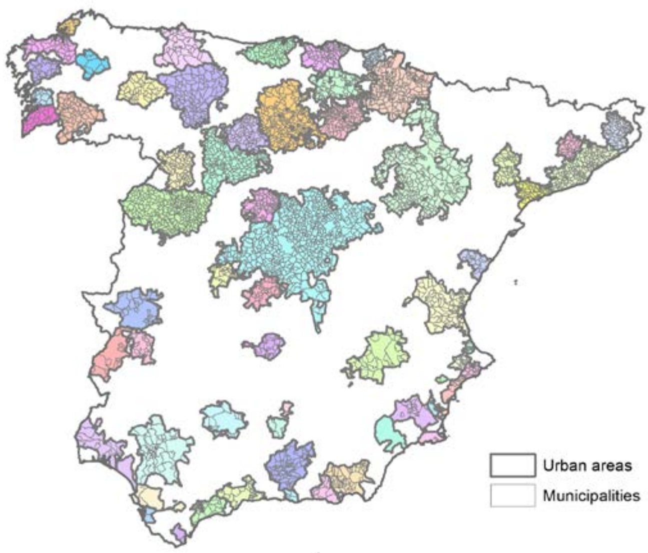
Table 1. Definitions and sources of the explanatory variables

	Definition	Sources
Population growth 1930-1990	Population growth rate (endogeneity: proxy it with mean decennial percentage population growth for the six decades 1930-1990)	Historical Census of Population. National Institute of Statistics (INE)
Vehicles per household	Average number of vehicles per household in 1991	Census of Population and Housing, 1991. National Institute of Statistics (INE).
% Aged 25-45	[Population between 25 and 45 years old in 1991 / Total resident population in 1991]x100	
% Aged > 65	[Population over 65 years old in 1991 / Total resident population in 1991]x100	
Children family	Average number of children per household in 1991	
% Graduate	[Residents with a higher education degree in 1991 / Total resident population in 1991]x100	
Poverty	Poverty level proxied by the percentage of resident population without studies: [Resident population over 10 years old without studies or illiterate in 1991 / Resident population in 1991]x100	
% Immigrants	[Non-EU immigrant population in 1991 / Total resident population in 1991]x100	
% Employed in manufacturing	[Employed in manufacturing in 1991 / Employment 1991]x100	
% Employed in retail	[Employed in retail in 1991 / Employment 1991]x100	
% Employed in other services	[Employed in other services in 1991 / Employment 1991]x100	
Mean max temp	Mean annual maximum temperature (calculated from the climatic normals for individual weather stations)	Series climatológicas mensuales, 1990. Agencia Estatal de Meteorología. Ministerio de Agricultura, Alimentación y Medio Ambiente
Mean min temp	Mean annual minimum temperature (calculated from the climatic normals for individual weather stations)	
Average precipitation	Mean annual precipitation (calculated from the climatic normals for individual weather stations)	
% Open space	[Forest and agricultural area (including vineyards, rice fields, friut trees plantations, olive groves, etc) / Total land area]x100	Corine Land Cover, 1990.
% Leisure and sports facilities	[Leisure and sports facilities area / Artificial land area]x100	
Water availability (%)	[Wetlands area / Total land area]x100	
Inland water (%)	[Inland waters / Total land area]x100	
Terrain ruggedness index (km)	Municipal average value of the terrain ruggedness index developed by Riley et al (1999), calculated on the 200-meter elevation grid to give a summary statistic of differences in meters of elevation between points 200-meters apart.	Spanish 200-meter digital elevation model. Instituto Geográfico Nacional. GIS
Elevation range (km)	Elevation range for each municipality	Spanish 200-meter digital elevation model (MDT200, sistema geodésico de referencia ETRS89, por provincias). Instituto Geográfico Nacional. GIS
Road density (km/pop)	[Kilometers of roads (main and secondary roads 18th century) / Total resident population in 1990]	Instituto Geográfico Nacional. GIS
Distance to road (km)	Distance from municipality centroid to the nearest main or secondary road 18th century (km)	Instituto Geográfico Nacional. GIS
Distance to central city (km)	Average distance of each residential development centroid to the centroid of each urban area's central city.	Corine Land Cover, 1990. GIS
Left	Dummy=1 if the major belongs to a left party during the 1991-1994 term, 0 otherwise. Parties on the left are: PSOE, PCE, IC, and several left regionalist parties.	Dirección General de Política Interior, Ministerio del Interior. Consulta de resultados electorales. Elecciones
Property tax revenues as % of local revenues	Total local revenues from property taxes in 1995	Spanish Ministry of Finance (Liquidación de Presupuestos de las Entidades Locales,
Intergov. transfers as % of local revenues	Total local revenues from transfers in 1995	

Notes: physical geography variables and other relevant distance measurements have been calculated using Georaphical Information Systems. All data is at the level of municipality.

The final sample includes all municipalities within the urban areas for which we also have data on all the explanatory variables but the budgetary ones (3,131 observations out of 3,895). The inclusion of the two budget variables (transfers and property tax revenues as % of local revenues) would have reduced the sample size considerably. The reason is that official budget data was not available for many cities (especially the smaller ones) in the early nineties. The estimation without these two variables was preferable to such shrinkage in the sample, given the importance of space in our analysis. Nonetheless, additional estimations including the budget data yield to similar results (with a sample size of 1,914 observations)[25]. As for the time period covered, it is important to notice that in Spain, as in the rest of Europe, the annual rate of change in land-cover type (from rural to urban uses) peaked during the 1990s, and slowed down from 2000 to 2006. Indeed, 30% of the artificial surfaces in existence today were created during the nineties (EEA, 2006).

Map 1. Spanish urban areas



Own elaboration

6. Main results

Non-spatial linear regression parameters provide consistent estimates of the marginal impacts of explanatory variables on the dependent variable, which are identified with the partial derivative of the dependent variable relative to the explanatory variable (or change in own dependent variable holding all others' dependent variables constant). But models containing

[25] Results are available upon request to the authors.

spatial lags of the dependent variable require special interpretation of the parameters, as spatial regression models expand the information set to include information from neighbouring regions/observations. In such cases, the total derivative would be the combined effect of all dependent variable changes in the simultaneous equilibrium, as a change in the explanatory variable for a single region/observation can potentially affect the dependent variable in all other observations/regions (*spillover effects*). This impact includes the effect of *feedback loops* where observation i affects observation j and observation j also affects observation i as well as longer paths which might go from observation i to j to k and back to i (LeSage and Peace, 2009). Thus, the spatial lag model estimate of β obtained after spatially filtering the dependent variable is a consistent estimate of the direct, or marginal, impact of X on y in the equilibrium for the system. All the results presented here correspond to the post-estimation summary measures of the *direct impacts*.

The results of the GMM estimation (GS2SLS) of the model given by expressions (4) and (5) are presented in Table 2. Columns (1) through (5) in Table 2 are informative, but for the remainder of the paper only the results of Column (6) – direct impacts - will be discussed. To aid comparison across variables, we report standardised coefficients that measure the absolute change in the urban sprawl variable for a one standard deviation change in each explanatory variable.

The most important finding from Table 2 is that the estimated coefficient of sprawl interaction (Wy) is positive and statistically significant at well over a 99% confidence level, and occurs with a magnitude of around 0.32 regardless of which model specification (and even estimation method or weighting scheme) is considered[26].This finding provides evidence of spatial interaction in the levels of urban sprawl between neighbouring municipalities, which turn out to be strategic complements. A local government's level of urban sprawl is positively influenced by the degree of urban sprawl in neighbouring jurisdictions, with other causal factors remaining constant. This result could suggest that local authorities engage in strategic competition for the construction of new residential developments located on the urban fringe and mimic each other so as to attract new residents to their jurisdictions, hence promoting urban sprawl.

We now consider the impact of the control variables. In general, all variables considered have the expected sign and are consistent with a priori expectations derived from urban economics theory, although a few of them turn out to be not statistically significant.

We begin by examining the relationship between population growth and urban sprawl. As population growth might be endogenous, we replace it by the historical mean decennial percentage population growth for the seven decades from 1930 to 1990. This variable accounts for expected future population growth and, as such, has a negative and significant impact. Hence, areas that have historically seen high population growth rates do, indeed, see less sprawl. As

[26] See Appendix.

explained in Burchfield et al (2006), in fast growing cities rational agents anticipate that nearby vacant land will be developed sooner so that they are not willing to incur in additional commuting costs to gain access to this open space.

Auto-reliance has been considered as an additional cause of urban sprawl. According to the monocentric city model, rising incomes and declining transportation costs have fostered the use of the private motor vehicle which, in turn, enables individuals to commute longer distances, causing urbanisation to spread out more. As expected, number of vehicles per household has a positive and statistically significant impact on urban sprawl. A one-standard deviation increase in the number of vehicles per household increases the sprawl index by 2.4 points.

As regards the socio-demographic variables included in the model to account for demand factors and preferences, several interesting findings arise from our results. We find that more sprawling cities exhibit a higher percentage of elderly people and families with children, whereas younger and more educated citizens prefer to live in more compact locations. A one standard deviation increase in the percentage of population greater than age 65 increases the sprawl index by 1.6 points, and a one-deviation increase in the number of children per family increases sprawl by 0.34 points. However, a one-standard deviation increase in the percentage of graduated citizens reduces sprawl by 0.7 points. In addition, the results seem to suggest that people do indeed flee from blight, as the coefficients of the poverty and immigrants variables are both negative, although not statistically significant. The left government dummy, included in the model to account for the influence of politics on land-use decision-making, has a negative and significant effect, so locations that belong to a left party experience less urban sprawl than those where a right-wing party is present. This result is consistent with previous empirical studies where parties to the right of the political spectrum are expected to allow more land to be developed, thus promoting more scattered development.

We now turn to the interpretation of the link between employment structure and urban sprawl. Differences in type of employment are meant to pick up the independent influences that variation in different forms of non-residential activity has on the shape of urban development. As expected, differences in the economic base of municipalities in a given urban area influence its geographical footprint. Specifically, the greater the presence of manufacturing, retail and other services (employment sectors whose economics drive them to locate in more densely populated central places in urban areas to benefit from agglomeration economies), the lower the level of urban sprawl. A one-standard deviation increase in the percentage of employment in manufacturing, retail and other services decreases the sprawl index by 2.3, 1.7 and 2.4 points, respectively.

In order to further investigate the determinants of urban sprawl, an additional set of local amenity variables was added to the specification. The characteristics that make open space less attractive are expected to reduce urban sprawl. Indeed, an extremely hot or cold climate, as well as extremely rainy locations (proxied here by average rainfall) exhibit lower levels of urban sprawl. A one-standard deviation increase in mean maximum and minimum temperature reduces the sprawl index by 0.55 and 0.65 points, respectively, whereas a one-standard deviation increase in the average rainfall variable reduces sprawl by 0.8 points. In contrast, people care about the characteristics of their nearby residential landscape. We expect a positive amenity effect that arises from designating neighbouring land as preserved open space, as it can be associated with a scenic view, increased privacy or even guarantee of no neighbouring future development (Turner, 2005; Saiz, 2010). We find that the higher the percentage of open space, the higher the level of urban sprawl. In particular, a one standard deviation increase in the percentage of forests increases the sprawl index by 0.74 points. The percentage of land devoted to leisure and sports facilities also exhibit a positive impact on urban sprawl, although it is not statistically significant.

The next set of results relates to a range of geographical variables. Firstly, we consider the presence of water features by introducing two variables in our specification. On the one hand, the presence of aquifers (inland waters) can facilitate suburban development, as it lowers the cost of obtaining household water (Saiz, 2010). Our regression findings show a positive but non-significant impact of this variable on the sprawl index. On the other hand, land is undevelopable whenever water bodies are present and, therefore, sprawl is contained. We account for this possibility by including a variable that measures the percentage of surface occupied by wetlands and oceans. A one-standard deviation increase in this variable leads to a 1.2 point decrease in the sprawl index. Secondly, the presence of mountains also limits urban expansion, as they make development more costly. In contrast, small-terrain irregularities have the opposite effect, as hillsides where development is more costly alternate with flat portions where development is less costly. Two variables are included in our specification to account for these natural barriers to urban development. We introduce the elevation range as a measure of the presence of mountains, and we compute the terrain ruggedness index to account for the presence of small-scale terrain irregularities. We see that both variables have the expected effects, providing compelling evidence that physical geography does exert an influence on urban sprawl. Specifically, a one-standard deviation increase in the terrain ruggedness index increases the sprawl index by 1.64 points, while a one-standard deviation increase in the elevation range decreases the sprawl index by 0.65 points. These results are in line with those presented in Burchfield et al (2006) for the U.S. Physical geography of urban areas is a key element in explaining sprawl and, as in the U.S., it alone explains about 25 percent of the cross-city variation in the sprawl index.

There is vast literature focused on the relationship between public investments in infrastructure, the extension of highway networks and urban spatial structure. Thus, two additional variables that account for the amount of public spending in roads and highways are also considered. On the one hand, road density has a positive but meagre and non-significant impact on the sprawl index, consistent with previous empirical findings. For instance, Garcia-

López (2012) investigates the Barcelona Metropolitan Region and finds that improvements to the transportation infrastructure cause suburbanisation and influence its spatial pattern by attracting population to the suburbs. Baum-Snow (2007) also shows that transportation improvements do cause suburbanization in the U.S., as opposed to Burchfield et al (2006), who find a negative and non-significant impact of this variable on urban sprawl, arguing that more roads may facilitate suburban development, but sprawled development leads to a less dense road network. On the other hand, the distance of each municipality's centroid to the nearest road also has a positive but non-significant impact on the degree of scattered development. Note that, in order to avoid endogeneity problems due to reverse causation of urban sprawl and transportation improvements, a historical road map (main and secondary roads constructed before the end of the 18th century) has been used as a source of exogenous variation for the definition of both variables (i.e. road density and distance to road).

Finally, the distance from each residential development to the nearest urban centre is also expected to play an important role in determining the intensity of sprawl. Contrary to the U.S. case, where urban sprawl only increases as the distance from the central business district (CBD) grows (Schneider and Wookcock, 2008), European suburban development is also characterised for occurring as the distance to the CBD diminishes. In particular, scholars describe a dual model where new edge cities around traditional urban centres coexist with scattered residential developments located on the urban fringe. Catalán et al. (2008) highlight the importance of the existing urban fabric in the sprawl processes of Southern European cities. As urbanisation advances, much non-urban land disappears (crop land is the main non-urban loss and, to a lesser extent, forest land and shrub land), especially in the sub-centres and municipalities of the first metropolitan ring, which have the highest urban potential. Thus, proximity to the metropolitan urban core is crucial and can be justified on the grounds of higher costs of the compact urban model. Actually, the population shift towards the periphery of the urban areas may be the result of both personal choice and the high costs of living in central places. Perhaps this is one of the most important particularities of many Southern European cities, compared to the North-American urban context, where there is a strong social preference for single-family housing and the suburban life style. Our results show that this is indeed the case. The predominant pattern of urbanisation is diffuse settlements adjacent to concentrated urban centres. The parameter is negative and statistically significant, with a magnitude of 1.6. Hence, a one-standard deviation

Table 2. The determinants of urban sprawl. S2SLS n=3,131

	Spatial Lag			SAC (AR form for disturbance process)			SAC (general form for disturbance process)			Summary statistics	
	(1)		(2)	(3)		(4)	(5)		(6)	Mean	St. Dev.
Spatial dependence:	Estimated parameters		Direct impacts	Estimated parameters		Direct impacts	Estimated parameters		Direct impacts		
Rho	0.325***	(0.022)		0.316***	(0.034)		0.325***	(0.037)			
Lambda	--			0.076	(0.046)		--				
Control variables:											
Vehicles per household	2.434***	(0.217)	2.509***	2.461***	(0.269)	2.532***	2.434***	(0.288)	2.509***	0,991	0,274
Population growth 1930-1990	-4.414***	(0.229)	-4.548***	-4.391***	(0.868)	-4.517***	-4.413***	(0.879)	-4.548***	-2,519	19,981
% Aged 25-45	0.057	(0.253)	0.059	0.057	(0.219)	0.058	0.057	(0.224)	0.059	24,508	5,229
% Aged > 65	1.592***	(0.278)	1.641***	1.606***	(0.334)	1.652***	1.592***	(0.332)	1.641***	22,469	10,124
Children family	0.344**	(0.175)	0.355**	0.337**	(0.131)	0.346**	0.344***	(0.134)	0.355***	1,913	0,399
% Graduate	-0.694***	(0.207)	-0.716***	-0.705***	(0.213)	-0.725***	-0.694***	(0.216)	-0.716***	3,685	2,611
Poverty (% without studies)	-0.142	(0.192)	-0.147	-0.134	(0.149)	-0.138	-0.142	(0.154)	-0.148	29,284	19,837
% Immigrants	-0.429**	(0.181)	-0.442**	-0.421*	(0.222)	-0.433*	-0.429*	(0.229)	-0.442*	0,169	0,511
% Employed in manufacturing	-2.347***	(0.197)	-2.419***	-2.342***	(0.235)	-2.409***	-2.347***	(0.241)	-2.419***	19,207	13,559
% Employed in retail	-1.710***	(0.182)	-1.762***	-1.717***	(0.180)	-1.766***	-1.709***	(0.184)	-1.762***	12,464	7,161
% Employed in other services	-2.380***	(0.217)	-2.452***	-2.422***	(0.232)	-2.492***	-2.380***	(0.234)	-2.452***	21,561	9,248
Mean max temp	-0.556**	(0.227)	-0.573**	-0.576**	(0.240)	-0.592**	-0.556**	(0.232)	-0.573**	19,944	1,817
Mean min temp	-0.652***	(0.190)	-0.672***	-0.678***	(0.198)	-0.698***	-0.652***	(0.202)	-0.672***	9,031	2,182
Average precipitation	-0.793***	(0.198)	-0.818***	-0.821***	(0.243)	-0.845***	-0.793***	(0.239)	-0.818***	48,839	25,382
% Open space	0.741***	(0.242)	0.764***	0.811***	(0.226)	0.835***	0.741***	(0.223)	0.763***	7,538	12,778
% Leisure and sports facilities	0.133	(0.245)	0.137	0.164	(0.241)	0.168	0.133	(0.229)	0.137	0,003	0,027
% Water availability	-1.216***	(0.180)	-1.253***	-1.232***	(0.199)	-1.267***	-1.216***	(0.203)	-1.253***	0,151	1,704
% Inland water	0.419*	(0.250)	0.432	0.457	(0.391)	0.470	0.419	(0.364)	0.432	0,025	0,173
Terrain ruggedness index (km)	1.640***	(0.281)	1.690***	1.705***	(0.291)	1.754***	1.640***	(0.296)	1.690***	0,006	0,008
Elevation range (km)	-0.656**	(0.282)	-0.676**	-0.764**	(0.288)	-0.786**	-0.656**	(0.287)	-0.676**	0,351	0,327
Road density (km/pop)	0.032	(0.170)	0.033	0.045	(0.128)	0.048	0.032	(0.133)	0.033	0,006	0,008
Distance to road (km)	0.214	(0.594)	0.220	0.299	(0.532)	0.308	0.214	(0.533)	0.220	8,228	20,171
Distance to central city (km)	-1.634***	(0.185)	-1.683***	-1.663***	(0.301)	-1.711***	-1.634***	(0.305)	-1.683***	3,728	10,546
Left	-0.704**	(0.343)	-0.725**	-0.711**	(0.334)	-0.732**	-0.704**	(0.340)	-0.725**	0,475	0,499
Constant	67.293***	(4.241)		68.434***	(5.125)		67.293***	(5.307)			
Wald test that rho and lambda are both zero:				167.3 (p-val: 2.8715e-38)							

The dependent variables is the sprawl index, which has mean 85.79 and standard deviation 16.06. Coefficients give the impact on the index of one-standard-deviation increase in the corresponding variable. Column (1) reports the spatial lag results (S2SLS with heteroskedastic innovations of unknown form). Column (3) reports the S2SLS results with Spatial HAC standard errors for the specification with spatial lag and spatial error dependence (It assumes a very general form for the disturbance process). Column (5) reports S2SLS results with Spatial HAC standard errors for the specification with spatial lag and spatial error dependence (AR process assumed for disturbances). Columns (2), (4) and (6) report the corresponding direct impacts. Numbers in brackets report heteroskedastic-consistent standard errors (HAC standard errors in Columns (3) and (5)). ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

increase in the distance from suburban residential developments to the centre of the urban area diminishes the intensity of urban sprawl.

Note that our results are robust to a variety of changes to the specification in addition to those discussed throughout this section (i.e. different weighting schemes and model specification). Our regressions include all municipalities regardless of size. If we include the initial population of each location in our specification, this variable is not significant and the rest of our results are not affected. Similarly, the inclusion of other insignificant variables, such as a beach dummy or distance to the coast, do not change the robustness of the results reported here.

We also run the equation with an extended sample of municipalities, i.e. both urban and rural areas, and excluding the central city of each urban area. The results obtained were quite similar to those reported in the paper.

Finally, we considered the effect of local public finance on urban sprawl. To do so, we included two additional variables to our specification (%property tax and %transfers) with the consequent reduction in the sample size (see Section 4). Nevertheless, our main results hold.

7. Conclusions

The spatial econometrics literature emphasizes the importance of spatial-based interactions in regression analysis, as spatial units are not independent. There is extensive theoretical and empirical literature related to urban sprawl. Previous research has, however, considered municipalities as isolated entities and the spatial correlation in the levels of urban sprawl between neighbouring jurisdictions has never been analysed.

This paper aims to fill this gap. Our main objective is to empirically asses the presence of local spatial interaction on urban development patterns of Spanish urban areas and whether this interaction might be due to the strategic behaviour of neighbouring local governments. Spain constitutes a good testing ground for our hypothesis, as land use policies are basically a local responsibility and most of the land-use conversion experienced during the nineties took the form of scattered, spatially expansive urban growth. In such a decentralised setting, land-use decision-making of neighbouring jurisdictions might exhibit a certain degree of correlation as local incumbents do not make policy decisions in isolation. The structure of local public finance along with housing and land-use policies make land a valuable commodity, creating the economic incentives for local governments (in terms of higher incomes and tax revenues) that use their control over land allocation to influence development patterns and even engage in strategic interaction with neighbouring jurisdictions.

The empirical research is conducted on a cross-section of 3,131 municipalities, covering almost all Spanish urban areas, for the period 1990-2000. As to the sprawl variable, we follow the pioneering work of Burchfield et al. (2006), and we use GIS techniques to define sprawl as the percentage of open space in the square kilometre surrounding an average residential development. In addition, we compile a database of variables on land use and topographic features, complemented with additional information on weather conditions, social, demographic, political and economic variables which are then used in a spatial regression model.

The empirical analysis shows that there exists spatial interaction in the levels of urban sprawl between nearby locations. A local government's level of urban sprawl is positively influenced by the degree of urban sprawl in neighbouring jurisdictions, with other causal factors remaining constant. There is not a generally accepted strategy in the spatial econometrics literature to determine the model specification. According to the most recent papers on the topic, we have performed our analysis according to different approaches, yielding similar results. The existence of a spatial lag parameter was justified on the grounds of economic theory and supported by econometric strategy. Moreover, the magnitude of the estimated parameter is consistent across different specifications, confirming the robustness of our result.

According to urban economics theory, urban development patterns are also influenced by a myriad of factors. Our results show that this is indeed the case. Car-based living, population growth, the structure of employment, certain socio-economic characteristics, and the flight from blight are driving forces of urban sprawl. Amenity variables (in terms of climate and open space preservation) also play an important role. Our results are consistent in the main with those for the U.S. case (see, for example, Wassmer, 2008; Burchfield et al, 2006). In addition, physical geography remains a key element in explaining sprawl and, as in its U.S. counterpart, it explains about 25 percent of the variation in the sprawl index.

The location of suburban development within an urban area is perhaps one of the most important particularities of many Southern European cities, compared to the North-American urban context. The regression findings indicate that European suburban development is characterised by occurring as the distance to the CBD diminishes. This result is in line with previous literature that highlights the importance of the existing urban fabric in the sprawl processes of Southern European cities, where proximity to the metropolitan urban core is crucial.

Urban development patterns differ among regions and it is important to identify such differences. In addition, a better understanding of the drivers of this phenomenon in each particular region is crucial for formulating efficient public policies. This paper has contributed additional empirical evidence of the causes of sprawl while providing new evidence for a European country, thus enabling comparison with previous studies for US cities.

8. References

Arribas-Bel, D.; Nijkamp, P.; and Scholten, H. (2011) Multidimensional urban sprawl in Europe: a self-organizing map approach. *Computers, Environment and Urban Systems* 35 (4), 263-275.

Battha, B.; Saraswati, S.; and Bandyopadhyay, D. (2010) Urban sprawl measurement from remote sensing data. *Applied Geography* 30, 731-740.

Baum-Snow, N. (2007): Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122 (2), pp.775-805.

Boix, R..; Veneri P.; Almenar V. (2012), "Polycentric metropolitan areas in Europe: towards a unified proposal of delimitation", in Esteban Fernández Vázquez and Fernando Rubiera Morollón (eds.), *Defining the Spatial Scale in Modern Economic Analysis, p.45-70.* Springer. *ISBN 978-3-642-31993-8.*

Brueckner, J. K. 1998. Testing for strategic interaction among local governments: The case of growth controls. *Journal of Urban Economics* 44: 438-67.

Brueckner, J.K. (2001): Urban Sprawl: Lessons from Urban Economics, en Gale, W.G., Pack, J.R. (Eds), *Brookings-Wharton Papers on Urban Affairs*, Brookings Institution, Washington, D.C., p.65-89.

Brueckner, J. K. (2003) Strategic interaction among local governments: an overview of empirical studies. *International Regional Science Review* 26 (2), 175-188.

Brueckner, J.K.; and Fansler, D.A. (1983): The economics of urban sprawl: Theory and evidence on the spatial sizes of cities, *The Review of Economics and Statistics*, volume 65 (3), p.479-482.

Brueckner, Jan K. and Robert W. Helsley. 2012. "Sprawl and blight." *Journal of Urban Economics*, 69, 205-213.

Brueckner, J.K., and Kim, H. (2003): Urban Sprawl and the Property Tax, *International Tax and Public Finance*, volume 10, p.5-23.

Brueckner, J.;Thisse J.; and Zenou, Y. (2000) Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review* 43, 91-107.

Bruegmann R. (2005) *Sprawl: A Compact History*, University of Chicago Press: Chicago.

Buettner, T.; Ebertz, A. (2012) Urban sprawl and interjurisdictional competition.

Burchfield, M.; Overman, H.; Puga, D.; Turner, M. (2006): Causes of sprawl: a portrait from space, *Quarterly Journal of Economics*, pp.587-633.

Carruthers, J.I., Ulfarsson, G.F. (2003) "Urban sprawl and the cost of public services" *Environment and Planning B: Planning and Design,* 30, 503-522.

Catalán, B.; Saurí, D.; Serra, P. (2008) "Urban sprawl in the Mediterranean? Patterns of growth and change in the Barcelona Metropolitan Region 1992-2000" *Landscape and Urban Planning,* 85, 174-184.

Cullen, Julie B. and Steven D. Levitt. 1999. "Crime, urban flight and the consequences for cities." *Review of Economics and statistics,* 81, 159-169.

Downs, Anthony. 1999. "Some realities about Sprawl and Urban Decline." *Housing Policy Debate*, 10 (4), 955-974.

Duranton, G., and Turner, M. (2011) The fundamental law of road congestion: evidence from US cities. *American Economic Review* 101, 2616-2652.

EEA (2006) Urban sprawl in Europe. The ignored challenge. *European Environmental Agency Report* 10/2006.

EEA (2010) The European Environment. State and Outlook 2010. *European Environmental Agency*.

Elhorst, J. P. (2010): "Applied spatial econometrics: raising the bar", *Spatial Economic Analysis*, 5 (1), 9-28.

Ewing, R.; Richardson, H.W.; Bartholomew, K.; Nelson, A.; Bae, C. (2014). Compactness vs. Sprawl revisited: converging views. CESifo working paper 4571.

Fingleton, B.; LeGallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science* 87 (3), 319-339.

Frenkel, A. and Ashkenazi, M. (2008). Measuring urban sprawl: how can we deal with it? *Environment and Planning B: Planning and Design*, 35(1):56.

Galster, G., Hanson, R., Ratcliffe, M., Wolman, H., Coleman, S., and Freihage, J. (2001). Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept. *Housing Policy Debate* Washington, 12(4):681-718.

García-López, M. A. (2012) Urban spatial structure, suburbanization and transportation in Barcelona. *Journal of Urban Economics* 72, 176-190.

Gibbons, S.; Overman, H. (2012). Mostly pointless spatial econometrics. *Journal of Regional Science*, 52 (2), 172-191.

Glaeser, E.L., Kahn, M.E. (2004): Sprawl and Urban Growth, in Henderson, J.V., Thisse, J.F. (Eds), *Handbook of Urban and Regional Economics*, volume IV, Elsevier, Amsterdam, p.2498-2527.

Glaeser, E.L.; Kolko, J.; Albert Saiz, A. 2001. "Consumer city," *Journal of Economic Geography*, 1(1), 27-50.

Gordon, P.; Richardson, H. W. (2000). Critiquing sprawl's critics. *Policy Analysis*, 365, 1–18.

Hayden D. and Wark J. (2006), A Field Guide to Sprawl, W. W. Norton & Company: 7-8)

Heubeck, S. (2009) Competitive sprawl. *Economic Theory*, 39, 443-460.

Hoover, K., Perez, S., (2004), Truth and robustness in cross-country growth regressions. *Oxford Bulletin of Economics and Statistics* 66, 765–798.

Hortas-Rico and Solé-Ollé (2010) Does urban sprawl increase the costs of providing local public services? Evidence from Spanish municipalities. *Urban Studies* 47(7), 1513-1540.

Hortas-Rico, M. (2013) Urban sprawl and municipal budgets in Spain: a dynamic panel data approach. *Papers in Regional Science* (forthcoming).

Hortas-Rico, M. (2014) Sprawl, Blight and the role of Urban Containment Policies. Evidence from U.S. cities. *Journal of Regional Science* (forthcoming).

Jaeger, J.; Bertiller, R.; Schwick, C.; and Kienast, F. (2010) Suitability criteria for measures of urban sprawl. *Ecological Indicators*, 10, 397-406.

Kelejian HH, Prucha IR. 1998. A generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* 17: 99-121.

Kelejian HH, Prucha IR (2010), Specification and Estimation of Spatial autoregressive Models with Autoregressive and Heteroscedastic Disturbances." *Journal of Econometric*s 157(1), 53–67.

LeSage, J.P.; Pace, R.K. (2009): *Introduction to spatial econometrics,* Boca Raton, Taylor & Francis.

Manski, C. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60: 531-42.

McGrath, Daniel T. 2005. “More evidence on the spatial scale of cities.” *Journal of Urban Economics*, 58, 1-10.

Mieszkowski, P., Mills, E.S. (1993): The causes of Metropolitan Suburbanization, *Journal of Economic Perspectives*, volume 7 (3), p.135-147.

Mills, Edwin S. and Richard Price. 1984. “Metropolitan suburbanization and central city problems.” *Journal of Urban Economics,* 15, 1-17.

Mur, J.; Angulo, A. (2009). Model selection strategies in a spatial setting: Some additional results. *Regional Science and Urban Economics*, 39, 200-213.

Nechyba, T.; Walsh, R. (2004): Urban Sprawl, *Journal of Economic Perspectives*, volume 18, (4), p.177-200.

OSE (2006) Cambios de ocupación del suelo en España. Implicaciones para la sostenibilidad. *Observatorio de la sostenibilidad en España. Instituto Geográfico Nacional. Ministerio de Fomento.*

Pace, LeSage and Zhu (2013) Interpretation and Computation of Estimates from Regression Models using Spatial Filtering," *Spatial Economic Analysis*, vol. 8(3), p. 352-369, September

Piras, G. (2010). sphet: Spatial Models with Heteroskedastic Innovations in R. *Journal of Statistical Software* 35 (1), 1-21.

Saiz, A. (2010): The geographic determinants of housing supply, *Quarterly Journal of Economics,* 125 (3), pp.1253-1296.

Sierra Club (1999). *The dark side of the American Dream: The costs and consequences of suburban sprawl.* San Francisco, CA. (http://www.sierraclub.org).

Slack E (2002) Municipal finance and the pattern of urban growth. *C.D. Howe Institute Commentary. The Urban Papers* 160: 1-25

Song, Y.; Knaap, G. (2004) Measuring urban form, *Journal of the American Planning Association*, 70 (2), 210-225.

Song, Y.; Zenou, Y. (2006) Property tax and urban sprawl. Theory and implications for US cities. *Journal of Urban Economics* 60: 519–534.

Schwarz, N. (2010) “Urban form revisited: Selecting indicators for characterizing European cities.”, *Landscape and Urban Planning*, 96, 29.47.

Torrens, P. (2008). A Toolkit for Measuring Sprawl. *Applied Spatial Analysis and Policy*, 1(1):5{36.

Turner, M. (2005) “Landscape preferences and patterns of residential location”, *Journal of Urban Economics*, 57, 19-54.

Tsai, Y. (2005) Quantifying Urban Form: Compactness versus ‘Sprawl’, *Urban Studies* 42 (1), 141-161.

Wassmer, R. W. (2008), Causes of Urban Sprawl in the United States: Auto reliance as compared to natural evolution, flight from blight, and local revenue reliance. *Journal of Policy Analysis and Management*, 27: 536–555.

Wilson, J. D., (1999), Theories of tax competition, *National Tax Journal* 52, 269{304.

Wolman, H.; Galster, G.; Hanson, R.; Ratcliffe, M.; Furdell, K.; and Sarzynski, A. (2005), The Fundamental Challenge in Measuring Sprawl: Which Land Should Be Considered? *The Professional Geographer*, 57(1), 94–105.

Wu, J. (2006): Environmental amenities, urban sprawl and community characteristics, *Journal of Environmental Economics and Management* 52, pp. 527-547.

Appendix. Table A1. The determinants of urban sprawl. OLS and ML estimation. N=3,131

	OLS		Spatial lag ML			Spatial Durbin ML		
	(1) Estimated parameters		(2) Estimated parameters		(3) Direct impacts	(4) Estimated parameters		(5) Direct impacts
Spatial dependence:								
Rho	--		0.355***	(0.015)		0.405***	(0.020)	
Control variables:								
Vehicles per household	2.223***	(0.299)	2.454***	(0.214)	2.545***	2.868***	(0.259)	2.880***
Population growth 1930-1990	-5.767***	(1.120)	-4.286***	(0.216)	-4.446***	-3.776***	(0.239)	-4.106***
% Aged 25-45	-0.096	(0.264)	-0.053	(0.251)	-0.056	-0.055	(0.250)	-0.081
% Aged > 65	2.543***	(0.506)	1.503***	(0.272)	1.559***	2.069***	(0.304)	2.106***
Children family	0.452**	(0.148)	0.334**	(0.174)	0.347*	0.196	(0.196)	0.231
% Graduate	-0.562*	(0.249)	-0.707***	(0.205)	-0.733***	-0.707***	(0.207)	-0.703***
Poverty (% without studies)	-0.498**	(0.162)	-0.109**	(0.189)	-0.113	-0.223	(0.207)	-0.256
% Immigrants	-0.734*	(0.285)	-0.400*	(0.179)	-0.41**	-0.213	(0.184)	-0.282
% Employed in manufacturing	-3.047***	(0.319)	-2.282***	(0.192)	-2.367***	-1.627***	(0.251)	-1.756***
% Employed in retail	-1.981***	(0.232)	-1.684***	(0.180)	-1.747***	-1.632***	(0.181)	-1.712***
% Employed in other services	-2.835***	(0.287)	-2.336***	(0.215)	-2.424***	-2.461***	(0.219)	-2.448***
Mean max temp	-1.342***	(0.247)	-0.482**	(0.223)	-0.500**	-0.762	(0.925)	-0.757
Mean min temp	-0.681***	(0.206)	-0.648***	(0.189)	-0.673***	-1.139	(0.866)	-1.142
Average precipitation	-1.299***	(0.254)	-0.746***	(0.195)	-0.774***	-0.862	(1.062)	-0.853
% Open space	0.751**	(0.246)	0.740**	(0.239)	0.768**	2.197***	(0.375)	2.002***
% Leisure and sports facilities	0.096	(0.311)	0.137	(0.243)	0.142	0.323	(0.251)	0.210
% Water availability	-1.421***	(0.220)	-1.196***	(0.178)	-1.241***	-1.188***	(0.189)	-1.211***
% Inland water	0.568	(0.493)	0.405	(0.248)	0.420*	0.625**	(0.276)	0.557**
Terrain ruggedness index (km)	1.767***	(0.314)	1.628***	(0.279)	1.689***	1.904***	(0.418)	1.871***
Elevation range (km)	-0.290	(0.300)	-0.690**	(0.278)	-0.716**	-1.657***	(0.338)	-1.452***
Road density (km/pop)	0.107	(0.165)	0.024	(0.169)	0.025	0.181	(0.172)	0.123
Distance to road (km)	0.888	(0.565)	0.150	(0.589)	0.156	3.494**	(1.170)	3.097**
Distance to central city (km)	-1.845***	(0.335)	-1.614***	(0.184)	-1.674***	-1.657***	(0.185)	-1.682***
Left	-1.135**	(0.363)	-0.663**	(0.339)	-0.688**	-0.956**	(0.347)	-0.682**
Lag. Vehicles per household						-1.067***	(0.388)	
Lag. Population growth 1930-1990						-1.131***	(0.414)	
Lag. % Aged 25-45						0.194	(0.481)	

Appendix. Table A1 (continued)

Lag. % Aged > 65					-0.538	(0.496)
Lag. Children family					0.206	(0.318)
Lag. % Graduate					0.319	(0.379)
Lag. Poverty (% without studies)					-0.178	(0.328)
Lag. % Immigrants					-0.471	(0.327)
Lag. % Employed in manufacturing					-0.376	(0.352)
Lag. % Employed in retail					-0.024	(0.342)
Lag. % Employed in other services					-1.109**	(0.369)
Lag. Mean max temp					0.351	(0.955)
Lag. Mean min temp					0.442	(0.885)
Lag. Average precipitation					0.418	(1.081)
Lag. % Open space					-2.465***	(0.478)
Lag. % Leisure and sports facilities					-1.045**	(0.464)
Lag. % Water availability					0.299	(0.267)
Lag. % Inland water					0.798**	(0.366)
Lag. Terrain ruggedness index (km)					-1.035*	(0.533)
Lag. Elevation range (km)					2.324***	(0.443)
Lag. Road density (km/pop)					-0.545	(0.336)
Lag. Distance to road (km)					-4.595**	(1.404)
Lag. Distance to central city (km)					0.477*	(0.247)
Lag. Left					0.385	(0.586)
Constant	104.859***	(3.632)	63.763***	(3.868)	56.861***	(5.215)
Robust LM lag	170.513 (p-value: 2.2e-16)					
Robust LM error	18.814 (p-value: 1.441e-05)					
LR test for rho			497.37 (p-value: 2.22e-16)		360.55 (p-value: 2.22e-16)	
Log Likelihood / AIC			-11,228.93 / 22,512		-11,157.38 / 22,417	
LM test for residual autocorrelation			2.4732 (p-value: 0.1158)		46.001 (p-value: 1.182e-11)	

The dependent variables is the sprawl index, which has mean 85.79 and standard deviation 16.06. Coefficients give the impact on the index of one-standard-deviation increase in the corresponding variable. Column (1) reports the OLS results (heteroskedastic-consistent standard errors in brackets). Column (2) reports Maximum Likelihood results for the spatial lag specification (direct impacts shown in Column (3)). Column (4) reports Maximum Likelihood results for the spatial Durbin specification with spatial lag and spatial dependence in the explanatory variables (direct impacts provided in Column (5)). ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.