UNIVERSITAT DE BARCELONA

Essays on Urban Economics

María Sánchez Vidal

2016

PhD in Economics | María Sánchez Vidal

UNIVERSITAT DE BARCELONA

PhD in Economics

Essays on Urban Economics

María Sánchez Vidal

PhD in Economics

Thesis title:

Essays on Urban Economics

PhD student:

María Sánchez Vidal

Advisors:

Jordi Jofre Monseny

Elisabet Viladecans Marsal

Date:

January 2016

UNIVERSITAT DE BARCELONA

A l'avi i la mamà

Acknowledgements

I would certainly not have been able to write this thesis without the assistance and support of many people. I would like to express my gratitude to all of them for providing me with the best possible academic and personal support.

First and foremost, I would like to thank my supervisors Jordi Jofre and Elisabet Viladecans. I will always be grateful to them for teaching me and generously supplying me with advice as I have undertaken this PhD. Their dedication and insightful comments have contributed greatly to the writing of this thesis. I also wish to thank them for their patience and support, above and beyond the merely academic side of this process. My gratitude is also extended to Rafael González, with whom I had the pleasure of collaborating in the writing of the first chapter of this thesis.

I am also grateful to both the Barcelona Institute of Economics (IEB) and the Department of Public Policy, Political Economy and Spanish Economy for funding my research and the conferences I have attended over these years. These institutions have provided the perfect environment in which to conduct research and in which to enjoy the teaching of economics to younger generations. Within the IEB, I particularly want to thank Susana Cobos and Mª Àngels Gómez for much more than their administrative assistance. Their professional and personal support at every step in this process has been essential to its success.

Part of the work carried out during these years has been done while first visiting, and later working, in the Spatial Economics Research Centre (SERC) at the London School of Economics. I am deeply indebted to Henry Overman for giving me the opportunity to visit SERC for six months. I also want to thank Steve Gibbons and Olmo Silva for having the confidence to appoint me to the post of Research Officer at the Centre. I am very grateful to them both for their understanding, support and patience during the last few months of this project.

Since the day I started my PhD, I have had the opportunity to meet some lovely, interesting people without whom these years would not have been the same. A special mention to my classmates on the Master's program who set out on this road with me and who have followed my every step since. Among them, I especially wish to thank Marta Gonzalez for being such a lovely, caring person.

I would also like to thank Simon Planells, with whom I shared my office, for the wonderfully fun times we spent there.

During the time I spent in Barcelona, I shared my room and experiences with many great people. The list of people to acknowledge is too long but I am grateful to all of them. Two people deserve a special mention. First, Elena Costas, whose support through times both good and bad was essential to undertake this project. She made me realise that a friendship can be forged in a matter of weeks. And, second, Catarina Alvarez, who always had time to listen to me and to take me shopping when I needed someone by my side.

At the LSE I have had the opportunity to work with some very interesting and intelligent people. I am particularly grateful to Giulia Faggio and Rosa Sanchis for all their help, support, and more importantly, their friendship. Also in London, nothing would have been the same without the dinners, TV series-marathons and fun lazy Sundays shared with my flatmates Maria and Diego.

I also want to thank Elena Costas (again), Marta Curto and Paula Salinas for making me realise that academic research can have an impact beyond University. I have greatly enjoyed all the coffees shared, and meetings and skype calls in relation to this great project that we are currently developing together.

None of this would have been possible without the unconditional support of my dear school friends. A special mention to Maria Coroleu and Blanca Puigdollers, who kept asking questions about my thesis until the end, even when I came close to boring them to tears. I also want to thank Javi Reig for his understanding and support over the first years of this PhD.

Last, but not least, I wish to thank my wonderful family, Joaquín, Oti, Beatriz and Natalia for their constant encouragement. I would also like to thank my step-father Pere for his understanding and patience during the whole process. I am fully aware that living with a PhD student can sometimes be difficult. Last and by no means least, the people to whom I dedicate this thesis are my mother, Maria José, and my grandfather, Pedro. Without my mother's encouragement I would never have enrolled on a PhD. I owe them everything that I have achieved until now, and not only in my professional life. They have been my biggest fans. Unfortunately, my grandfather has not had the chance to see me finish this project but, wherever he might be, I am sure he is proud of me.

Table of contents

Chapter 1: Introduction 1

Chapter 2: Sequential city growth in the US: Does age matter? 9

1. Introduction 9
2. Data 12
3. Empirical analysis 15
4. Results 20
5. Nonparametric analysis 26
6. Conclusions 30

References 31

Chapter 3: Big plant closures and agglomeration economies 34

1. Introduction 34
2. Data 37
   - 2.1 (International relocation) plant closures 37
   - 2.2 Employment outcomes 38
3. Matching procedure 39
4. Results 42
   - 4.1 Local employment effects in the industry affected by the plant closure 43
   - 4.2 Effects on other manufacturing industries and services 48
   - 4.3 The effects of plant closures on incumbents and new entrants 49
5. Conclusions 51

References 52

Appendix 55

Chapter 4: Small shops for sale! The effects of big-box openings on grocery stores 58

1. Introduction 58
2. The institutional setting 61

3. Data and Sample 63
4. Identification strategy 67
5. Results..73
   5.1 The big-box opening effects on grocery stores closures....................73
   5.2 Heterogeneous effects of big-box opening effects on grocery stores closures 78
6. Conclusions...81
References 82

Chapter 5: Concluding Remarks 85

References 89

Chapter 1

Introduction

The urban economics literature reports that the equilibrium city size depends on the trade-off between agglomeration economies, on the one hand, and commuting and housing costs, on the other. Studies, such as those developed by Roback (1982), Fujita and Thisse (2002) and Duranton (2008), present frameworks that show that city formation is the result of this trade-off. In other words, the larger the city, the greater the benefits in terms of higher wages and productivity (see Behrens and Robert-Nicoud, 2015, and Combes and Gobillon, 2015, for full reviews on the benefits of agglomeration economies). However, at the same time, as a city becomes bigger, housing prices rise and travel times and commuting costs increase (see Duranton, 2014, for a review of all the costs associated with large cities).

In practice, big cities coexist alongside smaller cities and both are subject to processes of growth and decline. Several papers have examined the statistical patterns associated with city growth and the shape of the city size distribution. Specifically, what this literature argues is that each city has different fundamentals (in terms, for example, of productivity). Then, small multiplicative and cumulative random shocks in these fundamentals result in an equilibrium city size distribution. Eeckout (2004) and Rossi-Hansberg and Wright (2007) provide good examples of these multiplicative random growth models that require Gibrat's law to hold: that is, that the growth of a city is independent of its initial city size. Although random growth theories make this claim and some authors (including Giesen and Sudekum, 2011, and Ioannides and Skouras, 2013) indeed report that random growth holds for different countries and periods of time, others (including Black and Henderson, 2003, and Bosker et al., 2008) fail to find the same evidence.

The second chapter of this thesis studies how cities of different ages grow. More specifically, using data from US cities, it studies whether cities of different ages present parallel growth or, on the contrary, there is a mean reversion process (smaller cities grow more). The analysis is performed focusing on the role played by the new-born cities created during the twentieth century. By means of parametric and nonparametric methods two main results are obtained. The first finding is that there are differences in city growth rates according to the age of the city. In general, when a city is born it has a very high growth rate but, as the decades pass, it matures and its growth rate stabilises or even declines.

These results are very much in line with those of contemporary studies, including Desmet and Rappaport (2015) and Giesen and Sudekum (2014), which also find that there is a correlation between city age and city size and that younger cities initially grow faster and, as they become older, parallel growth (i.e. Gibrat's law) emerges. In line with Duranton and Puga (2015), the findings in this chapter also help to understand better the mechanisms and timings of urban growth based on the aforementioned trade-off and the growth driven by random shocks, that is, in the initial decades they show cities' non-random growth but as they become older parallel growth emerges.

As introduced above, one of the main drivers of city growth is the emergence of agglomeration economies as cities become larger. The benefits that accrue from these agglomeration economies are productivity gains that are capitalized into higher wages. Many studies have sought to evaluate the effects of agglomeration economies on productivity levels. For instance, Combes et al. (2011) examine the causal relationship between agglomeration economies and productivity growth. Other studies focus on agglomeration effects on wages in large cities. The most recent studies in this line are De la Roca and Puga (2012) and D'Costa and Overman (2014). Both studies find that workers in bigger cities obtain a wage premium, and that this wage premium increases with city size. A full review of the extensive theoretical and empirical literature examining the benefits of agglomeration economies can be found in Behrens and Robert-Nicoud (2015) and Combes and Gobillon (2015), respectively.

The mechanisms via which agglomeration economies increase productivity and wages have been extensively studied in the urban economics literature. The first proposal was made by Marshall (1890) who identified three main mechanisms. First, the concentration of firms in a geographical area allows firms to share input suppliers. This is known as input-output sharing. Second, a thick labour market facilitates the flows of workers across firms in the presence of firm-specific shocks and improves the matching between workers and firms. These are the labour market pooling effects. Finally, the last mechanism, known as knowledge spillovers, concerns the importance of firms' proximity in facilitating the flow of innovative ideas. More recently, Duranton and Puga

(2004) revisited these proposals and distinguished three possible effects: sharing, matching, and learning. The first implies that firms that are in close proximity to each other gain from sharing a greater variety of inputs, risks and local indivisible goods. The second refers to the improvement in the matching between firms and workers and the latter corresponds to the diffusion of knowledge. In addition, as documented by such studies as Jofre-Monseny et al. (2011), Jofre-Monseny et al. (2014), Ellison et al. (2010) and Faggio et al. (2015), the effects of agglomeration are stronger in some industries than in others and the relative importance of the mechanisms driving them also varies. However, the implication of all these mechanisms is that all the firms in an area are interconnected. Therefore, if an external shock affects one of them, it is likely to affect them all.

Arguments of this type have often been used by policy makers for offering financial incentives to big firms in order to attract them to their cities. If a big firm locates its business in a particular area, it is likely that its suppliers will also set up in the same area, thus generating input-output linkages. Indeed, input-output sharing has commonly been used to justify the bargaining processes between jurisdictions anxious to attract a new firm. One of the more recent studies to tackle this question is that of Greenstone et al. (2010), who examine the effects of million-dollar plant openings on the productivity of incumbent firms. They show that plant location is the result of a bidding process between different areas keen to attract a new plant precisely because of the alleged agglomeration benefits for the rest of the firms in the area.

The third chapter of this thesis studies the effects when a firm decides to leave the area, above all its impact on local employment. Following the same rationale as that of plant establishments, if an area's input-output linkages are strong, the closure of a big plant may lead to the disappearance of its suppliers. This means that policy makers are presumably willing to provide firm owners with financial incentives to prevent such closures. This chapter, therefore, analyses the effects of large manufacturing plant closures on local employment. Specifically, it estimates the net employment effects of the closure of 45 large manufacturing plants in Spain, which relocated abroad between 2001 and 2006. Differences-in-differences specifications are estimated in which locations that experience a closure are matched to locations with similar pre-treatment employment levels and trends.

The results show that when a big plant closes, for each job directly lost in the plant closure, between 0.3 and 0.6 jobs are actually lost in the local economy. The adjustment is concentrated in incumbent firms in the industry that suffered the closure, providing indirect evidence of labour market pooling effects. This means that a specific firm shock such as a closure, rather than causing additional job losses due to input-output linkages, represents an opportunity for laid off workers to change jobs. Additionally, no employment effects are found in the rest of the manufacturing industries or in the services sector. Thus, these findings suggest that traditional input-output analyses tend to overstate the net employment losses of large plant closures. Therefore, this chapter seems to show that agglomeration economies, in the form of labour pooling, play a role in adjusting local labour markets to external shocks.

Apart from the importance attached to cities as places providing higher productivity, a more recent strand in the literature has focused on the role of consumption amenities in cities. In fact, Glaeser et al. (2001) and Carlino and Saiz (2008) argue that the success of cities depends on their role as centres of consumption. Additionally, Shapiro (2006) reports that a city's employment growth, as well as being generated by gains in productivity, is also dependent on improvements in the quality of life. Others, such as Couture (2015), demonstrate the importance of non-tradable consumption in explaining the value of cities. A greater variety of amenities is associated by consumers with a higher quality of life. Indeed, several recent papers have focused on the identification of this revealed quality of life. Ahlfeldt (2013), Ahlfeldt and Holman (2015) and Albouy (2015) represent the latest attempts at quantifying the value of consumption amenities adopting a variety of strategies.

Some of the highest valued amenities tend to be located in city centres. For instance, Ahlfeldt and Holman (2015) identify architectural amenities in England as a proxy for quality of life and the amenities that people value most seem to be city centre monuments. However, it is not only the landscape in the city centre that is attractive. If we take European cities as our example, it is easy to identify city centres as the places in which most of the commercial activity is conducted. As such, these areas are usually considered as being the centre of a city's commercial amenities.

One of the main threats to city centre shopping areas is the opening of big-box stores, that is, stores located out-of-town, offering a huge variety of products in a one-stop shop. In line with the above argument, the opening of a big-box may harm a city centre's existing amenities and have a negative effect on the number of grocery stores that remain open. Moreover, the opening of a big-box store may also create negative externalities in the area in the form of traffic congestion or environmental pollution (Cheshire et al., 2014). However, from the perspective of productivity, big-box stores tend to push prices down and might be more productive sites.

Therefore, although it might be better for a city's productivity to open a big-box, in terms of local consumption amenities, the question of the impact on the citizens' quality of life is more controversial. As a consequence, over the past years, many European countries have implemented policies to restrict the entry of these big box stores (see Bertrand and Kramarz, 2002; Schivardi and Viviano, 2011, and Cheshire et al., 2015, for the effects of regulations of this type on retail productivity in France, Italy and the UK, respectively).

In this context, the fourth chapter of this thesis evaluates the effects of big-box openings on the closure of grocery stores at the municipality level. To estimate these effects, a discontinuity in a commercial regulation in Spain is used as the source of exogenous variation for the period 2003 to 2011. This regulation, which varies by region, establishes entry barriers on big-box stores in municipalities of less than 10,000 inhabitants. This study first tests whether there is a discontinuity of the number of big box-openings when crossing the population threshold from regulated to non-regulated areas. This first stage shows that non-regulated places experience 0.3 more big-box openings than the regulated areas. Then, using this discontinuity as an instrument to examine the effects of these openings on the number of grocery stores, it is shown that four years after the big-box opening, between 20% and 30% of the grocery stores in the municipality disappear. In addition, there do not seem to be significant differences between big-box stores opened in the city centre or in the suburbs, at least in the short run. However, when looking at the typology of big-boxes, the conventional ones (those selling well-known brands) seem to compete more with grocery stores than do the discount big-boxes (those selling their own brands at lower prices) and, thus, they are more instrumental in forcing them to close down.

References

Ahlfeldt, G. (2013): "Urbanity". Discussion Paper No' 0136. Spatial Economics Research Centre, London School of Economics.

Ahlfeldt, G. and Holman, N. (2015): "Distinctively different: A new approach to estimating architectural amenities". Discussion Paper No' 0171. Spatial Economics Research Centre, London School of Economics.

Albouy, D. (2015): "What are cities worth? Land rents, local productivity, and the total value of amenities". *Review of Economics and Statistics*, forthcoming.

Behrens, K. and Robert-Nicoud, F. (2015): "Agglomeration theory with heterogeneous agents". In Duranton, G., Henderson, J.V. and Strange, W. (Eds.), Handbook of Regional and Urban Economics, Vol.5. North-Holland, Amsterdam.

Bertrand, M. and Kramarz F. (2002): "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry," *Quarterly Journal of Economics*, 117(4), 369-413.

Black, D. and Henderson, V. (2003): "Urban Evolution in the USA", *Journal of Economic Geography*, 3, 343-372.

Bosker, M., Brakman S., Garretsen H. and Schramm, M. (2008): "A century of shocks: the evolution of the German city size distribution 1925–1999", *Regional Science and Urban Economics*, 38, 330–347.

Carlino, G.A. and Saiz, A. (2008): "City beautiful". Working Paper 08-22, Federal Reserve Bank of Philadelphia.

Cheshire, P., Hilber, C. and Kaplanis, I. (2015): "Land use regulation and productivity—land matters: evidence from a UK supermarket chain", *Journal of Economic Geography*, 15, 43–73.

Cheshire, P., Hilber, C. and Sanchis-Guarner, R. (2014): "Do English Planning Policies Make Shopping More Sustainable?", London School of Economics, mimeo.

Combes, P.P, Duranton, G. and Gobillon, L. (2011): "The Identification of Agglomeration Economies" *Journal of Economic Geography*, 11, 253-266.

Combes, P.P. and Gobillon, L. (2015): "The empirics of agglomeration economies". In Duranton, G., Henderson, J.V. and Strange, W. (Eds.), Handbook of Regional and Urban Economics, Vol.5. North-Holland, Amsterdam.

Couture, V. (2015): "Valuing the Consumption benefits of Urban Density", University of California, Berkeley, mimeo.

D'Costa, S., Overman, H. (2014): "The urban wage growth premium: sorting or learning?". *Regional Science and Urban Economics*, 48, 168–179.

De la Roca, J., and Puga, D. (2012): "Learning by working in big cities". Discussion Paper 9243, Centre for Economic Policy Research.

Desmet, K. and Rappaport, J. (2015): "The settlement of the United States, 1800 to 2000: The Long Transition to Gibrat's Law". *Journal of Urban Economics*, forthcoming.

Duranton, G. (2008): "Viewpoint: from cities to productivity and growth in developing countries". *Canadian Journal of Economics*, 41(3), 689-736.

Duranton, G. (2014): "Growing through cities in developing countries". World Bank Research Observer.

Duranton, G. and Puga, D. (2004): "Micro-foundations of urban agglomeration economies," Handbook of Regional and Urban Economics, in: J. V. Henderson & J. F. Thisse (ed.), Handbook of Regional and Urban Economics, edition 1, volume 4, chapter 48, 2063-2117.

Duranton, G. and Puga, D. (2015): "The Growth of Cities". In Aghion, P. and Durlauf, S.N. (Eds.), Handbook of Economic Growth, vol.2, Elsevier, Amsterdam, 781-853.

Eeckhout, J. (2004): "Gibrat's law for (All) Cities", *American Economic Review*, 94 (5), 1429-1451.

Ellison, G., Glaeser, E.L. and Kerr, W.R. (2010): "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100(3), 1195-1213.

Faggio, G., Silva, O. and Strange, W.C. (2015), "Heterogeneous agglomeration", *Review of Economics and Statistics*, forthcoming.

Fujita, M. and Thisse, J.F. (2002): "Economics of agglomeration: Cities, Industrial Location and Regional Growth". Cambridge University Press (Cambridge).

Giesen K. and Südekum J. (2011): "Zipf's Law for cities in the regions and the country", *Journal of Economic Geography*. 11(4), 667–686.

Giesen K. and Südekum J. (2014): "City Age and City Size", *European Economic Review*, 71,193-208.

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

Greenstone, M., Hornbeck, R. and Moretti, E., (2010): "Identifying agglomeration spillovers: evidence from winners and losers of large plant openings", *Journal of Political Economy*, 118 (3), 536-598.

Ioannides, Y. and Skouras, S. (2013): "US city size distribution: Robustly Pareto, but only in the tail". *Journal of Urban Economics*, 73(1), 18-29.

Jofre-Monseny, J., Marín-López, R. and Viladecans-Marsal, E. (2011): "The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms", *Journal of Urban Economics*, 70, 61-74.

Jofre-Monseny, J., Marín-López, R. and Viladecans-Marsal, E. (2014): "The determinants of localization and urbanization economies: Evidence from the location of new firms in Spain", *Journal of Regional Science,* vol. 54, 313-337.

Marshall, A., (1890): "Principles of Economics". Macmillan, London.

Roback, J., (1982): "Wages, rents and the quality of life", *Journal of Political Economy*, 90, 1257–1278.

Rossi-Hansberg, E. and Wright, M.L.J. (2007): "Urban structure and growth". *Review of Economic Studies*, 74(2), 597-624.

Schivardi, F. and Viviano, E. (2011): "Entry Barriers in Retail Trade," *Economic Journal*, 121(551), 145-170.

Shapiro, J. M. (2006): "Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital", *Review of Economics and Statistics*, 88(2), 324-335.

Chapter 2

Sequential city growth in the US: Does age matter?[1]

1. Introduction

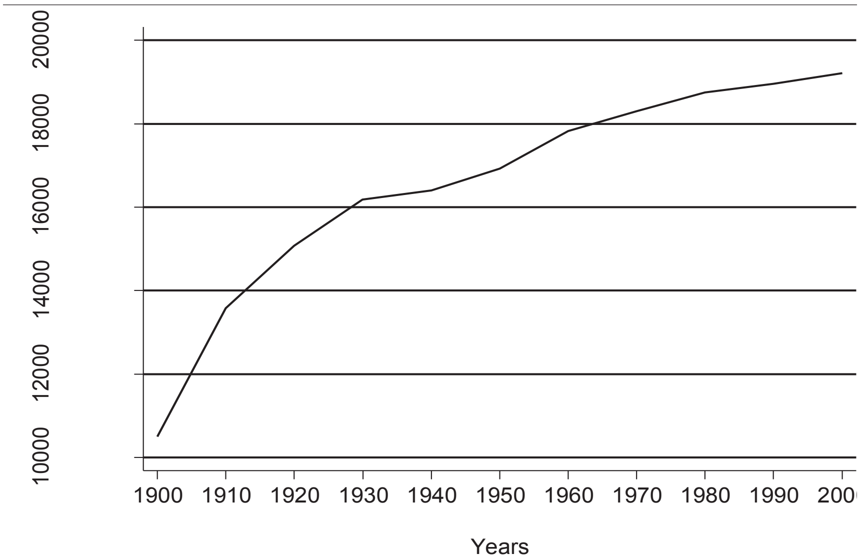
The dynamics of city size distribution and, in particular, the analysis of Gibrat's law – that a city's population growth rate is independent of its initial size – has attracted the attention of researchers for many years. In fact, there are many studies evaluating the performance of Gibrat's law for different countries and periods. Ioannides and Overman (2003) find that Gibrat's law holds for the US, Eeckhout (2004) concludes the same when including all cities without size restrictions, and so does Giesen and Südekum (2011) for the case of Germany. Others such as Black and Henderson (2003) or Bosker et al. (2008) find that this is not the case for either the US or West Germany respectively. Despite the amount of literature quantifying the size effect on growth, there is little evidence of the effect of a city's age on its growth. In this context, this paper adopts parametric and nonparametric techniques to evaluate the age-dependent patterns of urban growth using data from US cities and Metropolitan Statistical Areas (MSAs) for the period 1900 to 2000. Moreover, the non-parametric analysis provides additional empirical evidence for the above mentioned theories regarding the acceptance or rejection of Gibrat's law focusing on the role of new-born cities.

The inclusion of new cities is of special relevance for the US which saw its cities grow in number from 10,496 to 19,211 over the 20th century. At the same time, these cities increased in population and size. Figure 1 shows the evolution of the total number of US cities throughout the twentieth century as well as Figure 2 shows it for MSAs. At first glance, we can see that the number of cities grows over time but this growth is not the same throughout the period. In fact, the graph shows that this growth is concave, being higher during the first third of the century[2] and becoming more stable in the years after, while in

[1] The paper in this chapter is coauthored with Rafael González-Val and Elisabet Viladecans-Marsal. It has already been published in the *Regional Science and Urban Economics*, 44, 29-37.
[2] In fact, 62.26% of the new cities in the whole century were born in the first three decades, while the average rate of new creations for the rest of the period stands at around 5% per decade.

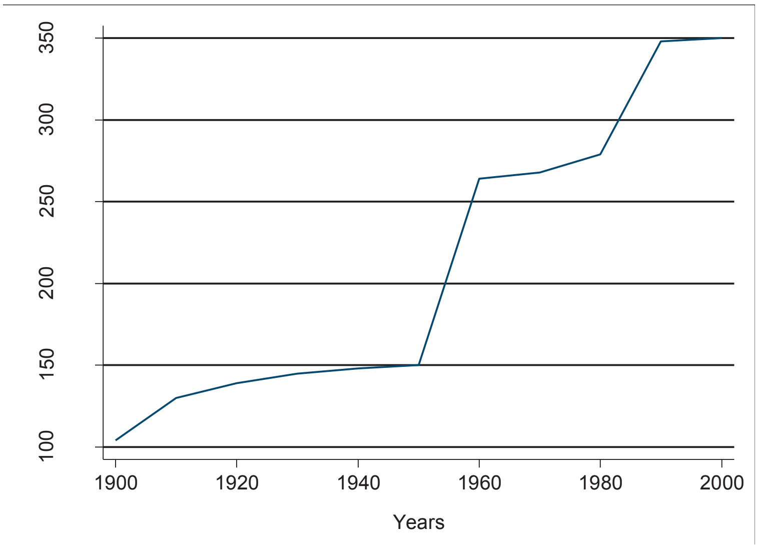
the case of MSAs we can observe less concavity. There are many examples of cities appearing during the 20th century. For instance, Long Beach in the state of New York, was incorporated in 1922, and today is the 15th biggest city in the state (the 18th in 2000). With a population of 35,462 inhabitants (2000), it enjoyed an annual growth rate of between 4.5 and 5.5% during its first three decades of existence, though this rate slowed down to 0.5% in the 1990s. The second half of the twentieth century is characterized by a suburbanization process and a proliferation of cities in the south of the country. Good examples of this phenomenon are Carson City and San Marcos, two cities in California, which are suburbs of Los Angeles and San Diego respectively. They were created during the 1960s as a consequence of the Sun Belt development. Carson City was incorporated in 1968, grew at an annual rate of 1.3% during its first decade of existence and then at a slower rate up to 2000. The case of San Marcos differs slightly. The decline in its growth rate with the passing decades was similar to that of Carson City, but its annual growth rates have been much higher: ranging from 15% on average for the first decade of its existence to 3% over the last decade, growing from a settlement of just 3,896 inhabitants in 1970 to 54,977 in 2000. These are just three examples from our dataset but there are almost 9,000 similar cases.

Figure 1. Evolution of the number of cities in the U.S. over the 20th century



Note: Data from incorporated places. Own calculations based on the censuses published by the US Census Bureau.

Figure 2. Evolution of the number of MSAs in the U.S. over the 20th century



Note: Data from MSAs. Own calculations based on the censuses published by the US Census Bureau.

However, we are not the first to analyse new cities. Previous works by Dobkins and Ioannides (2000) and Henderson and Wang (2007) also include new cities in their datasets when they cross a particular threshold. Nevertheless, the inclusion of all new cities without any threshold restriction is only considered by Giesen and Südekum (2014) who use data about the exact foundation dates of 7,000 American cities for the period 1790 to 2000, and Desmet and Rappaport (2015) whose data involves US counties and MSAs from 1800 to 2000. Our work is closely related to both studies. Giesen and Südekum (2014), by means of a theoretical model, find that the distribution of city sizes is systematically related to the country's city age distribution. They point out that young cities initially grow faster but in the long run all the cities grow at the same rate (Gibrat's law). Desmet and Rappaport (2015) argue that in earlier periods smaller counties converge and larger ones diverge but, taking into account the changes in age composition over time, both convergence and divergence dissipates and Gibrat's law gradually emerges. The results of our paper are very much in line with theirs. We find that young small cities tend to grow at faster rates but, as decades pass, their growth stabilises or even declines. Moreover, this high level of growth rates is spread across ages but is especially important in the first years of existence. After that, Gibrat's law tends to hold more firmly.

Our work thus shows a sequential growth pattern of cities according to their age. To grow sequentially means that, within a country, a few cities initially grow much faster than the rest, but at some point their growth slows and other cities start to grow in their turn, and so on. This fact has been theoretically documented by Cuberes (2009) and Henderson and Venables (2009) using theoretical models in which cities grow sequentially, allowing for the entrance of new cities to the sample in the case of Henderson and Venables (2009) or for exogenous population shocks in the case of Cuberes (2009). The only empirical approach to these theories until now has been Cuberes (2011) who, drawing on data for cities from 54 countries and on data for metropolitan areas from 115 countries, shows that urban agglomerations have followed a sequential growth pattern. This study focuses on the sequential pattern driven by the size of the city, however, while our work traces the age-dependent patterns.

Furthermore, we reproduce the analysis for metropolitan areas, the same geographical unit used in Cuberes (2011). Our results do not confirm our earlier findings for cities, however. This could reflect the fact that a metropolitan area is an aggregation of different cities; even if the area is new, the cities within it might not be. Moreover, it is not possible to know how old an area is since it does not enter the sample until it reaches the minimum population threshold of 50,000 inhabitants. As such, larger - and, therefore, more mature - cities within the area, have lower growth rates than smaller cities within the same area, and the aggregate effects may disappear.

The rest of this paper is structured as follows. Section 2 presents the data. Section 3 explains the parametric empirical methodology and Section 4 discuss its main results. Section 5 provides the nonparametric analysis and its results. Section 6 concludes.

2. Data

We use data from US cities and Metropolitan Statistical Areas (MSAs) for the whole of the 20th century. The database is the same as that used by González-Val (2010) with the addition of extra periods in the MSA dataset. The information for both geographical units was obtained from the annual census published by the US Census Bureau. A city can be defined in many different ways. Here, for our analysis, we use the word to mean 'incorporated place'. According to the census, an incorporated place is *a type of governmental unit incorporated under state law as a city, a town (except in New England, New York and Wisconsin), a borough (except in Alaska and New York city), or a village and having legally prescribed limits, powers and functions*. The Census Bureau recognises incorporated places in all American states except Hawaii, which is thus excluded from our sample. In addition, the territory of Puerto Rico and the state of Alaska are excluded as they (together with Hawaii) were not annexed until the second half of the 20th century. As Eeckhout (2004) stresses, the whole sample of cities in each state without size restrictions needs to be considered since otherwise a truncated distribution can produce biased results.

To take into account that part of the population which lives outside cities, we also use data from MSAs. This allows us to compare results of both geographical units. In line with Ioannides and Overman (2003), for the period from 1900 to 1950, we use data from Bogue (1953). This is based on the definition of Standard Metropolitan Areas (SMAs)[3] for 1950, used to reconstruct the population for the period 1900 to 1940. This means, however, that in 1900 some of the SMAs were below the 50,000-inhabitant threshold, and these are excluded until they reach that limit. For the period 1950 to 2000 our MSA data is taken from the Census Bureau.

As Glaeser and Shapiro (2003) explain, MSAs are multi-county units that capture labour markets and, as such, might serve as more effective economic units than incorporated places. Yet, the use of MSAs gives rise to a problem that is directly related to their definition: an MSA usually comprises a group of counties that requires a central city with a minimum of 50,000 inhabitants (a criterion that has changed over the period of analysis). Using MSAs represent another more specific problem for the analysis. As Dobkins and Ioannides (2001) show, the US system is characterised by the entry of new cities that can have an impact on the city size distribution. As we are particularly interested in these cities, the data for incorporated places provides more information than those on the MSAs. However, MSAs are larger geographical areas and include a large proportion of the population living in rural areas. Despite the fact that the sample of incorporated places accounts for a lower percentage of the total

---

[3] The definition of a metropolitan area was first issued in 1949 under the name Standard Metropolitan Area (SMA). It changed to Standard Metropolitan Statistical Area (SMSA) in 1959 and in 1983 was replaced by Metropolitan Statistical Area (MSA).

population, however, it is considerably more urban (94.18% in 2000) than that of the MSAs (88.35%).

Table 1 shows the descriptive statistics for the population of incorporated places in each decade of the twentieth century, while Table 2 presents the same statistics for the MSAs, the minimum threshold being 50,000 inhabitants. It is easy to see that the number of cities and MSAs increases over time as does their average size. In fact, new-born cities represent 42.52% of the total sample of incorporated places for all decades (with the average starting size for the whole sample of new-born cities being 1,294.36 inhabitants), while the number of new MSAs amounts to 180, which represents 49.85% of the sample. The number of cities in 2000 is almost twice that of 1900; the number of MSAs has increased more than threefold. This is clearly indicative of the importance of taking the appearance of new units (cities or MSAs) into consideration when studying the US population growth process. Moreover, the average size of cities and MSAs increases over time showing that the increase in urban population is even faster than the creation of new units. What these tables illustrate, therefore, is the urbanization process that the US has experienced over the last century.

Table 1. Descriptive statistics for incorporated places

Year	Cities	Mean Size	S.D.	Min	Max
1900	10,496	3,468.27	42,617.51	7	3,437,202
1910	13,577	3,610.36	50,348.78	7	4,766,883
1920	15,073	4,087.61	57,540.69	3	5,620,048
1930	16,183	4,771.31	68,462.35	1	6,930,446
1940	16,400	4,977.44	72,001.37	1	7,454,995
1950	16,923	5,662.07	76,487.59	2	7,891,957
1960	17,825	6,455.86	75,195.01	1	7,781,984
1970	18,302	7,149.50	75,690.26	4	7,895,563
1980	18,752	7,431.72	69,475.36	2	7,071,639
1990	18,953	7,998.27	72,178.75	2	7,322,564
2000	19,211	8,939.77	78,175.03	1	8,008,278

Notes: (1) Alaska, Hawaii and Puerto Rico are excluded (2) The minimum size is that provide by the US Census Bureau. We include all the cities without size restrictions to avoid truncated distributions that lead to biased results.

Table 2. Descriptive statistics for MSAs

Year	MSAs	Mean Size	S.D:	Min	Max
1900	104	280,915	586,361	52,577	5,048,750
1910	130	307,261	719,325	50,731	7,049,047
1920	139	362,905	847,072	51,284	8,490,694
1930	145	445,147	1,063,769	50,872	10,900,000
1940	148	473,984	1,125,419	51,782	11,700,000
1950	150	570,480	1,127,541	56,141	12,900,000
1960	264	477,991	1,095,872	51,616	13,000,000
1970	268	561,378	1,318,920	53,766	16,100,000
1980	279	617,269	1,455,040	57,118	18,900,000
1990	348	588,405	1,457,107	51,359	19,500,000
2000	350	658,734	1,510,498	52,457	18,300,000

Notes: (1) The minimum threshold is 50,000 inhabitants, (2) Alaska, Hawaii and Puerto Rico are excluded

3. Empirical analysis

In the context of the city size distribution analysis and, in particular, related to the sequential city growth literature, we seek to test which US cities grew the most during each decade of the 20th century, using a panel dataset. In line with this literature, we expect new-born cities to grow rapidly during the first decades of their life before stabilising (or even declining) in the decades that follow. In order to test this hypothesis, we estimate the following model:

$$g_{i,t} = \alpha + \sum_{k \geq 1} \beta_k d_{k,i,t} + \gamma \text{city size}_{i,t-1} + \delta_t + \varphi_s + \eta_r + \mu_s + \varepsilon_{i,t} \tag{1}$$

where the dependent variable g_{it} is the growth rate for each city (or MSA) *i* at time *t* calculated as $g_{it} = \ln p_{it} - \ln p_{it-1}$, being *p* the population. The variable d_k is a dummy capturing the age of the cities. The sub index *k* represents the number of decades[4] that a city is present in our sample. Therefore, d_1 (d_k when $k = 1$) is equal to one when the city is new (first decade of existence) and zero otherwise. A city is considered new when it records a positive population in one decade while having no population previously. Additionally, d_2 (d_k when $k = 2$

[4] As our data is divided by decades (not years), 1910 includes cities created from 1901 to 1910. In 1920 we find cities incorporated from 1911 to 1920. The same holds for the other eight decades of the century.

) is equal to one if the city has existed for two decades and zero otherwise, d_3 (d_k when $k=3$) is equal to one if the city has existed for three decades and zero otherwise and so on. Therefore, this dummy variable measures the age of cities, from new-born (d_1) to nine decades old (d_9), for the entire period of analysis. The variable *city size* controls for one decade lag of city size, δ_t is a time fixed effect, φ_s is a state fixed effect (in city estimations we also add county fixed effects), η_r is a region fixed effect and μ_s is a dummy capturing other location fixed effects. $\varepsilon_{i,t}$ is the error term.

Table 3 shows the evolution of the nine age dummies over the 20[th] century. For each decade, d_1 is the number of new cities created in that decade so that in 1910 a total of 3,291 new cities were born; in 1920 the number was 1,747 new cities, and so on. For each decade, d_2 is the number of cities in their second decade of existence. For instance, in 1950 there were 489 cities with two decades of existence; in 1960 there were 627 and so on. Column d_3 shows the cities in their third decade of existence, column d_4 the ones with four decades of existence, and the same until d_9. The total number of cities by age (independent of the year of their creation) for the entire twentieth century is displayed at the bottom of each column of Table 3. This number is the sample size used in the nonparametric analysis (Section 5). Moreover, we can trace the cities' evolution from the decade they first appeared until the end of the period by observing the diagonals in Table 3. In fact, if d_1 shows the number of new cities per decade, d_2 those with two decades of existence, d_3 those with three decades each decade and so on, then, we can trace the 3,291 new-born cities in 1910 by observing the number of cities corresponding to d_2 in 1920, d3 in 1930, etcetera. By construction, the numbers in the diagonal cannot increase over time; however, it becomes apparent that some cities disappeared during the century because the elements in the diagonals are not always the same. This can be explained by a variety of causes including hurricanes, the death of a town's benefactor, or the fact that when some cities expanded their borders and absorbed others. This disappearance was largely concentrated in western states and on the Great Plains, especially during the first half of the 20[th] century[5]. However, the number of cities disappearing from the sample always represents less than 3% of the total number of US cities (1,667 disappearances throughout the whole period), even in the first half of the century. In other words, if we calculate the decade average net and gross creation rates of cities over the 20[th],

[5] See Blanchard (1960) for further discussion of ghost towns in the US.

century we find that they are not very different, the decade average net creation rate of cities being 6.51% and the decade average gross creation rate of cities being 7.46%.

Table 3. Evolution of cities over the 20th century

year	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
1910	3,291	0	0	0	0	0	0	0	0
1920	1,747	3,229	0	0	0	0	0	0	0
1930	1,267	1,711	3,171	0	0	0	0	0	0
1940	505	1,245	1,684	3,132	0	0	0	0	0
1950	646	489	1,210	1,657	3,088	0	0	0	0
1960	1,046	627	470	1,164	1,614	3,025	0	0	0
1970	756	1,025	619	459	1,155	1,597	3,010	0	0
1980	553	750	1,008	612	457	1,143	1,588	2,987	0
1990	313	553	750	1,008	612	457	1,143	1,588	2,987
Total	10,124	9,629	8,912	8,032	6,926	6,222	5,741	4,575	2,987

Source: Self elaboration with US Census Bureau data

According to the sequential growth hypothesis, β_k must be positive and significant during the first decades following the birth of a city but, as the decades pass, we expect this coefficient to decrease, even possibly acquiring a negative value. However, in order to avoid any bias in these estimations, we need to add a number of controls that capture the time or space effects that might influence the results. Thus, we incorporate time and state fixed effects into our estimation. On top of the state fixed effects, for the city estimations, we include county fixed effects in order to control for a smaller geographical area.

Black and Henderson (2003) found that US cities with coastal locations grow faster, and they incorporated regional variables in their analysis so as to capture their market potential. Other studies, including Rappaport and Sachs (2001), Mitchener and McLean (2003) and Bleakley and Lin (2012), also point out that having access to navigable waters plays an important role in accounting for population distribution and growth. Thus, to control for these characteristics, we include a dummy variable that captures access to navigable waters (including access to rivers, lakes and oceans) at the state level, and four dummy variables, one for each of the major US regions: the Northeast, the Midwest, the South and the West.

Moreover, Duranton (2007) points out that cities grow or decline following gains or losses in their industries, and so we include one more control variable to capture changes in industrial composition in the US over the course of the 20^{th} century. As Kim and Margo (2004) explain, during the first half of the twentieth century, the rise of the industrial economy and the manufacturing (or 'Rust') belt saw people move westwards. Since 1950, thanks to the diffusion of air conditioning and milder winters, the population has grown in the southern part of the country, leading to the creation of the Sun Belt[6]. Thus, we include two dummies at the state level, one for each of the rust and sun belts, in order to control for these regional and industrial impacts on the population growth rate.[7]

In order to evaluate the isolated effect of city age on a city's growth and to distinguish them from those of city size on its growth, we also include a variable consisting on one decade lag city size ($\ln p_{it-1}$) as a control.[8] Then, if the age coefficient is still significant it will mean than, even after controlling for initial size, age still matters. An additional advantage of including this variable is that we are able to test the mean reversion hypothesis. When the coefficient of this variable is negative, we can assume mean reversion (convergence) in the steady state while a positive coefficient may indicate divergence (regression away from the mean). A non-significant coefficient can be interpreted as indicative of independence between growth and initial size, supporting Gibrat's law and, therefore, rejecting the mean reversion hypothesis. In the previous literature, authors such as Black and Henderson (2003) and Henderson and Wang (2007) found that the smallest cities grow faster, supporting the mean reversion hypothesis.

In our analysis, we have also introduced the age of a city, which is correlated with its size (Giesen and Südekum, 2014). Therefore, in the parametric analysis, it may be difficult to distinguish between the net effect of city age on growth from that of city size on growth. As a consequence, in order to evaluate the isolated effect of size on growth for all ages, we perform a nonparametric analysis in Section 5. We only include the city size variable in our parametric specification as a control to any possible bias. More specifically, the

[6] Other works such as Rappaport (2007) also study population mobility compared to weather conditions.

[7] Other possible drivers of cities' growth such as human capital cannot be included in our analysis due to historical data constraints.

[8] An alternative approach will be to include in the regressions interactions between the age dummies and the size variable. We estimated these regressions and the results were robust.

nonparametric analysis aims to evaluate the exact size effect on growth for every decade of a city's existence. We can therefore also examine whether we can accept the mean reversion hypothesis or reject it (being Gibrat's law the one holding in this latter case) and whether there are relevant differences in the impact of city size on growth across different ages. Moreover, the city size may, in some cases, be a source of possible endogeneity. However, our results regarding the effect of a city's age on its growth are robust whether including the city size variable or not.

We reproduce the analysis for the MSAs in order to test whether the growth pattern for the city data still applies when aggregating the geographical units in MSAs. Table 4 shows the evolution of the nine age dummies for the MSAs during the 20th century. There are two main differences between Tables 3 and 4: first, MSAs do not disappear from the sample (once an MSA reaches the minimum population threshold it never falls below it) and, secondly, the falling trend in the appearance of new MSAs is not as clear as that for the cities (as can be seen in Figure 2 as well). The former is a consequence of the MSA definition: to become an MSA a minimum population of 50,000 in the central city is required. Thus, MSAs tend to account more for larger cities with high levels of capital stock and scale externalities that make them unlikely to disappear[9]. The second distinctive characteristic is attributable to a change in the criteria used to define an MSA in 1960 (47.2% of the MSAs were created that decade).

Table 4. Evolution of MSAs over the 20th century

Year	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
1910	26	0	0	0	0	0	0	0	0
1920	9	26	0	0	0	0	0	0	0
1930	6	9	26	0	0	0	0	0	0
1940	3	6	9	26	0	0	0	0	0
1950	2	3	6	9	26	0	0	0	0
1960	114	2	3	6	9	26	0	0	0
1970	4	114	2	3	6	9	26	0	0
1980	11	4	114	2	3	6	9	26	0
1990	69	11	4	114	2	3	6	9	26
Total	244	175	164	160	46	44	41	35	26

[9] See Henderson and Wang (2007) for further explanations of why larger cities did not lose a high proportion of their population over the period 1960-2000.

4. Results

In this section we present the results of the estimation of Eq. (1). Table 5 shows the results for cities while Table 6 presents those for the MSAs. All regressions include the nine age dummy variables. The control variables are sequentially introduced from regression (1) to (6). For both geographical units (cities and MSAs), the regressions corresponding to each column represent the same specification with only the unit of analysis being changed from cities to MSA.

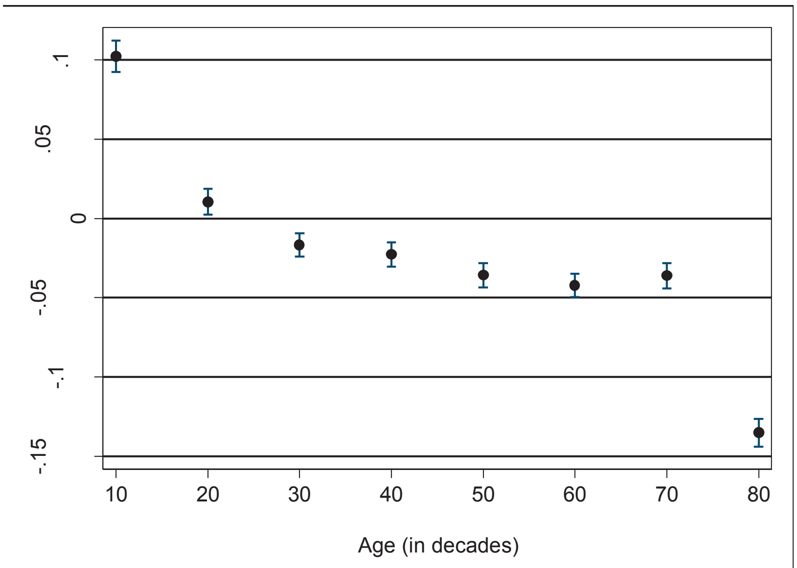
The coefficients can be interpreted as the additional average impact, measured in logarithmic points, on the growth rate of a specific city *i* (or MSA) depending on the age of that city (or MSA) compared to the average growth rate for the entire pool of cities within the period, represented by the constant of the model. As explained above, d_1 represents the city when it is newly born, d_2 when it has existed for one decade, d_3 two decades and so on, so that d_9 represents the most mature cities. Therefore, the coefficient associated with d_1 represents the additional average impact on growth of being a new-born city with respect to the average growth for the entire pool of cities within the twentieth century, that associated with d_2 represents the additional average impact of two-decades old cities with respect to the rest of the cities, and so on. For that reason, we are interested in the trend presented by the coefficients from d_1 to d_9, as this represents the dynamic effects of a city's age on their growth with respect to the average growth rate for the whole period. As this period is the whole of the 20th century, this average growth rate can be interpreted as the long term growth rate, while the coefficients of our dummy variables show the dynamics of growth.

Table 5 presents the results for cities where the total number of observations corresponds to the sum of all the cities that grow during the 20th century. Column (1) presents the results of estimating Eq. (1) by OLS without any control variable. Column (2) shows the same estimation but including the city size control. The coefficients of the nine dummies in both specifications follow the expected pattern: they are significantly positive for d_1 and become smaller until they record negative values. However, note that the results from (1) and (2) might lack precision as there may well be a considerable amount of missing and uncontrolled information in these specifications. In order to alleviate possible biases, we estimate equations (3), (4), (5) and (6) using different control variables. In column (3) we estimate the same equation but taking into account the possibility that time effects might be driving some of the results.

However, the coefficients are similar to those estimated in the previous regressions. In fact, the d's are significant and decreasing in most cases.

Column (4) presents the results for the city fixed effect estimation. Here, the interpretation of the coefficients is different from those of the other five regressions. Now, the estimated parameters show how a new-born-city *i* grows in decade *t>1* in comparison with the way new-born city *i* grew in decade *t*. An analysis of the coefficients reveals that the trend followed is the same as that in the previous estimations (the coefficient associated with d_1 being higher than that associated with d_2 and so on), indicating that the growth of a new-born city is greater than that of the same city once it becomes mature. However, the overall size of the coefficients is smaller than before. In fact, the first two dummies are not significant because they are indeed the base category[10] but from d_3 to d_9 they become significantly negative. In column (5) we estimate the same model but we include state fixed effects and county fixed effects to control for the spatial dimension of the data. The results, again, present the same pattern with significantly positive coefficients associated with d_1 and a decreasing trend until d_9. It is not, in fact, a perfectly decreasing trend because with the passing decades, city growth tends to stabilise and only declines at the end of the period. This trend can be observed in Figure 3, which plots the estimated parameters in column (5) in Table 5.

Figure 3. Estimated parameters (β_k) for regression (5) in Table 5



[10] We estimated the same regression without the incumbent cities and the results were robust.

Table 5. Estimation for cities

Dependent variable: population growth at the city level						
Decades of existence	(1) OLS	(2) OLS	(3) OLS	(4) FE	(5) OLS	(6) OLS
d_1	0.142***	0.169***	0.183***	-0.079	0.111***	0.106***
	(0.021)	(0.021)	(0.021)	(0.079)	(0.010)	(0.011)
d_2	0.048**	0.07***	0.094***	-0.129	0.017**	0.023***
	(0.020)	(0.019)	(0.018)	(0.079)	(0.008)	(0.007)
d_3	0.017	0.036***	0.057***	-0.144*	-0.01	-0.004
	(0.017)	(0.015)	(0.013)	(0.079)	(0.009)	(0.008)
d_4	0.003	0.019	0.025**	-0.159**	-0.016	-0.026***
	(0.016)	(0.015)	(0.012)	(0.079)	(0.010)	(0.009)
d_5	-0.016	-0.0007	-0.001	-0.173**	-0.029***	-0.044***
	(0.003)	(0.012)	(0.010)	(0.079)	(0.007)	(0.008)
d_6	-0.025*	-0.009	0.007	-0.160**	-0.036***	-0.028***
	(0.013)	(0.012)	(0.010)	(0.079)	(0.011)	(0.008)
d_7	-0.028**	-0.013	0.009	-0.155*	-0.031***	-0.009*
	(0.011)	(0.010)	(0.010)	(0.079)	(0.011)	(0.046)
d_8	-0.096***	-0.082***	-0.007	-0.170**	-0.131***	-0.024***
	(0.013)	(0.011)	(0.010)	(0.079)	(0.008)	(0.005)
d_9	-0.02	-0.005	-0.006	-0.162**	--	--
	(0.013)	(0.012)	(0.012)	(0.079)		
Constant	0.075***	-0.106***	-0.120***	1.672***	0.086***	0.028***
	(0.008)	(0.029)	(0.029)	(0.004)	(0.028)	(0.008)
City size $_{t-1}$		0.025***	0.028***	-0.219***	0.006	0.007**
		(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
City fixed effects	No	No	No	Yes	No	No
Time effects	No	No	Yes	Yes	No	Yes
County effects	No	No	No	No	Yes	Yes
State effects	No	No	No	No	Yes	Yes
Region effects	No	No	No	No	No	Yes
Navigable waters	No	No	No	No	No	Yes
Sun & Rust Belts	No	No	No	No	No	Yes
Observations	160,292	160,292	160,292	160,292	130,836	130,836
R-squared	0.019	0.034	0.061	0.194	0.156	0.174

Notes: (1) Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1), (2) d9 is not included in regressions (4) and (5) due to collinearity problems with the county effects, (3) The number of observations varies between regressions (5) and (6) and the rest of the specifications due to the exclusion of d9 for the whole panel.

Finally, column (6) shows the results when estimating Eq. (1) including all the control variables: initial city size, time, state, county and regional fixed effects. We also include the other geographical dummy variables: access to navigable waters and belonging to the Sun or Rust belts. This regression is our preferred one. As in the previous cases, the coefficients follow the same decreasing trend allowing us to demonstrate that when a city is born, its growth is high and as the decades pass, the growth becomes more moderate and even declines. The average impact on growth of a city in the first decade of its creation is about 0.106 points higher with respect to the average growth rate of the whole century which corresponds to the estimated value of the constant in the model (0.026). One decade later, the coefficient falls significantly (from 0.106 to 0.023 points), although the impact on growth remains positive. Thus, the highest growth occurs during the first decade of a city's existence. However, if we focus on the coefficient associated with the last decade (-0.024), we see that its age impact on growth is 0.024 points lower than the average growth rate of the whole pool of cities. These results are consistent with the theories of sequential city growth, showing that in a certain decade, the new-born cities are the ones growing the most but, in the next decades their growth stabilizes and another group of cities (the new-born cities of each of these decades) start to grow at higher rates. Then, a sequential pattern is followed over the whole century.

Table 6 shows the results for MSAs, using the same specifications as in Table 5, being the total number of observations, as for the cities: the sum of all the MSAs that grow over the 20th century. In the first two columns (1) and (2) we cannot identify the same decreasing trend as the one found in the city estimations (Table 5), although these specifications might lack precision, as do the first two identified above for the cities. For this reason, we also estimate the model incorporating time fixed effects, city fixed effects and state fixed effects in columns (3), (4) and (5) respectively. None of these three regressions presents the same pattern of results as those for the cities in terms of a declining growth trend. Finally, column (6) includes, as before, state, time and region fixed effects and the geographical controls. As in the previous columns, almost none of the coefficients are statistically significant and the expected decreasing trend is neither observed. Thus, we can conclude that the MSAs do not present the same growth pattern as that presented by cities, and the aggregation of geographical units does not provide the same results.

Studies such as Cuberes (2011) and Desmet and Rappaport (2015) found that both cities and MSAs grow sequentially, which is not seen in our estimations which suggest that cities are the only ones following this sequential pattern. One reason may be that these results are sensitive to the unit and period of analysis and, although they use MSA data, their MSA definitions and the periods analysed differ from ours. Cuberes (2011) uses a worldwide dataset for many different periods and, although for the US he considers data from 1960 which is part of our period of analysis, he defines MSAs typically above a threshold which is not the same as ours (50,000 inhabitants). In the case of Desmet and Rappaport (2015), they work with a hybrid dataset of metro areas and the remaining US counties for the period 1800 to 2000, while our study starts in 1900; almost one hundred years later, and MSA growth patterns may have been different then.

A second plausible explanation for our results lies in the definition of an MSA. A metropolitan area typically comprises a group of counties with a central city with a minimum of 50,000 inhabitants and a number of other smaller places located at points in the orbit of this central city. According to the sequential growth literature, the central city (assumed to be older and therefore larger than most surrounding places) will present different growth patterns over the time period to those of other cities within the same MSA. More specifically, the central city will be more mature than the rest and its growth rate is therefore not expected to be as high. By contrast, there will be other smaller and younger cities that will grow more rapidly during the same period. As such, the final growth rate of the MSA is the average of many rates of different cities weighted by city size, with the largest city inside the metro area being the most mature in many cases. Therefore, although young cities may grow quickly, these high growth rates are not reflected in the average growth rate of the metro area, typically dominated by the central city growth rate. Finally, another plausible explanation is that in order to become an MSA, a city must have more than 50,000 inhabitants. A new MSA is therefore the consequence of reaching an arbitrary threshold and, its growth pattern at that moment might be very different than that of a new city.[11]

[11]Another explanation can be inferred from the city results. Using our estimation results, we calculate the expected time until the average entrant city (with a size of 1,294.36 inhabitants) should cross the 50,000 threshold and thus become an MSA. The results suggest that the average new born city would not cross the 50,000 inhabitants threshold for 90 years (which is the maximum time we can calculate). Therefore, it seems very clear that, on average, the cities that pass the threshold are mature and thus, there is no age impact for them.

Table 6. Estimation for MSAs

Dependent variable: population growth at the MSA level						
Decades of existence	(1) OLS	(2) OLS	(3) OLS	(4) FE	(5) OLS	(6) OLS
d_1	0.053***	0.035**	0.117***	0.007	-0.028***	0.035*
	(0.012)	(0.014)	(0.018)	(0.035)	(0.008)	(0.021)
d_2	0.067***	0.052***	0.121***	0.025	-0.042***	0.042**
	(0.015)	(0.016)	(0.020)	(0.033)	(0.015)	(0.021)
d_3	0.01	-0.001	0.082***	-0.014	-0.027*	-0.011
	(0.022)	(0.023)	(0.025)	(0.031)	(0.015)	(0.028)
d_4	0.028**	0.019	0.091***	0.003	-0.086***	-0.001
	(0.013)	(0.013)	(0.014)	(0.026)	(0.024)	(0.017)
d_5	0.180***	0.173***	0.151***	0.064*	-0.064***	0.067**
	(0.033)	(0.034)	(0.030)	(0.033)	(0.014)	(0.029)
d_6	0.003	0.001	0.069***	-0.006	0.084**	-0.011
	(0.018)	(0.019)	(0.018)	(0.023)	(0.033)	(0.019)
d_7	0.026	0.025	0.098***	0.035	-0.081***	0.019
	(0.021)	(0.022)	(0.022)	(0.025)	(0.017)	(0.022)
d_8	-0.047	-0.045	0.032	-0.021	-0.052***	-0.05
	(0.042)	(0.043)	(0.042)	(0.027)	(0.020)	(0.039)
d_9	-0.039**	-0.035*	0.053***	---	-0.122***	-0.022
	(0.017)	(0.019)	(0.018)		(0.017)	(0.018)
Constant	0.161***	0.136***	-0.043	1.722***	0.404***	0.201***
	(0.005)	(0.048)	(0.051)	(0.230)	(0.057)	(0.073)
MSA size $_{t-1}$		-0.013*	0.009	-0.138***	-0.028***	-0.002
		(0.003)	(0.009)	(0.020)	(0.004)	(0.004)
MSA fixed effects	No	No	No	Yes	No	No
Time effects	No	No	Yes	Yes	No	Yes
State effects	No	No	No	No	Yes	Yes
Region effects	No	No	No	No	No	Yes
Navigable waters	No	No	No	No	No	Yes
Sun & Rust Belts	No	No	No	No	No	Yes
Observations	1,975	1,975	1,975	1,715	1,975	1,975
R-squared	0.036	0.04	0.153	0.244	0.201	0.301

Notes: (1) Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1), (2) d9 is not included in regression (3) due to collinearity problems with the city-fixed effects, (3) The number of observations varies between regression (4) and the rest due to the exclusion of d9

Given these results about the impact of city age effect on the city growth, in order to differentiate between the direct effects of city size and those of the city's age, we conducted a nonparametric analysis, seen in Section 5. More precisely, we were interested in analysing whether there are systematic deviations from Gibrat's law for all cities at specific ages.

5. Nonparametric analysis

There are several studies that employ a nonparametric approach to evaluate the relationship between city growth and city size, examining whether Gibrat's law and mean reversion holds in the steady state. Ioannides and Overman (2003), for example, undertake such an analysis with a time-series dataset for metropolitan areas. This same methodology is adopted by Eeckhout (2004) and González-Val (2010). The former uses it to evaluate the impact of a city's size on its growth for all the cities in the US in two specific years: 1990 and 2000. González-Val (2010) uses the same database as that described here – without any size restriction – for all of the twentieth century. All three studies find that Gibrat's law holds (at least for means) for the data and periods analysed. On the other hand, Michaels *et al.* (2012) regress population growth on a full set of fixed effects for initial population density, using their self-made dataset of county subdivisions, and finding an increasing relationship between population growth and initial population density in intermediate population densities and regression to the mean for small cities.

However, our study is much more in line with Desmet and Rappaport (2015). Using data on counties and MSAs, they empirically document the relationship between the level of population and the growth rate of a city for every twenty-year period over the nineteenth and twentieth centuries. They find that, although Gibrat's law gradually emerges, it is never fully attained. We perform a similar analysis, consisting of the nonparametric estimation of city growth against city size for every decade of existence of the US cities. By doing so, we are able to differentiate between the city age and city size effects on growth, and come to conclusions about the acceptance or rejection of the mean reversion hypothesis. Moreover, we also examine whether our results are independent of the age of a city or, on the contrary, if they differ across ages.

Our nonparametric approach is conducted using the methodology developed by Ioannides and Overman (2003) and used in Eeckhout (2004) and

González-Val (2010), but differs in terms of the data we use. We include only the cities identified as new-born in each decade and estimate a pool for any possible city age, from one to nine[12]. Table 3 shows the sample size for each estimation. This means that in decade one, we include the total number of cities with one decade of existence; no matter the year in which they were created, in decade two, any city that has existed for twenty years, and so on. The regression we estimate is the following:

$$g_i = m(s_i) + \varepsilon_i,$$

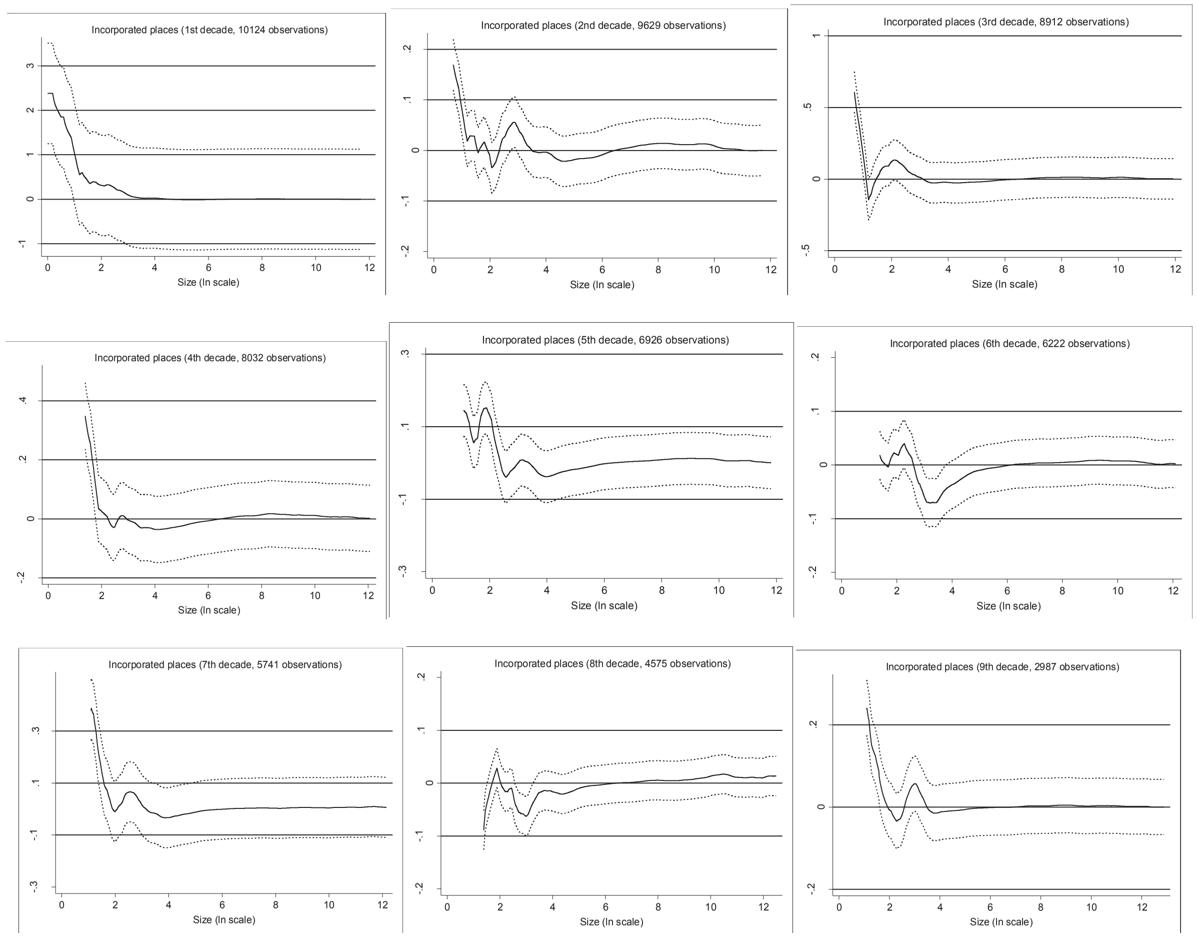
where g_i is the normalized growth rate, i.e., the difference between growth and the contemporary sample mean divided by the contemporary standard deviation, s_i is the logarithm of the population size of a city and ε_i is the error term. The aim of this approach is to provide an estimation of $m(s_i)$ without imposing any specific parametric functional form. The estimation of $m(s_i)$ is a local average that smooths the value around the point s.[13] In order to calculate the estimate we use the Nadaraya-Watson method, based on the following expression[14]:

$$\hat{m}(s) = \frac{n^{-1}\sum_{i=1}^{n} K_h(s - s_i) g_i}{n^{-1}\sum_{i=1}^{n} K_h(s - s_i)}$$

where K_h denotes the dependence of K on the bandwidth h, and where K is an Epanechnikov kernel[15]. Figure 4 shows the results for $\hat{m}(s)$ calculated for a bandwidth of $h = 0.5$[16] including only the new-born cities and considering all the possible city ages (from one to nine decades). Thus, for any age we estimate a pool of all the cities of that age along the whole period, including information from every decade in the twentieth century. The figure also displays the bootstrapped 95 per cent confidence bands, calculated using 500 random samples with replacement.

[12] We consider a city age up to nine decades (i.e., over the course of the twentieth century). To be able to consider a city with an age of ten decades, data for 1890 would be required.
[13] The smoothing is conducted using a kernel which is a symmetrical, weighted and continuous function around the point s.
[14]Employed here as in Härdle (1990).
[15] The results are robust to the use of a Gaussian kernel, as well as to the local polynomial fit technique.
[16] The results are robust to different bandwidths including the optimal one for each decade.

Figure 4. Growth and size by age

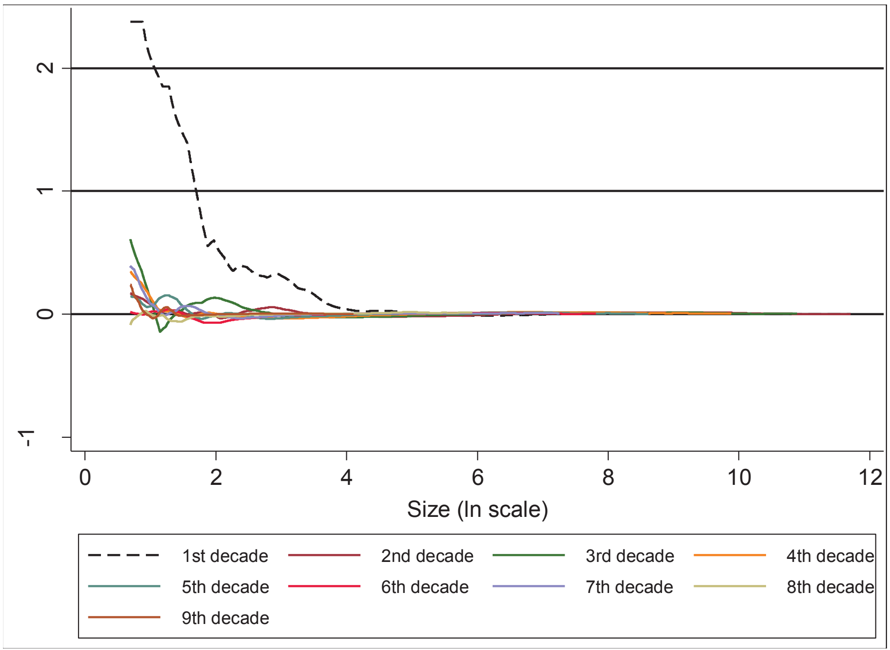


This type of analysis allows us to visually compute the temporal evolution of cities according to their size. If random growth is to be rejected, the average growth of the smallest cities should differ from that of the largest ones. If this were not the case, the figures would only present horizontal lines on the zero value of the growth axis (or random deviations that on average are around zero) and there would be no deviation from the mean. In Figure 4, it is immediately apparent that the smallest cities of all ages show higher growth rates than the rest of cities, and that the larger the city the lower its growth rate tends to be. However, as a city becomes bigger (city size increases), its average growth stabilises in the mean. Therefore, it seems that, for every decade of the 20th century, smaller cities tend to grow more.

However, if we plot all the decades together in a single graph, this conclusion can be refined. Figure 5 shows the nine different estimations of Figure 4 in a single plot. Although we can argue that there are still some differences in growth rates between the smaller cities and the rest, those differences are much larger for the youngest cities (those which are one decade old). In fact, in the figure, the dashed line corresponds to the youngest cities in the sample while the others correspond to the cities between two and nine decades old. By examining them, we clearly see that, apart from that corresponding to the youngest cities, the others look almost flat around the zero value of growth. We can therefore conclude that there are deviations from Gibrat's law for the smallest cities of all ages, and these are especially important when considering the youngest cities. In fact, as cities become older, Gibrat's law fits better[17]. Moreover, these results are consistent with our parametric findings in which we conclude that the impact of age on cities' growth is driven by their first decade of existence.

[17] From the parametric results of MSAs, which, almost by definition, represent more mature (and bigger) cities, we can clearly see in regression (6) that the initial MSA size coefficient is not statistically significant, showing that Gibrat's law fits better as cities become older (and bigger).

Figure 5. Population growth and size



This finding is in line with Giesen and Südekum (2014) which, by means of a theoretical model, find that cities grow at the same expected rate in the long run (Gibrat's law), but young cities can initially grow faster than the rest. We are also in line with the results of Desmet and Rappaport (2015) who, using different data, find that city size and growth are negatively correlated across small locations for the 19th and early 20th centuries, although Gibrat's law gradually emerges as time passes but it never completely holds.

6. Conclusions

In this paper we have drawn on data from US cities and MSAs in order to study the evolution of city growth throughout the twentieth century. More specifically, we have performed our analysis focusing on the role played by the new-born cities that have been created during the decades of our period of analysis. Using parametric and nonparametric methods we have obtained two main results. Our first finding is that there are differences in city growth rates according to the age of the city. In general, when a city is born it has a very high growth rate but, as decades pass, it matures and its growth rate stabilises or even declines.

Our second finding is related to the analysis of the dynamics of city size distribution, i.e. the study of Gibrat's law. We perform nonparametric regressions to examine the relationship between the time dimension of growth (the city's age) and the city's initial size. Our results confirm that there are deviations from Gibrat's law for the smallest cities of all ages but they are especially important for the youngest ones. In fact, as cities become older, Gibrat's law tends to hold better. Therefore, these results suggest that most of the growth differential across cities is driven by their first decade of existence, which is generally in line with our parametric results as well as with the recent literature analysing the impact of age on a city's growth.

Our results are very much in line with those presented by the city growth literature and, in particular, with those in studies of sequential city growth. Furthermore, our findings could provide interesting input for policy makers in developing countries such as China and India, which are now experiencing their own processes of urbanisation. In recent decades, both countries have experienced changes from rural to urban societies, i.e., the same pattern followed by the US in the 19th and 20th centuries. As urban policies slowly adjust to the dynamics of growth, and given the huge populations of both India and China, it must surely be in the best interests of policy makers in these countries to learn lessons from the experience of countries such as the US. In fact, if it is shown that there is a statistical regularity driving some of the population growth of cities, dependent on their initial size or age, some investment (especially in public infrastructure) could be performed strategically.

References

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

Blanchard, F.S. (1960): "Ghost towns of New England: their ups and downs", ed. Dodd, Mead & Company, New York.

Bleakley, H. and Lin, J. (2012): "Portage and path dependence", *Quarterly Journal of Economics*, 127(2), 587-644.

Bosker, M., Brakman S., Garretsen H. and Schramm, M. (2008): "A century of shocks: the evolution of the German city size distribution 1925–1999", *Regional Science and Urban Economics*, 38, 330-347.

Bogue, D. (1953): "Population Growth in Standard Metropolitan Areas: 1900-1950", Washington, DC: Housing and Home Finance Agency.

Cuberes, D. (2009): "A model of sequential city growth", *The BE Journal of Macroeconomics,* 9 (1) (Contributions), Article 18.

Cuberes, D. (2011): "Sequential City Growth: Empirical evidence", *Journal of Urban Economics*, 69, 229-239.

Desmet, K. and Rappaport, J. (2015): "The settlement of the United States, 1800 to 2000: The Long Transition to Gibrat's Law". *Journal of Urban Economics*, forthcoming.

Dobkins, L. and Ioannides, Y.M. (2000): "Dynamic evolution of the US city size distribution", in: Huriot, J.M., Thisse, J.F. (Eds.), The Economics of Cities. Cambridge University Press, Cambridge, pp. 217-260.

Dobkins, L. and Ioannides, Y.M. (2001): "Spatial Interactions among US cities: 1900-1990", *Regional Science and Urban Economics*, 31, 701-731.

Duranton, G. (2007): "Urban evolutions: the fast, the slow and the still", *American Economic Review*, 97(1), 197-221.

Eeckhout, J. (2004): "Gibrat's law for (all) cities", *American Economic Review*, 94 (5), 1429-1451.

Giesen K. and Südekum J. (2014): "City Age and City Size", *European Economic Review*, 71,193-208.

Giesen K and Südekum J. (2011): "Zipf's Law for cities in the regions and the country", *Journal of Economic Geography*, 11(4), 667-686.

Glaeser, E.L. and Shapiro, J. (2003): "Urban Growth in the 1990s: Is City Living Back?", *Journal of Regional Science*, 43 (1), 139-165.

González-Val, R. (2010): "The Evolution of the US City Size Distribution from a long term perspective (1900-2000)", *Journal of Regional Science*, 50 (5), 952-972.

Härdle, W. (1990): "Applied Nonparametric Regression". Econometric Society Monographs. Cambridge, New York and Melbourne: Cambridge University Press.

Henderson, J.V. and Venables, A.J. (2009): "Dynamics of city formation", *Review of Economic Dynamics*, 2, 233-254.

Henderson, J.V. and Wang, H.G. (2007): "Urbanization and city growth: the role of institutions", *Regional Science and Urban Economics*, 37 (3), 283-313.

Ioannides, Y.M. and Overman, H.G. (2003): "Zipf's Law for cities: An empirical examination", *Regional Science and Urban Economics*, 33, 127-137.

Kim, S. and Margo, R.A. (2004): "Historical perspectives on US economic geography", in Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Urban and Regional Economics, Vol 4. Amsterdam: Elsevier Science, North-Holland, Chapter 66, pp.2982-3019.

Michaels, G., Rauch, F. and Redding, S.J. (2012): "Urbanization and structural transformation", *Quarterly Journal of Economics*, 127(2), 535-586.

Mitchener, K. and McLean, J. (2003), The productivity of US states since 1880, *Journal of Economic Growth*, 8 (1), 73-114.

Rappaport, J. (2007): "Moving to nice weather", *Regional Science and Urban Economics*, 37(3), 375-398.

Rappaport, J. and Sachs, J. (2001): "The US as a coastal nation", *Journal of Economic Growth*, 8 (1), 5-46.

Chapter 3

Big plant closures and agglomeration economies[1]

1. Introduction

Local and regional governments around the world provide large plants with generous subsidies, often in the form of tax breaks. According to the *New York Times*, each year US local and State governments spend more than $80 billion on incentives targeted to individual firms[2]. In Europe, although government aid to firms is generally forbidden by EU legislation, national and regional governments do subsidize large plants by exploiting certain exemptions, including funds used to promote research and development, environmental protection and economic activities in lagging regions. Subsidies are frequently offered to attract new plants. For instance, Tesla Motors recently decided to locate an electric-car battery 'gigafactory' in Nevada (partly) because of a $1.25 billion tax deal. However, once a plant is operational, subsidies to avoid its relocation (or that of some of its activities) are also common. In fact, the $8.7 billion tax break that Boeing was recently offered to produce a new jet in Seattle is the largest incentive received by an individual firm in US history. In Spain, the Seat and Ford plants in Barcelona and Valencia have regularly held regional governments to 'ransom' under the threat of relocating production.

The welfare effects of subsidies targeted to individual firms are unclear (Wilson, 1999). Subsidies might cause inefficiencies if they shift plant locations to low productivity areas. However, as emphasized by Glaeser (2001) and Greenstone and Moretti (2004), subsidies can also be welfare enhancing. If the local labor supply curve slopes upward, inframarginal resident workers will gain by the presence of a large plant. In this context, subsidies can be seen as bids offered by different locations reflecting local welfare gains. A similar argument applies if large plants create significant (positive) local production externalities. Then, a subsidy will be efficient if it induces a plant to locate in an area in which the resulting local externality is especially large.

---

[1]The paper in this chapter is co-authored with Jordi Jofre-Monseny and Elisabet Viladecans-Marsal

[2]http://www.nytimes.com/interactive/2012/12/01/us/government-incentives.html

In the policy arena, the desirability of subsidies targeted to individual firms is often evaluated on a cost per job basis. An argument often made in justification of such subsidies is that large plants create employment in local supplier firms. In fact, input-output models predict (large) net employment effects of big plant openings/closures. However, the opening of a large plant might also tighten the local labor market and, thus, reduce employment in the rest of the local economy. The objective of this paper is to estimate empirically the net employment effects of large manufacturing plants. To the best of our knowledge, this is the first study to address this empirical question directly.

Specifically, we estimate the net local employment effects in Spain of the closure of 45 large manufacturing plants (median layoff of 264 jobs), which relocated abroad between 2001 and 2006. We match each municipality experiencing a closure to a small set of municipalities (four in the baseline analysis) that are very similar in terms of their 2000 employment levels. We also find that treatments and the selected controls do not differ in their pre-treatment employment trends, either. This lends empirical support to the hypothesis that the plant relocations examined here were the result of international strategies adopted by parent companies and did not respond to declining, area-specific employment trends. We run differences-in-differences specifications in which each treatment is matched to its controls by including case-specific fixed effects. The results show that when a plant closes, for each job directly lost in the plant closure, between 0.3 and 0.6 jobs are actually lost in the local economy. This is explained by local incumbent plant expansions in the industry that suffered the plant closure. We find no employment effects in the rest of manufacturing industries or in the services sectors. One implication of these findings is that they suggest traditional input-output analyses tend to overstate the net local employment losses of large plant closures. In fact, for our sample of closures, the input-output framework predicts that, for each job directly lost in the plant closure, one additional job will be lost in the local economy. Thus, in our application, the input-output prediction overestimates the negative employment consequences by an order of three. The fact that some fired workers are reemployed in local incumbent firms in the industry that suffered the closure provides indirect evidence of labor market pooling hypothesis, which states that industry concentration arises because of scale economies in the labor market[3]. Specifically, our results suggest that the

[3] Ellison *et al.* (2010), Jofre-Monseny *et al.* (2011) and Faggio *et al.* (2015) test the relative importance of labor market pooling vis-à-vis other agglomeration economies' mechanisms.

presence of same industry firms allow workers to change employers when firm specific shocks occur[4].

Fox and Murray (2004) and Edmiston (2004) study the employment effects of large plant openings in the US. Both studies conclude that such openings largely fail to create indirect jobs in the local economy. Here, our study seeks to complement these earlier reports by quantifying the effects of large plant closures. Note that the effects of openings and closures need not necessarily coincide if, for instance, a closure provides an opportunity for local incumbents to hire trained workers that have recently been laid off. Our study shows that plant closures do not, in fact, destroy indirect jobs and, moreover, that they actually generate jobs in local incumbent firms. As a consequence, the net employment effects of closures are smaller than the initial layoff itself. Greenstone *et al.* (2010) also study large plant openings in the US but focus on the impact on local productivity. In a unique empirical design, the authors use data on the subsidies offered to new plants by different local and State governments to define 'winning' counties (those attracting a plant) and 'losing' counties (those left as runners-up in the choice process). They find that the opening of a large plant increases the productivity of incumbent plants in the winning county relative to that of plants in the losing county. In line with our study, Hooker and Knetter (2001) and Poppert and Herzog (2003) estimate the local employment effects of closures but focus their attention on US military bases as opposed to manufacturing plants. They report that net employment effects are very similar to the number of jobs directly destroyed by the closure. Finally, Moretti (2010) develops a framework to estimate empirically the local impact of creating an additional job in a tradable industry on employment levels in the rest of local industries[5]. His estimates indicate that additional jobs in one part of the tradable sector have a negligible impact on jobs in other parts of the tradable sector but a large positive effect on those in the non-tradable sector, especially if these newly created positions are for skilled occupations that command higher wages. Our results can (partly) be reconciled with those reported in Moretti (2010) as net employment effects in the industry directly affected by the closure are much smaller than the closure layoffs themselves.

---

[4] Krugman (1991) formalizes this argument while Overman and Puga (2010) show that, in the UK, industries with more idiosyncratic volatility tend to be more geographically concentrated.

[5] Using this same framework, Faggio and Overman (2014) estimate the local labor market effects of public sector employment.

Following on from this introduction, the rest of the paper is organized as follows. Section 2 describes the data used throughout the paper with particular emphasis on individual plant closures. In Section 3 we explain how we select the control locations to match the areas experiencing a plant closure in terms of their respective pre-treatment employment levels. Section 4 introduces the empirical specifications used and presents the results. Finally, section 5 concludes.

2. Data

Our study examines the impact of 45 large plant closures in the manufacturing sector resulting from international relocations. In this section we first describe the characteristics and circumstances of these closures. Then, we turn our attention to the employment data sources that constitute our outcome of interest.

2.1 (International relocation) plant closures

Information on plant closures (and their corresponding job losses) is obtained by combining various data sources. Thus, we draw on information from the firms' international relocation dataset built by Myro and Fernández-Otheo (2008) and combine this with balance sheet data extracted from the *Sistema de Análisis de Balances Ibéricos* (SABI) and information obtained from newspapers and the trade unions. We restrict our attention to the 45 plant closures resulting from international relocations that occurred between 2001 and 2006 and which involved, at least, 100 job losses[6]. We exclude closures in the five largest Spanish municipalities (Madrid, Barcelona, Valencia, Seville and Zaragoza) as layoffs here are unlikely to represent a relevant shock to local employment. However, by so doing, only three closures are excluded.

For each closure, we collected the following information: firm's name, year of closure, number of workers laid off, activity (3-digit CNAE-93 classification), municipality of origin and the new country of destination[7]. Table A1, deferred to the Appendix, reports these plant-level data. Most of the closures in our dataset (49%) correspond to what the OECD classifies as medium-technology industries. The number of workers laid off ranges between 105 and 1,600, with a median of 264. In terms of their impact on the local

[6] Greenstone *et al.* (2010) examine evidence from 47 large plant openings in the US.
[7] CNAE-93 is the Spanish equivalent to the NACE classification.

economy, the layoffs represent, on average, 30 percent of local employment in the industry suffering the plant closure. In Spain, firms are among the smallest in OECD countries[8]. In fact, the average manufacturing plant employs 14 workers and, therefore, all the closures in our sample can be considered as being big[9].

The plant closures we analyze form part of international relocation processes. As Table A1 shows, most plants relocated to China or Eastern Europe. Using international relocation closures to estimate the effect of large layoffs on the local economy is helpful in terms of identification to the extent that these closures can be attributed directly to the parent companies' international strategy rather than the effects of declining local employment. As is shown below, we find no evidence that the areas experiencing closures present differential employment trends prior to the closure. Two other factors need to be borne in mind when interpreting the effects of these plant closures. First, the study period was characterized by economic growth. Between 2000 and 2008, the Spanish economy experienced an average annual growth rate of 3.1 percent; however, in the manufacturing sector, growth was much less vigorous with employment rising at an annual rate of 0.77 percent. Second, among the countries of the OECD, Spain's employment protection regulations represent some of the strictest. This holds also for collective dismissals[10]. In Spain, plant closures are accompanied by a bargaining process between the firm and trade unions mediated by the (regional) government. Anecdotal evidence suggests that deals generally involve severance payments above the (already very high) statutory level, early retirement packages and attempts by local and regional governments to re-locate workers within the local economy.

2.2 Employment outcomes

The outcome we examine is local employment at the industry level. We draw primarily on Social Security employment counts by industry and municipality. The data covers the universe of employees in Spanish municipalities at the 2-digit industry level. One caveat of this dataset is that it does not cover self-employed workers[11]. We follow employment outcomes in the period 2000 to

[8] Entrepreneurship at a Glance 2012 (OECD).
[9] Spanish Social Security for the year 2000.
[10] OECD Employment Outcome 2004.
[11] The data, in fact, exclude all workers in specific social security regimes which, in addition to the self-employed, include agricultural workers, and civil servants.

2008. Since we will study the impact of plant closures taking place between 2001 and 2006, this gives us a minimum of one pre-treatment year (2000) and two post-treatment years (2007 and 2008). Additionally, we use employment data from the 1990 Census of Establishments, which enables us to measure (and control for) local (pre-treatment) employment trends. We end the period of analysis in 2008 for two reasons. First, in 2009 the industry classification underwent a major overhaul and, second, 2008 was the last year of economic growth in Spain with output growing at 0.9 percent[12,13].

3. Matching procedure

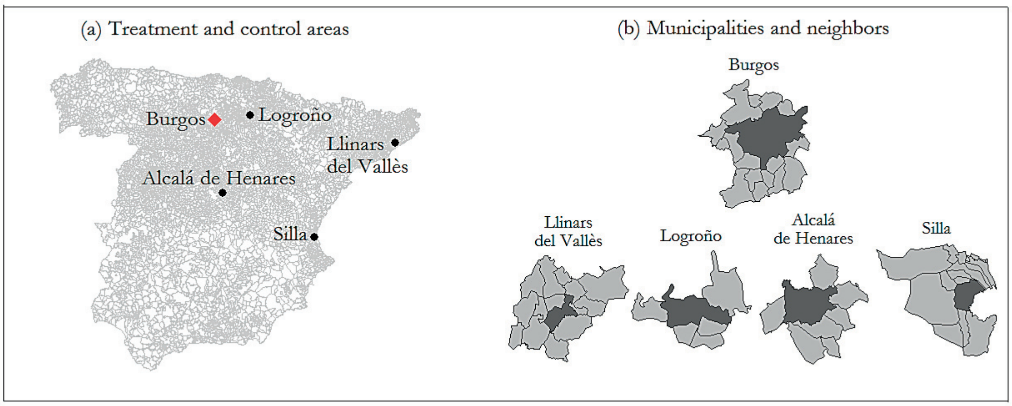
Most of the 8,122 municipalities in Spain are quite small, which suggests the impact of a plant closure might extend beyond a municipality's borders. Therefore, we construct a 10-km ring around each municipality in order to capture a municipality's immediate neighbors. This ring is built by calculating air distances between municipality centroids and the resulting area serves as our baseline geographical unit. We define a treated area as one suffering a plant closure between 2001 and 2006 and we select four appropriate controls using a matching procedure based on employment characteristics measured in 2000. Each treatment and its corresponding controls constitute what we label here as a case. Figure 1 illustrates the case of *La Cellophane Española,* a rubber and plastics plant in Burgos that closed in 2001. Panel (a) shows the geographical location of treatment and controls (*Llinars del Vallès*, *Logroño*, *Alcalá de Henares* and *Silla*). Panel (b) zooms in to show that the five areas are in fact the sum of the municipality itself (dark gray) and its neighbors lying within a 10-km ring (light gray).

---

[12] From 2009, the industry classification adopted was CNAE-2009.

[13] In 2009 there was a sharp drop in output of 3.8 percent (EUROSTAT).

Figure 1. A plant closure example: Treatment and control areas



Note: The example corresponds to Cellophane Española, a rubber and plastics plant in Burgos closing in 2001. Panel a shows the location of treatment and control areas within Spain while Panel b shows the selected municipalities (dark gray) and neighbors (light gray).

The matching procedure applied operates in two steps[14]. First, for each municipality in Spain, we compute its total level of employment in 2000 by adding to its own employment level that of its neighbors. Then, we rank the 8,122 Spanish municipalities and create six categories (<5, 5-10, 10-20, 20-50, 50-100 and >100 thousand employees). We restrict the matching procedure to municipalities within the same total employment category. Thus in the case illustrated in Figure 1, *Burgos*, *Llinars del Vallès*, *Logroño*, *Alcalá de Henares* and *Silla* have an employment level of between 50 and 100 thousand jobs, if we consider number of jobs in the municipality itself (dark gray) together with the number of jobs in the neighboring municipalities (light gray). In the second step, the target is to make treated and control areas similar in terms of employment levels in 2000 in the specific industry affected by the closure. To do so, we compute the distance for this industry between the level of employment in each potential control and each treated area. This is done in two dimensions: first, we only consider employment at the level of the municipality and, second, we add to this figure the jobs in the neighboring municipalities. Then, we compute the

[14] We do not use propensity score matching because our sample only contains plants that eventually closed due to an international relocation strategy. As such, we cannot predict where these plant closures might occur. An alternative matching procedure, and one that is more similar to the one used here, is the synthetic control algorithm, which matches pre-treatment trends in the dependent variable (see Abadie and Gardeazabal, 2003). However, this method is more appropriate for cases in which the treatment affects a large aggregate, such as a region or a country. In our case, we are able to choose our counterfactuals from a pool of more than 8,000 municipalities and so building a synthetic control is unnecessary.

following Euclidean distance $\sqrt{(I_m)^2+(I_a)^2}$, where I_m and I_a are the employment deviations in the industry affected by the plant closure at the municipality and area (municipality and neighbors) levels, respectively. Among the control municipalities whose employment level in this industry is higher than that of the treated municipality, i.e. $I_m > 0$, we select the two controls with the smallest Euclidean distance. We apply the same procedure to the control municipalities whose employment level in the affected industry is lower, i.e. $I_m < 0$. In the case illustrated in Figure 1, *Llinars del Vallès* and *Silla* are the two closest matches having higher levels of employment than Burgos in the rubber and plastics industry in 2000. Analogously, *Logroño* and *Alcalá de Henares* are the two closest matches with lower levels of employment in this industry. While we allow municipalities to be the controls for more than one treatment, we do not always find four controls for all cases. As a result, we have 217 (as opposed to 225) case-municipality observations.

In order to validate this matching procedure, we regress predetermined employment variables on a treatment indicator variable, while controlling for case fixed-effects. The results are reported in Table 1.

Table 1. Differences between treatments and controls. Pre-treatment employment levels in 1990 and 2000

	Employment in the affected industry				Overall employment	
	1990	2000	1990	2000	1990	2000
	Municipality		Area (Municipality & neighbors)		Area (Municipality & neighbors)	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatments	-60.03 (308.7)	-70.07 (264.9)	-40.55 (338.4)	-67.43 (276.1)	14,704 (20,118)	19,541 (28,205)
Case dummies	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.799	0.795	0.877	0.881	0.682	0.684
Observations	217	217	217	217	217	217

Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

The dependent variables in columns 2, 4 and 6 are the employment outcomes for the year 2000 that are directly used in the matching procedure. These results validate the matching insofar as the treated and control areas do not present statistically significant differences for any of the variables used to perform the matching. In columns 1, 3 and 5 we measure the same employment outcomes in 1990, namely, the level of employment in the affected industry at the municipality and area levels, and total employment at the area level[15]. The results indicate that employment levels in 1990 in treatments and controls were also similar, suggesting common pre-treatment employment trends. Figure 2 illustrates this point by plotting the evolution in employment in the industry suffering a plant closure for the treatment and control groups, where both time and employment levels have been normalized for the year of plant closure.

Figure 2. Employment in the plant closing industry

Note: The year and employment level are normalized to zero at the plant closure year

4. Results

Using this matched sample, we use differences-in-differences specifications to estimate the effects of big plant closures on local employment. We focus our attention primarily on the employment changes that occurred between 2000 and 2008.

[15] The 1990 employment outcomes are drawn from *Censo de Locales del INE 1990*.

4.1. Local employment effects in the industry affected by the plant closure

In this section we seek to estimate the impact of a plant closure on the employment in the industry suffering that closure. We estimate variants of the following equation:

$$\Delta employment_{ij} = a_c + \beta\, job\ losses_{ij} + X'_{ij}\delta + u_{ij} \qquad (1)$$

where $\Delta employment_{ij}$ is the job change in area i and industry j between 2000 and 2008 and, thus, u_{ij} denotes shocks in employment changes. The key explanatory variable is *job losses*, which is defined as the layoff count associated with the particular plant closure. If $|\beta|$ equals 1, then each job lost as a result of the closure translates simply as one job lost in the local industry affected by that closure. We label $|\beta|$equal to unity as 'the mechanical effect', as this is the expected outcome if the closure had zero impact on the rest of the firms in the affected industry. However, if $|\beta|> 1$, then each job lost as a result of the closure generates additional job losses in the affected industry and area. A possible mechanism accounting for such an outcome is the one often used to justify subsidies, namely, that large plants create indirect jobs through the purchase of inputs from local suppliers[16]. Alternatively, if $|\beta|< 1$, then each job lost as a result of the closure creates jobs in the local industry affected by the closure. In the presence of workers that are imperfectly mobile across locations and industries, a significant collective dismissal would reduce labor market tightness and increase employment in all other local firms. In terms of control variables, case fixed-effects (a_c) are included to account for case industry employment trends while, in some specifications, the 1990 and 2000 (pre-treatment) employment outcomes used in the matching procedure are further included (X'_{ij}) as controls. The baseline results are reported in the first two columns of Table 2.

[16] The presence of agglomeration economies would also be consistent with $|\beta|> 1$ as the productivity of local firms (and labor demand) would depend positively on local employment size.

Table 2. Impact of a plant closure on the affected industry.

	A: 2000-2008 long differences			B: 2000-2008 yearly differences		
	Industry affected by plant closure		Pooled industries	Industry affected by plant closure		Pooled industries
	(1)	(2)	(3)	(4)	(5)	(6)
Job losses	-0.521**	-0.628***	-0.556**			
	(0.228)	(0.231)	(0.227)			
Job losses (-3)				0.001	0.029	0.070
				(0.132)	(0.117)	(0.069)
Job losses (-2)				-0.025	0.000	-0.017
				(0.096)	(0.096)	(0.097)
Job losses (-1)				-0.021	0.002	-0.036
				(0.071)	(0.068)	(0.053)
Job losses (0)				-0.700***	-0.687***	-0.728***
				(0.168)	(0.178)	(0.133)
Job losses (+1)				0.046	0.059	0.072
				(0.095)	(0.09)	(0.049)
Job losses (+2)				-0.061	-0.061	-0.087
				(0.103)	(0.103)	(0.118)
Job losses (+3)				-0.087	-0.088	-0.039
				(0.064)	(0.065)	(0.064)
Case fixed-effects	Yes	Yes	No	No	No	No
Pre-treatment employment controls	No	Yes	Yes	No	Yes	Yes
Case year fixed-effects	No	No	No	Yes	Yes	No
Case industry fixed-effects	No	No	Yes	No	No	No
Case industry year fixed-effects	No	No	No	No	No	Yes
Area fixed-effects	No	No	Yes	No	No	Yes
R-squared	0.649	0.797	0.799	0.189	0.194	0.165
Observations	217	217	4,991	1,720	1,720	39,792

Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The dependent variable in columns 1 to 3 is the change in employment between 2000 and 2008 at the 2-digit industry level. The dependent variable in columns 4 to 6 are 2000-2008 yearly changes. Columns 1, 2, 4 and 5 include only the treated industry for each case while columns 3 and 6 include all manufacturing industries. Pre-treatment employment controls are the 2000 and 1990 levels at the appropriate industry level as well as in total employment. There are 23 (2-digit) industries in columns 3 and 6.

The first column shows the estimates of a specification that only includes case fixed-effects. The results imply that a job lost as a result of a large plant closure reduces employment in the affected industry and area by -0.521, implying that the closure spurs employment growth in local firms operating in the same industry and area as the closing plant. In the second column, we add the pre-treatment employment levels (X'_{ij}) to the case fixed-effects. Specifically, we include the 2000 and 1990 industry and overall employment levels. As expected, the main estimate of interest, β, is not greatly affected by the inclusion of these pre-treatment outcomes (the point estimate is -0.628) as these controls are orthogonal to treatment status as shown in Table 1. In the third column of Table 2, we estimate a slightly different model by pooling all manufacturing industries so as to account for (possible) area specific trends in employment. Here, the specifications include case industry fixed-effects and area fixed-effects. The results yield a point estimate of -0.556, confirming that when a large plant closes, employment in the rest of the firms within the local area and sector increases rather than decreases. This finding provides indirect evidence of labor market pooling effects. As first put forward by Marshall (1890), industry concentration creates scale economies by allowing workers to move between firms when idiosyncratic shocks at the firm level occur.

As discussed above, input-output analyses have often been used to predict the net employment effects of large plant openings/closures. For our sample of plant closures, a traditional input-output analysis predicts that for each job directly lost in the closure, another (indirect) job is lost in the local economy[17]. As such, our results seem to suggest that input-output analysis performs very poorly in predicting local employment responses to plant closures. Specifically, the traditional input-output analysis predicts a reduction in net employment that is three times greater (in absolute terms) than that observed.

We check the robustness of our results to the specific matching procedure adopted in two ways. First, we re-run the baseline specification selecting only the two closest controls (as opposed to four). The results, reported in columns 1 to 3 in Table A2 (deferred to the Appendix), are largely unchanged, suggesting that our findings do not hinge on the exact number of controls selected. Second, we run a placebo exercise in which we drop the actual treatment and randomly assign it to any of the four controls. The results,

[17] This is the average effect across the 45 closures using the 2005 Catalan Input–Output Table built by Statistics Catalonia (IDESCAT)

presented in columns 4 to 6, are reassuring as none of the coefficients of interest are statistically significant.

In the baseline regressions (panel A in Table 2), we focus on changes in employment in an eight-year time window. We do this as opposed to examining yearly changes for two reasons. First, (potential) anticipation effects might mean that employment falls in the year(s) prior to a plant closure. Second, the local response to a plant closure might take more than one year to take effect. To determine whether these possibilities are relevant in our application, in panel B of Table 2 we examine yearly employment changes between 2000 and 2008. In these regressions, we include the main explanatory variable (*job losses*) in the year the closure occurs as well as three lags and leads of this variable. In terms of control variables, Panels A and B adhere to the same logic, although the addition of the time dimension changes the nature of the fixed-effects that can be accounted for. Specifically, column 4 only includes case year fixed-effects while column 5 includes both these and the pre-treatment employment controls, namely, the 2000 and 1990 industry and overall employment levels. In column 6, we pool all manufacturing industries and, in addition to the pre-treatment employment controls, we introduce case industry year fixed-effects and area fixed-effects. We find no statistically significant results for any of the lag and lead variables. This finding suggests that anticipation effects are not especially relevant in our application and that the bulk of the adjustment takes place within a year of plant closure. These results are largely consistent with Figure 2 in which we show the evolution in the level of employment in the treated and control groups. However, the contemporaneous closure point estimates are slightly higher (in absolute value) than those found using 2000-2008 differences. Specifically, the point estimates using yearly variation range between -0.687 and -0.728. This is consistent with a slight recovery in employment levels in the treated areas in the years after the plant closure.

In section 3, when describing the matching procedure used, it was acknowledged that the effects of a plant closure might extend beyond the borders of a municipality. In Table 3 we explore in depth the geographical scope of the effects under study. To this end, we estimate variants of the following specification:

$$\Delta employment_{mj} = a_c + \beta_0 \, job\ losses_{mj} \, I_0 + \beta_{10} \, job\ losses_{ij} I_{10} + \gamma I_0 + X'_{mj}\delta + u_{mj} \quad (2)$$

where $\Delta employment_{mj}$ is the 2000-2008 change in the number of jobs in municipality *m* and industry *j*. Note that there are four types of municipality. Returning to the example illustrated in Figure 1, there is one treated area (*Burgos*) and four control areas (*Llinars del Vallès*, *Logroño*, *Alcalá de Henares* and *Silla*). In turn, each area comprises the municipality itself (dark gray) and the municipalities within a 10-km radius of it (light gray). Hence, we have treated municipalities, treated neighbors, untreated municipalities and untreated neighbors. I_0 indicates if the municipality itself is a treatment or a control (dark gray municipality) while I_{10} takes the value of one for the remaining municipalities within the treated and control areas (light gray municipalities). In the regressions we interact these indicators with our main explanatory variable and, thus, we estimate the employment effect in the municipality directly affected by the closure (β_0) and in the municipalities within a 10-km radius of the plant that has been closed down (β_{10}). Since the number of jobs in the plant being closed down does not form part of the neighbors' employment figures, no effects being recorded in neighboring municipalities implies $\beta_{10} = 0$. The results are presented in Table 3.

Table 3. The geographical scope of the employment effects of a big plant closure. 2000-2008 long differences.

	Industry affected by plant closure		Pooled industries
	(1)	(2)	(3)
Job losses in own municipality (β_0)	-0.800***	-0.515***	-0.634***
	(0.140)	(0.122)	(0.121)
Job losses in neighboring municipality (β_{10})	0.023	-0.018	-0.01
	(0.024)	(0.021)	(0.021)
Case fixed-effects	Yes	Yes	No
I_0 indicator	Yes	Yes	Yes
Pre-treatment employment controls	No	Yes	Yes
Case industry fixed-effects	No	No	Yes
Area fixed-effects	No	No	Yes
R-squared	0.14	0.454	0.491
Observations	2,514	2,514	57,822

Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The dependent variable is the change in employment between 2000 and 2008 at the industry and municipality level. I_0 as defined in the text. Columns 1 and 2 include only the treated industry for each case, while column 3 includes all manufacturing industries in each municipality. Pre-treatment employment controls are the 2000 and 1990 levels at the appropriate industry level as well as in total employment at the municipality level. There are 23 (2-digit) industries in columns 3.

Here again column 1 only includes case fixed-effects and the indicator variable I_0. Column 2 additionally includes, as controls, 1990 and 2000 (pre-treatment) employment levels measured here at the municipality level. Finally, column 3 pools the data from all manufacturing industries. We find no evidence that the effects of a big plant closure extend beyond the municipality in which the closure has occurred. Hence, our finding that plant closures spur employment growth in local firms operating in the same industry and area is driven solely by the behavior of firms located in the same municipality as that which has suffered the plant closure[18].

4.2 Effects on other manufacturing industries and services

According to input-output predictions, a plant closure has a negative impact on the employment in other industries. To determine whether this prediction is supported by the data, in columns 1 and 2 of Table 4 we evaluate the effects of plant closures on employment in manufacturing industries (excluding for each case, the industry directly affected by the closure). Analogously, we test in columns 3 and 4 whether the layoffs caused by the plant closure reduce employment in the services sector. The results are reported in Table 4.

Table 4. Impact of a plant closure on other industries 2000-2008

	Other manufacturing industries		Services	
	(1)	(2)	(3)	(4)
Job losses	0.111 (0.089)	-0.003 (0.008)	0.000 (0.000)	0.001 (0.003)
Case industry fixed-effects	Yes	Yes	Yes	Yes
Pre-treatment employment controls	No	Yes	No	Yes
R-squared	0.498	0.787	0.626	0.806
Observations	4,774	4,774	3,255	3,255

Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The dependent variable is the change in employment between 2000 and 2008 at the industry and area level. Pre-treatment employment controls are the 2000 and 1990 levels at the appropriate industry level as well as in total employment. There are 23 (2-digit) industries in columns 1 and 2 and 15 in columns 3 and 4.

[18] Additional evidence that interactions between firms are highly localized has been provided by Rosenthal and Strange (2003) and Arzaghi and Henderson (2008) for the US and by Viladecans-Marsal (2004) and Jofre-Monseny (2009) for the Spanish case.

Table 4 reports the outcomes of specifications in which the 2000-2008 employment change at the (2-digit) industry level is regressed on the job losses attributable directly to the closure and case industry fixed-effects. In columns 2 and 4 we also include pre-treatment employment controls. All the coefficients in Table 4 are statistically insignificant and close to zero, suggesting that plant closures have no effect on employment levels outside the industry directly affected by the closure. Since one job directly lost in the closure reduces employment in that industry by less than one job, it is important to keep in mind that the regressions reported in Table 4 measure the impact of net job reductions in the affected industry. This goes some way to reconciling our results with those reported by Moretti (2010), which suggest that reductions in tradable jobs reduce employment in the non-tradable industries.

4.3 The effects of plant closures on incumbents and new entrants

The results reported in section 4.1 indicate that for each job lost due to a plant closure only around 0.6 jobs are lost in the affected industry. This suggests that jobs are created in the industry and area directly affected by the closure. In this regard, it is interesting to determine whether these jobs are created by incumbent or new firms. To answer this question we draw on data from the SABI (firm-level) database. Although SABI does not cover the universe of Spanish firms, its coverage is extensive (around 80 percent of the firms on the Social Security register) and it does include the self-employed[19]. We identify in the SABI database all firms reported as being active in the industry affected by the plant closure. This means the industry definition applied here is somewhat wider than that used above as a firm might be active in more than one industry. Columns 1 to 3 in Table 5 re-estimate the baseline analysis using local employment levels built with the SABI database. We exclude the jobs in the plant closed down and, thus, the 'mechanical effect' now becomes zero.

[19] SABI is a firm and not a plant database. Nevertheless, the Spanish economy is dominated by small and medium sized firms. In fact, only 1.1 percent of the firms in Spain in 2006 were multi-plant firms (Encuesta sobre Estrategias Empresariales, 2008).

Table 5. Impact of a plant closure on the affected industry. SABI database. 2000-2008 changes.

	Overall			New firms			Incumbent firms		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Job losses	0.519**	0.511**	0.618**	0.013	0.009	0.022	0.533**	0.520**	0.595**
	(0.243)	(0.239)	(0.247)	(0.038)	(0.038)	(0.05)	(0.243)	(0.237)	(0.242)
Case fixed-effects	Yes	Yes	No	No	No	No	No	No	No
Pre-treatment employment controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Case industry fixed-effects	No	No	Yes	No	No	No	No	No	No
Area fixed-effects	No	No	Yes	No	No	Yes	No	No	Yes
R-squared	0.318	0.341	0.367	0.597	0.627	0.507	0.311	0.327	0.354
Observations	217	217	4,991	217	217	4,991	217	217	4,991

Note: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The dependent variable is the change in employment between 2000 and 2008 at the 2-digit industry level computed using the SABI database and excluding the plant forced to close. Columns 1, 2, 4, 5, 7 and 8 include only the treated industry for each case while columns 3, 6 and 9 include all manufacturing industries. Pre-treatment employment controls are the 2000 and 1990 levels at the appropriate industry level as well as in total employment. There are 23 (2-digit) industries in columns 3, 6 and 9.

The results indicate that for each job lost due to a plant closure, between 0.5 and 0.6 jobs are created in the local industry affected by the closure. These point estimates are slightly higher than those recorded in Table 3, which lie between 0.3 and 0.5. This result is, however, consistent with the broader industry definition used in the SABI database and the fact that SABI also includes the self-employed. Importantly, the results obtained with this alternative dataset confirm our main qualitative results, namely, that the net employment effects of large plant closures are not as high as the direct job losses associated with the closure itself. In columns 4 to 9 in Table 5 we re-run the analysis, breaking down the changes in levels of employment between incumbent firms (columns 4 to 6) and new entrants (columns 7 to 9). According to the results, the impact on jobs is concentrated in the incumbents, that is, in firms that existed before the plant was closed down.

5. Conclusions

Local and regional governments around the world use subsidies to attract large plants. Similarly, large incumbent plants will often try to hold regional governments to 'ransom' under the threat of relocating production. The argument frequently made to justify such subsidies is that large plant closures have marked effects on employment that can extend beyond those of the collective dismissal itself. Indeed, the input-output framework has been used in predicting very large net employment losses. In this paper, we have empirically estimated the 'real' net local employment responses to large manufacturing plant closures.

Specifically, we have estimated the employment effects of the closure of 45 large manufacturing plants in Spain, which relocated to low-wage countries between 2001 and 2006. We match each municipality experiencing a closure to a small set of comparable municipalities in terms of employment level and mix in the year 2000. We find that treatments and controls do not differ in their 1990-2000 (pre-treatment) employment trends, thereby lending credence to the identification assumption underpinning our differences-in-differences estimates. Our results show that when a plant closes, for each job directly lost in the plant closure, only between 0.3 and 0.6 jobs are actually lost in the local economy, with the adjustment being concentrated in local incumbent firms in the industry having suffered the closure. One implication of these findings is that they suggest traditional input-output analyses tend to overstate the net employment losses of large plant closures. In our application, the input-output prediction overestimates the negative employment consequences by an order of three.

A couple of considerations are worth making regarding the external validity of our findings. First, among the countries of the OECD, Spain's employment protection regulations are among the strictest. At the same time, following a big plant closure, Spain's regional governments often intervene to facilitate the re-employment of some of the dismissed workers in local firms. Hence, employment responses may differ in contexts with less government intervention. Second, the closures we analyze occurred in a period (2001-2006) in which the Spanish economy was growing. It could well be that the consequences of massive layoffs are far more negative in stagnant economies. This said, our findings suggest that, in normal times, local employment responses do not seem to justify the payment of large subsidies to avoid the relocation of large manufacturing plants.

References

Abadie, A. and Gardeazabal, J. (2003), "The economic costs of conflict: A case study of the Basque Country", *The American Economic Review*, 93(1), 113-132.

Arzaghi, M. and Henderson, J.V. (2008), "Networking Off Madison Avenue", *Review of Economic Studies*, 75(4), 1011-1038.

Edmiston, K.D. (2004), "The net effects of large plant locations and expansions on county employment", *Journal of Regional Science*, 44(2), 289-320.

Ellison, G., Glaeser, E. and Kerr, W. (2010), "What causes industry agglomeration? Evidence from coagglomeration patterns", *American Economic Review*, 100, 1,195-1,213.

Encuesta sobre Estrategias Empresariales. 2008. Las empresas industriales en 2006. Madrid. Fundación SEPI.

Faggio, G. and Overman, H. (2014), "The effect of public sector employment on local labour markets", *Journal of Urban Economics*, 79, 91-107.

Faggio, G., Silva, O. and Strange, W.C. (2015), "Heterogeneous agglomeration", *The Review of Economics and Statistics*, *forthcoming*.

Fox, W.F. and Murray M.N. (2004), "Do economic effects justify the use of fiscal incentives?", *Southern Economic Journal*, 71(1), 78-92.

Glaeser, E.L (2001), "The economics of location-based tax incentives", Discussion Paper No. 1932. Cambridge, MA: Harvard Institute of Economic Research.

Greenstone, M. and Moretti, E., (2004), "Bidding for industrial plants: Does winning a 'million dollar plant' increase welfare?", *mimeo*.

Greenstone, M., Hornbeck, R. and Moretti, E., (2010), "Identifying agglomeration spillovers: evidence from winners and losers of large plant openings", *Journal of Political Economy*, 118 (3), 536-598.

Hooker, M.A. and Knetter, M.M. (2001), "Measuring the economic effects of military base closures", *Economic Inquiry*, 39(4), 583-598.

Jofre-Monseny, J. (2009), "The scope of agglomeration economies: Evidence from Catalonia", *Papers in Regional Science*, vol. 88(3), 575-590.

Jofre-Monseny, J., Marín-López, R. and Viladecans-Marsal, E. (2011), "The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms", *Journal of Urban Economics*, 70, 61-74.

Krugman, P. (1991), *Geography and Trade*, MIT Press, Boston.

Marshall, A. (1890), *Principles of Economics*, MacMillan, London

Moretti, E. (2010), "Local Multipliers", *American Economic Review*, 100(2), 373-377.

Myro, R. and Fernández-Otheo, C.M. (2008), "Profitability of Spanish foreign direct investment", *Journal GCG Georgetown University-Universia*, 2(3), 76-99.

Overman, H. and Puga, D. (2010), "Labor pooling as a source of agglomeration: An empirical investigation", in Glaeser, E. (Ed.), *Agglomeration Economics*, Chicago University Press, Chicago.

Poppert, P.E. and Herzog Jr., H.W. (2003), "Force reduction, base closure, and the indirect effects of military installations on local employment growth", *Journal of Regional Science*, 43(3), 459-482.

Rosenthal, S.S. and Strange, W.C. (2003): "Geography, industrial organization, and agglomeration", *Review of Economics and Statistics*, 85(2), 377-393.

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

Wilson, J.D. (1999), "Theories of tax competition", *National Tax Journal*, 52(2), 269-304.

Appendix

Table A1. Big Plant Closures Sample

Case	Firm	Municipality	2-digit Industry Classification	Year	Nº of Job losses	Destination
1	Jumberca S.A.	Badalona	29 - Machinery and equipment	2002	201	China
2	Proflex S.A.	Calaf	24 - Chemicals and chemical products	2004	105	Czech Republic
3	Torcidos Ibéricos S.A.	Castellbell i el Vilar	17 - Textiles	2005	116	India
4	Braun Española S.L.	Esplugues de Llobregat	29 - Machinery and equipment	2006	684	China
5	DB Apparel Spain S.A.	Igualada	17 - Textiles	2003	255	Morocco
6	Tenería Moderna S.A.L.	Mollet del Vallès	19 - Leather and leather Products	2003	131	--
7	Hilados y Tejidos Puigneró S.A.	Sant Bartomeu del Grau	17 - Textiles	2002	502	--
8	Galler Textiles S.A.	Sant Boi de Llobregat	17 - Textiles	2003	313	Thailand
9	ZF Sistemas de dirección Nacam S.L.	Sant Boi de Llobregat	34 - Motor vehicles, trailers and semi-trailers	2006	185	Germany/France
10	José Ribatallada S.L.	Cerdanyola del Vallès	15 - Food products and beverages	2005	117	--
11	Celestica S.L.	Cerdanyola del Vallès	30 - Office machinery and computers	2004	320	Czech Republic
12	Selecciones Americanas S.A.	Sitges	18 - Wearing apparel, dressing and dyeing of fur	2005	124	China
13	IMC Toys S.A.	Terrassa	36 - Furniture and other manufacturing	2003	139	China
14	Autotex S.A.	Viladecavalls	17 - Textiles	2004	189	Czech Republic
15	TRW Automotive España S.L.	Burgos	34 - Motor vehicles, trailers and semi-trailers	2005	318	Poland/Czech Republic
16	La Cellophane Española S.A.	Burgos	25 - Rubber and plastics products	2001	310	--
17	Delphi Automotive Systems España S.L.	Puerto Real	34 - Motor vehicles, trailers and semi-trailers	2006	1,600	Morocco
18	Panasonic Iberia S.A.	Celrà	29 - Machinery and equipment	2004	214	China
19	Tybor S.A.	Massanes	17 - Textiles	2003	149	China
20	La Preparación Textil S.A.	Ripoll	17 - Textiles	2004	145	China
21	Promek S.L.	Azuqueca de Henares	34 - Motor vehicles, trailers and semi-trailers	2004	350	Poland/Czech Republic
22	Moulinex España, S.A.	Barbastro	29 - Machinery and equipment	2003	270	China
23	JoyCo España S.A.	Alcarràs	15 - Food products and beverages	2004	213	China
24	Lear Corporation Spain S.L.	Cervera	31 - Electrical machinery and apparatus	2001	1,280	Poland
25	Delphi Componentes S.A.	Agoncillo	34 - Motor vehicles, trailers and semi-trailers	2001	578	Poland
26	Electrolux España S.A.	Fuenmayor	29 - Machinery and equipment	2005	454	Hungary
27	Yoplait España S.L.	Alcobendas	15 - Food products and beverages	2001	185	France

28	Sanmina-SCI España S.L.	Leganés	32 - Radio, television and communication equipment	2001	250	Hungary
29	Vitelcom Mobile Technology S.A.	Málaga	32 - Radio, television and communication equipment	2004	433	Korea
30	Calseg S.A.	Artajona	19 - Leather and leather Products	2001	150	Tunisia
31	Findus España S.L.	Marcilla	15 - Food products and beverages	2001	471	Italy/UK
32	Viscofan S.A.	Pamplona	25 - Rubber and plastics products	2006	742	Brazil/Czech Republic
33	TRW Automotive España S.A.	Orkoien	34 - Motor vehicles, trailers and semi-trailers	2002	382	Poland
34	Valeo Sistemas de Conexión Eléctrica S.L.	San Cibrao das Viñas	31 - Electrical machinery and apparatus	2004	264	Poland
35	MMN&P Acconta S.A.	Segovia	34 - Motor vehicles, trailers and semi-trailers	2001	190	Morocco
36	Levi Strauss de España S.A.	Ólvega	17 - Textiles	2003	561	Poland/Hungary
37	Delphi Packard España S.L.	Ólvega	34 - Motor vehicles, trailers and semi-trailers	2001	560	Morocco/Romania
38	GDX Automotive Ibérica S.L.	Valls	25 - Rubber and plastics products	2005	153	Germany/Czech Republic
39	Sanmina-SCI España S.L.	Toledo	32 - Radio, television and communication equipment	2005	430	Thailand/China
40	Alcatel Lucent España S.A.	Toledo	32 - Radio, television and communication equipment	2002	150	Hungary
41	Grupo Tavex S.A.	Alginet	17 - Textiles	2006	300	Brazil/Mexico
42	Bayer Cropscience S.A.	Quart de Poblet	24 - Chemicals and chemical products	2006	300	Portugal
43	Valeo España S.A.	Abrera	31 - Electrical machinery and apparatus	2001	406	Morocco/Tunisia
44	IAR Ibérica S.A.	Montcada i Reixac	29 - Machinery and equipment	2004	423	Hungary
45	Fisipe Barcelona S.A.	El Prat de Llobregat	17 - Textiles	2006	270	China

Notes: (1) Source: Authors' own elaboration.
(2) In cases 6,7,10 and 16 we have been unable to identify the country to which the firm relocated.

Table A2. Impact of a plant closure in the affected industry.2000-2008 employment changes. Robustness checks.

	Industry affected by plant closure		Pooled industries	Industry affected by plant closure		Pooled industries
	(1)	(2)	(3)	(4)	(5)	(6)
Job losses	-0.597**	-0.771***	-0.645**	0.227	0.074	0.040
	(0.288)	(0.276)	(0.269)	(0.214)	(0.251)	(0.232)
Case fixed-effects	Yes	Yes	No	No	No	No
Pre-treatment employment controls	No	Yes	Yes	No	Yes	Yes
Case industry fixed-effects	No	No	Yes	No	No	No
Area fixed-effects	No	No	Yes	No	No	Yes
R-squared	0.596	0.787	0.822	0.626	0.841	0.832
Observations	131	131	3,013	172	172	3,956

Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The dependent variable is the change in employment between 2000 and 2008 at the 2-digit industry level. Columns 1, 2, 4 and 5 include only the treated industry for each case while columns 3 and 6 include all manufacturing industries. Pre-treatment employment controls are all the outcomes examined in Table 1.

Chapter 4

Small shops for sale! The effects of big-box openings on grocery stores

1. Introduction

In recent years many governments have adopted restrictive policies in response to the opening of big-box stores. Before 1990, many European countries underwent increasing market liberalization, as a consequence of which the retail sector, and the food retail sector in particular, expanded greatly with the opening of many new supermarkets. In the Spanish case, the five biggest supermarket chains opened their first stores in the 1970s and by 1990 they accounted for 45% of the market, according to figures published by the Spanish Ministry of Economy[1]. In this way, a highly traditional sector, made up primarily of city centre grocery stores, found itself up against a new type of competitor. The economic consequences of the opening up of these new supermarkets, typically out-of-town big-boxes, became an important policy concern in most countries. In particular, the main concern was (and still is) the impact of these stores on the quality of cities and their market structure (see, for example, Basker, 2007, for an analysis of the impact of the growth of Wal-Mart, one of the biggest big-box chains in the US). However, the proponents of big-box stores argue that they tend to push prices down and, so, consumers tend to be better off when they locate in their municipalities. In response, throughout the 1990s, many European countries, most notably the UK, Italy and France, introduced stringent policies to restrict the entry of big-box stores, or, at least, implemented controls on the type of store that could be built and where they could locate.

In this paper, I exploit a similar regulation introduced in Spain in 1997 to evaluate the effects of the entry of big-box stores on traditional grocery stores. More specifically, by implementing a 'fuzzy' Regression Discontinuity Design, I test whether the opening of big-box stores is causing grocery stores to close. If this is the case, and given that grocery stores are typically located in city centres, the opening of big-box stores would be 'hollowing out' city centres. The results show that non-regulated municipalities experience 0.3 more big-box

[1] Informe de Distribución Comercial 2003 (http://www.comercio.mineco.gob.es/es-ES/comercio-interior/Distribucion-Comercial-Estadisticas-y-Estudios/Pdf/InformeDistribucion_2003.pdf)

openings than regulated municipalities, and, as a consequence, four years after the first big-box opening, between 20 and 30% of the grocery stores in the area disappear, offering clear evidence that city centres are losing part of their economic activity. I also examine whether these effects differ according to the location of the big-box (city centre vs. out-of-town) and the typology of the big-box opened (conventional vs. discount). To this end, I exploit the possibility that big-boxes located in the city centre, and therefore closer to the grocery stores, have a different impact to that of big-boxes opened in the suburbs. I also analyse whether conventional big-box stores, selling well-known brands, have a different impact to that of discount stores, selling their own brands at lower prices. The results show that there does not seem to be a significant difference between big-box stores operating downtown and those operating in the suburbs, at least in the short run. However, in the case of the typology, results show that it seems to be the conventional supermarkets that are competing with grocery stores and forcing them to pull down their shutters.

Several papers have examined the impact of planning (and/or commercial) regulations in the retail sectors of various countries. For instance, Bertrand and Kramarz (2002) exploit a French regulation requiring regional approval for the opening of large retail stores. They show that this barrier to entry and high levels of concentration among large retail chains significantly reduce retail employment, stemming its growth rate. Schivardi and Viviano (2011) exploit a similar regulation in Italy and, using political variables as instruments, find that this entry barrier is associated with substantially larger profit margins and lower productivity of incumbent firms. Griffith and Harmgart (2008), for the UK case, build a theoretical model allowing for multiple store formats and introduce a restrictive planning regulation. They report that planning regulations have an impact on market equilibrium outcomes, although not as great as suggested by the previous literature. Haskel and Sadun (2012), also focusing on the UK retail sector, find that by preventing the emergence of more productive, large format stores and by increasing the costs of space, planning policies impede the growth of the sector's total factor productivity (TFP). The same results are reported by Cheshire et al. (2015) in their examination of the effects of 'Town Centre First' policies in the UK's large supermarket sector. They find that such policies directly reduced output by forcing stores onto less productive sites.

The issues addressed in this paper are closely related to another branch of the literature examining the effects of big-boxes on grocery stores, but more specifically focused on the role of competition and its impact on employment. Most studies here have analysed the impact of Wal-Mart stores in the US. Basker (2005) reports an instantaneous positive effect of a Wal-Mart opening on retail employment, although the effect is halved five years after the opening. Others, including Neumark *et al.* (2008), using an instrumental variables approach, show that Wal-Mart openings have a negative effect on retail employment and wages in US counties. Haltiwanger *et al.* (2010) use data from grocery stores in the Washington DC metropolitan area to evaluate the effects of the first Wal-Mart opening on grocery stores and small supermarkets. They find negative effects of the big-box on other retailers, especially for those located closest to the Wal-Mart facility. The same results are reported by Ellickson and Grieco (2011) in their analysis of a panel dataset for the years 1994 to 2006 for the whole country. Finally, Jia (2008) also evaluates the effects of Wal-Mart openings on grocery stores but, in line with the present paper, focusing on their exit decisions. The study develops an empirical model to assess the effects on discount grocery stores of big-box store openings.[2]

However, the European food retail sector works very differently from that in the US, given the continent's different city structures and the agglomeration forces operating in its cities. Sadun (2015) is the only paper, to date, to analyse the European case. In a study of UK retailers, the author finds that following the introduction of stringent policies, supermarket chains adapted the size of their outlets to the regulation resulting in stores that can compete even more directly with the grocery stores, and so harming them even more than before the policy. Adopting a theoretical perspective, Uschev *et al.* (2015) build a model in which, combining spatial and monopolistic competition, they find that downtown retailers gradually disappear when a big-box is sufficiently large.

The main contribution of this paper is that it is, to the best of my knowledge, the first attempt to study the direct effects of big-box store openings on grocery stores using a quasi-experimental design, in this case that of a Regression Discontinuity Design. Previous papers, exploiting similar regulations, use political variables as their instruments to evaluate the causality of the effects (see Sadun, 2015). The novelty of this paper is that the source of exogenous variation is generated by the commercial regulation itself, thanks to the fact that this regulation varies across the regions and across the

[2] Other studies of the impact of Wal-Mart stores, including Basker (2005) and Basker and Noel (2009), focus on other outcomes such as grocery store prices.

municipalities within each region. Therefore, it is unnecessary to rely on any other external source of exogenous variation. In addition, this is the first paper to show the impact of the opening of big-box stores on grocery store closures drawing on all available data for big-box openings and, hence, distinguishing the effects by location and typology of these stores. Previous studies in the US have been limited to the role played by Wal-Mart stores. Moreover, this is the first European study to focus specifically on the number of grocery stores forced out of the market, given that the only other paper available (Sadun, 2015) focuses on the employment effects of the opening of big-box stores. The results reported here show that, following the introduction of stringent policies, non-regulated municipalities experienced more grocery store closures than were suffered by regulated municipalities, pointing to the policies' effectiveness in saving existing businesses. These findings seem to complement those reported by Sadun (albeit focused more specifically on employment), suggesting that restrictive policies in the retail sector may have a different impact in southern Europe to the effects described in the UK. Finally, my results are in line with the theoretical findings of Uschev *et al.* (2015) who conclude that big-box stores may contribute to the 'hollowing out' of the city centres.

The rest of the paper is structured as follows. Section 2 presents the institutional setting as well as the regulation exploited while Section 3 introduces the different data sources. Section 4 states the empirical strategy used and presents the results for the first stage estimations, i.e. the effect of the commercial regulation on big-box openings. Section 5 shows the results of the effect of big-box openings on grocery stores and reports some robustness tests and heterogeneous effects. Section 6 concludes.

2. The institutional setting

Between 1985 and the mid-1990s, Spain experienced a change in its market structure with the complete international liberalisation of the retail sector, affecting above all the food retail trade (Matea and Mora-Sanguinetti, 2009, show an increase in restrictiveness from the late 1990s with respect to the previous decade). Thus, a market that had previously been dominated by grocery stores saw the arrival of the supermarket, most belonging to foreign chains. These changes ushered in a major policy debate between those in favour and those opposed to trade liberalisation and free market entry, a debate that became even more heated when the supermarket chains began opening large out-of-town stores. The detractors of such stores argue that big-box openings create enormous externalities for the local community, including more pollution, distortions to the existing retail market structure and the hollowing-out of city centres. One of their chief arguments is that these stores affect the pre-existing body of firms, especially small, traditional businesses, causing their eventual disappearance from the area. Thus, to prevent this from happening and in response to the growing unrest in the sector, in 1996, the Spanish parliament passed a law aimed, among other things, at restricting the entry of big-box stores.[3] [4] The law required a developer seeking to open a big-box store in Spain to obtain a second licence, in this case from the regional government, in addition to the municipal licence. The fact that the two licences (municipal and regional) have to be solicited from two different entities means that big-box developers incur an additional entry cost vis-à-vis grocery stores. While this is not a monetary cost, it does represent a considerable cost in terms of time and uncertainty given the amount of red tape developers have to contend with in applying for this second licence.

The key to this new regulation lies in its definition of what should be considered a "big-box store". The central government opted to define a big-box as one with at least 2,500 m^2. However, nine (out of Spain's seventeen) regions chose to strengthen the law by further limiting the number of square metres. This they did in line with the population of their municipalities. Thus, in smaller cities a more restrictive definition was placed on the size of big-box stores, making their market entry even more difficult. Each region set their own arbitrary population thresholds, introducing the corresponding measures between 1997 and 2004[5]. Here, therefore, in order to identify the causal effects of big-box openings on grocery stores in an operative way, I focus on those municipalities centred on the lowest population threshold as defined by most of the regions: namely, 10,000 inhabitants. This means that, for all regions, municipalities below the 10,000 population threshold restrict the opening of big-box stores, while municipalities above this threshold are non-regulated. Note, that three regions did in fact define lower thresholds but these are discarded because they do not provide enough observations to perform the

[3] Retail Trade Law 7/1996 of 15 January 1996

[4] The law also regulated store opening hours as well as licences for hard discount stores.

[5] Note that the adoption of the regulation was not a party political issue as the nine regions were governed by different parties with different ideologies at the time of its introduction. Four regions had a socialist party in office, three were governed by a conservative party and the other two regions were governed by regional nationalist parties.

analysis. Additionally, most Spanish municipalities are very small (almost 60% have less than 5,000 inhabitants), which means establishing a threshold above 10,000 would only capture restrictions for a specific set of large cities. Thus, using a larger threshold would not be operative here. For the same reason, there will be more observations to the left of the threshold than there are to the right. Table 1 shows the specific details of the regulations – size restrictions and the year they were introduced – for the nine regions included in the analysis. Note that the definition of a big-box varies across the regions, ranging from 600 to 1,500 m^2. In the empirical analysis I use each region's specific definition, but I also include region fixed effects in all the estimations. As such, the analysis undertakes a within region comparison where the size threshold is the same for all municipalities in that region, independently of the regulation.

Table 1: Commercial regulations per region for the 10,000 inhabitant threshold

Region	Size restrictions	Year of introduction
Andalusia	> 1000 m^2	2002
Castile and Leon	> 1000 m^2	1997
Castile-la Mancha	> 750 m^2	2004
Catalonia	> 800 m^2	2001
Extremadura	> 750 m^2	2002
Balearic Islands	> 600 m^2	2001
La Rioja	> 1000 m^2	1997
Community of Madrid	> 1500 m^2	1999
Basque Country	> 800 m^2	2001

Note: The table shows the definition of big-box store used in each of the nine regions that strengthened the central law and the year this regional law was introduced for the 10,000 inhabitant threshold.

3. Data and sample

I use two different datasets to perform the analysis. First, data concerning the openings of big-box stores are drawn from a private dataset compiled by Alimarket, S.A, a company that generates information (from sources that range from news articles to databases) for different industries in Spain. I draw specifically on their food and beverages dataset and use their 2011 Census of Chain Supermarkets in Spain. For each big-box, this census contains information on its date of opening, exact location, size (in square meters) and the chain to which they belong. Although this is not a panel dataset, the time dimension can be added by exploiting the information on the date each big-box store was opened. This means that, as with any census, the dataset only contains information on the stores surviving in 2011. However, the closure of a big-box store, especially in the period analysed, is highly unlikely.[6] It should be stressed at this juncture that information regarding the number of licences per municipality is unavailable, which means little can be said about the administrative process for the granting of licences. Indeed, I am only able to observe those that met with success (i.e. the actual number of big-box openings per municipality and year).

For information on grocery stores (i.e., the outcome variable), I use the *Anuario Económico de España* (AEE), a municipality dataset, for the period 2003 to 2011. This dataset includes detailed local demographic and economic variables for municipalities with more than 1,000 inhabitants. More specifically, in the case of the food retail sector, it records the exact number of stores in each Spanish municipality and year, classifying them in two categories: traditional stores (i.e. grocery stores) and supermarkets (i.e. chain stores, not necessarily big-boxes). The number of traditional stores is used to identify the effects of big-box openings on grocery store closures. According to the literature (for example, Bertrand and Kramarz, 2002) and anecdotal evidence from local planners in Spain (provided by Matea and Mora-Sanguinetti, 2009), four years would appear to be the plausible, average time lag between applying for a licence to build a big-box store and its eventual opening. This means the effects of the 1997 regional regulation would not make themselves manifest until 2001 and so the period of analysis should start in 2001. However, the AEE only began distinguishing between grocery stores and supermarkets in 2003, further restricting the period of analysis from 2003 to 2011, the latter year corresponding to the Alimarket Census.

Other variables may, at the same time, be influencing the numbers of big-box openings and grocery stores. In order to control for this, local economic and socio-demographic variables extracted from the Spanish National Institute of Statistics (INE) 2001 Census are used. Specifically, I use an index representing the average economic activity of each municipality, computed by the INE using data about the occupation and professional activity of the

[6] Using the 2007 Census of Chain Supermarkets it can be verified that between 2007 and 2011 there were no big-box closures, that is, those stores operating before 2007 remained in the sample in 2011.

population in the municipality. Additionally, I also use two indicators of level of education achieved: compulsory education and post-compulsory education, defined as a percentage of the overall local population. Finally, a variable showing the share of immigrants as a percentage of the overall population is included as is another variable capturing the importance of the services sector, i.e., the share of the services sector within a municipality's total activities. In addition to the Census data, a variable capturing the surface of the municipality (km^2) is included. Table 2 shows the descriptive statistics for the outcome variable, i.e. number of grocery stores at the municipality level, as well as for the control variables. Their values are all presented around the threshold (+/- 3,000 inhabitants from the 10,000 threshold).

Table 2. Outcome and control variables - Descriptive statistics around the threshold (+/- 3,000 inhabitants of the 10,000 threshold)

	Obs.	Mean	S.D.	Min	Max
Outcome					
Number of grocery stores	795	58.94	35.12	5	236
Controls					
Economic activity	795	0.919	0.157	0.61	1.25
Compulsory education (%)	795	47.13	10.36	22.19	72.27
Post-compulsory education (%)	795	34.21	8.73	10	62.51
Square kilometres	795	119.26	124.96	2	586
Immigrants (%)	795	2.48	3.53	0.02	21.92
Unemployment rate (%)	795	15.98	9.74	4.07	61.23
Importance of the services sectors (%)	795	50.38	12.40	20.32	81.77

Source: Based on AEE and Census data. Notes: (1) The outcome variable is defined using AEE data and represents the universe of grocery stores at the municipality level. (2) The control variables are all extracted from the 2001 Census. (3) The variable *Economic activity* represents the average of an index of the economic activity of each municipality. It is computed using data on the occupation and professional activity of the population in the municipality. The variables *Compulsory education, Post-compulsory education* and *Immigrants are* computed as a percentage of the overall population. The *Importance of the services sectors* variable is computed as a percentage of the overall activities within a municipality.

As discussed above, there is, on average, a four-year lag between the developers applying for a license and the big-box being opened. Therefore, as I only observe the date of opening but the regulation applies from the moment the developers request the licence, each opening has to be matched with its corresponding population at a point four years earlier – that is, I match the openings from 2003 to 2011 with population data from 1999 to 2007, respectively, as extracted from INE data. The initial pooled sample size comprises a total of 2,020 municipalities per year belonging to the nine regions that strengthened the central law. I restrict the sample to municipalities with between 1,000 and 50,000 inhabitants that did not have a big-box store before the onset of my period of analysis[7]. This means discarding 656 municipalities from the sample. I also exclude a further 83 municipalities that crossed the threshold three, two or one year(s) prior to the opening. Finally, I only include municipalities once the region in which they lie has implemented the regulation; thus, for each year, I only include the regulated regions' municipalities. This means I only estimate the post-regulation effect.[8]

Table 3. Sample size

Year	Observations	Big-Box Openings
2003	241	5
2004	241	6
2005	544	11
2006	1,113	41
2007	1,113	85
2008	1,281	49
2009	1,281	45
2010	1,281	55
2011	1,281	20
Total		317

Note: The initial sample comprised the 2,020 municipalities belonging to the nine regions that strengthened the central law. However, the sample shown here is a restricted sample based on the following criteria: municipalities with less than 50,000 inhabitants and having a big-box store before the period of analysis have been discarded. This means eliminating 656 municipalities from the sample. The 83 municipalities that crossed the threshold three, two or one year(s) prior to the opening have also been excluded. Finally, municipalities are only included once their region has implemented the regulation; thus, for each year, the sample consists only of the regulated regions' municipalities.

[7] Note that municipalities with less than 1,000 inhabitants are also excluded from the sample due to AEE data availability.

[8] It would have been interesting to estimate the before- and after-policy effects but, as the study period starts in 2003, I lack pre-regulation data for three of the regions. Table 3 reports the number of municipalities, i.e. the sample size, and the number of big-box openings per year.

4. Identification strategy

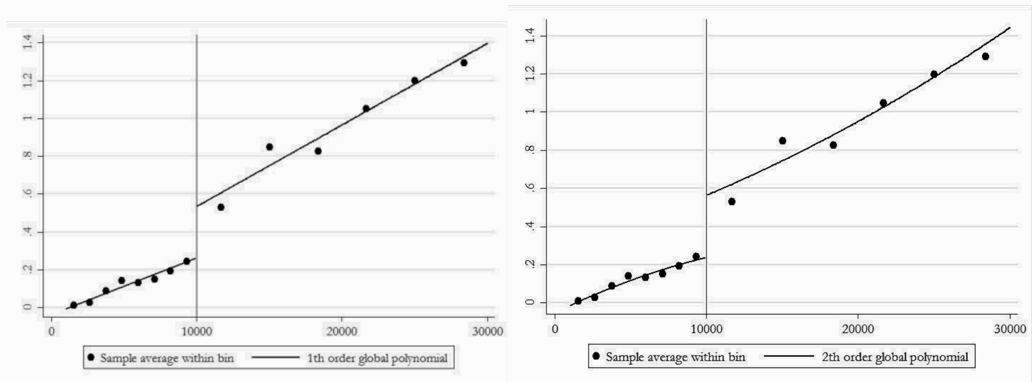
I use a Regression Discontinuity Design (RDD) framework to estimate the effects of big-box openings on grocery store closures. As discussed, to build a big-box store in a municipality of less than 10,000 inhabitants, a second regional licence is required. However, this licence should be seen as an additional barrier to entry, since it is by no means a binding constraint. In a "sharp" RDD, the treatment jumps from zero to one at the threshold. In a setting such as the one described here, this would mean that non-regulated areas (those with more than 10,000 inhabitants) are the only ones in which big-box stores open. However, as this is not the case, the setting requires the use of a "fuzzy" RDD, the crucial assumption being that there is a discontinuity in the probability of assignment at the threshold (see Imbens and Lemieux, 2008 and Lee and Lemieux, 2010 for a fuller discussion of "sharp" and "fuzzy" RDDs). In other words, the probability of establishing a big-box store jumps on crossing the threshold from regulated to non-regulated municipalities. This is the so-called 'first stage' that is used afterwards as an instrument in a two-stage least squares (2SLS) regression to identify the causal effect. In this section, I begin by examining this first stage; that is, testing whether there are systematically more openings in non-regulated municipalities than there are in their regulated counterparts around the threshold.

The "fuzzy" RDD relies on the assumption that the probability of assignment to treatment jumps at a particular threshold and, as such, this can be used as a source of exogenous variation. However, this assumption needs to be tested. Before empirically estimating the existence of such a jump, I first examine it graphically using the raw data. Figure 1 shows the jump in the number of big-box openings at the threshold. Panel (a) presents the results for a first order polynomial fit while panel (b) reports the results for a second order polynomial. In both cases we observe a jump at the threshold of around 0.3, meaning that, when crossing from regulated to non-regulated municipalities, there are, on average, 0.3 more big-box openings. We also see that there is very little difference when fitting different order polynomials. In order to assess this more formally, I estimate variants of the following equation:

$$\textit{big-box openings}_{it} = \alpha_{it} + \beta_{it} \cdot T_{it} + \gamma_{it} \cdot f(P_{i,t-4}) + \delta_t + \theta_r + X_{it}'\omega + \varepsilon_{it} \qquad (1)$$

where $\textit{big-box openings}_{it}$ is the number of big-box openings in municipality i up to time t, that is, the change in the stock of big-box stores up to time t. The variable that identifies the jump in treatment is T_{it}, which takes a value equal to one if the municipality is above the threshold and zero otherwise. The running variable is the four-year lagged population ($P_{i,t-4}$), which enters the equation using different polynomial degrees. The regression also includes a set of control variables (X'_{it}), region and time fixed effects to control for time invariant region characteristics and countrywide shocks, respectively. Additionally, the region fixed effect controls for the fact that the regulation varies by region; thus, by incorporating this fixed effect, I am performing a within-region analysis. The controls are included in order to capture variables that might affect both big-box store openings and the change in the number of grocery stores. These are the pre-regulation levels of population, economic activity, education levels, size of the municipality (in km^2), immigration level, unemployment rate and the importance of the services sector.

Figure 1: Jump in the number of big-box stores at the threshold



Panel (a) Panel (b)

Note: Panel (a) shows bin averages of the number of big-box openings using the raw data and adjusting a linear polynomial at each side of the threshold. Panel (b) shows the same but adjusting a quadratic polynomial at each side of the threshold.

Table 4 presents the results of this first stage equation, i.e. the effect of commercial regulation on the number of big-box openings. The first four columns show the results of estimating equation (1) using polynomial regressions while the last three present the results of estimating the same equation using local linear regressions. For the polynomial regressions, I use first- and second-degree polynomial fits, which according to Figure 1 would seem to fit the data properly.[9] Columns (1) and (2) show the results without the control variables while columns (3) and (4) report the results when including them. All the regressions seem to adapt well to the features presented by the raw data in Figure 1. The preferred estimation is the one in column (4), which presents a better fit and controls for observables that may be influencing both the outcome and the explanatory variable. Columns (5) to (7) report the results of local linear regression estimations using the Imbens and Kalyanaraman (2012) methodology. Column (5) presents the results for the optimal bandwidth while columns (6) and (7) show the results for half and twice the optimal bandwidth, respectively. All the results, with the exception of the half optimal bandwidth (owing to the small sample size), also show a jump in treatment at the threshold of around 0.3 – or slightly higher – coinciding with the graphical inspection.

Table 4. The effect of commercial regulations on big-box openings

	Dependent variable: Number of big-box openings						
	Polynomial Regressions				Local Linear Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T_{it}	0.219*	0.303***	0.277**	0.331***	0.429***	0.735***	0.385***
	(0.13)	(0.111)	(0.123)	(0.108)	(0.111)	(0.175)	(0.072)
Polynomials	1	2	1	2	--	--	--
Bandwidth	--	--	--	--	Optimal	-50%	+50%
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7,095	7,095	7,095	7,095	6,696	1,445	6,937

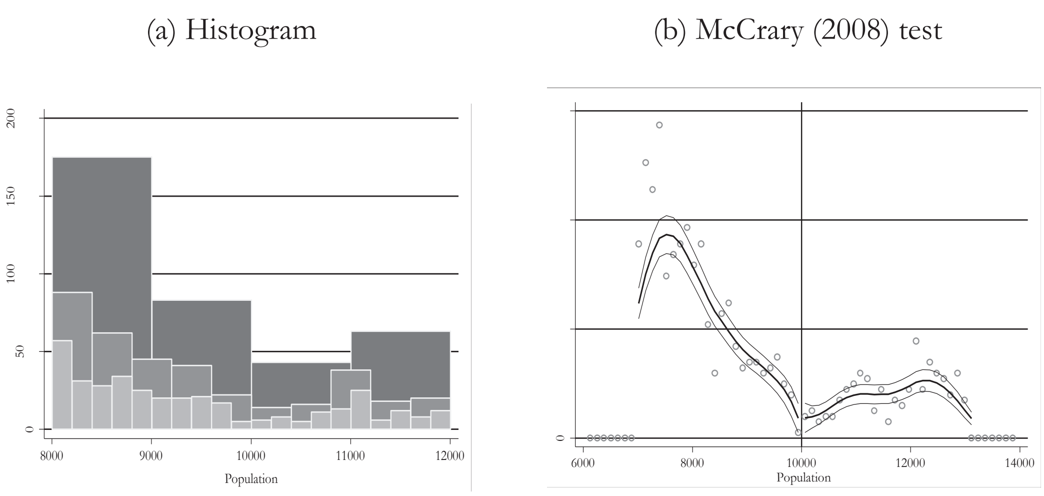
Notes: (1) Robust standard errors in parentheses, clustered at the municipality level (2) T is a dummy that takes a value equal to one if the municipality is above the 10,000 inhabitant threshold and zero otherwise. (3) All regressions include region and time fixed. (4) Columns (3) to (7) include pre-regulation levels of population, economic activity and education levels, size of the municipality (in km^2), immigration level, unemployment and importance of the services sector in order to control for trends. (5) *** $p<0.01$, ** $p<0.05$, * $p<0.1$

One assumption of the RDD strategy is that the 'forcing' variable must be continuous at the threshold. In order to reject any manipulation of this forcing variable, I inspect the histogram of the population around the threshold. A more formal way of assessing this is to run local linear regressions of the density of the forcing variable on both sides of the threshold, as proposed by

[9] I also estimated the regressions using a third-degree polynomial fit but the polynomial turned out to be non-significant.

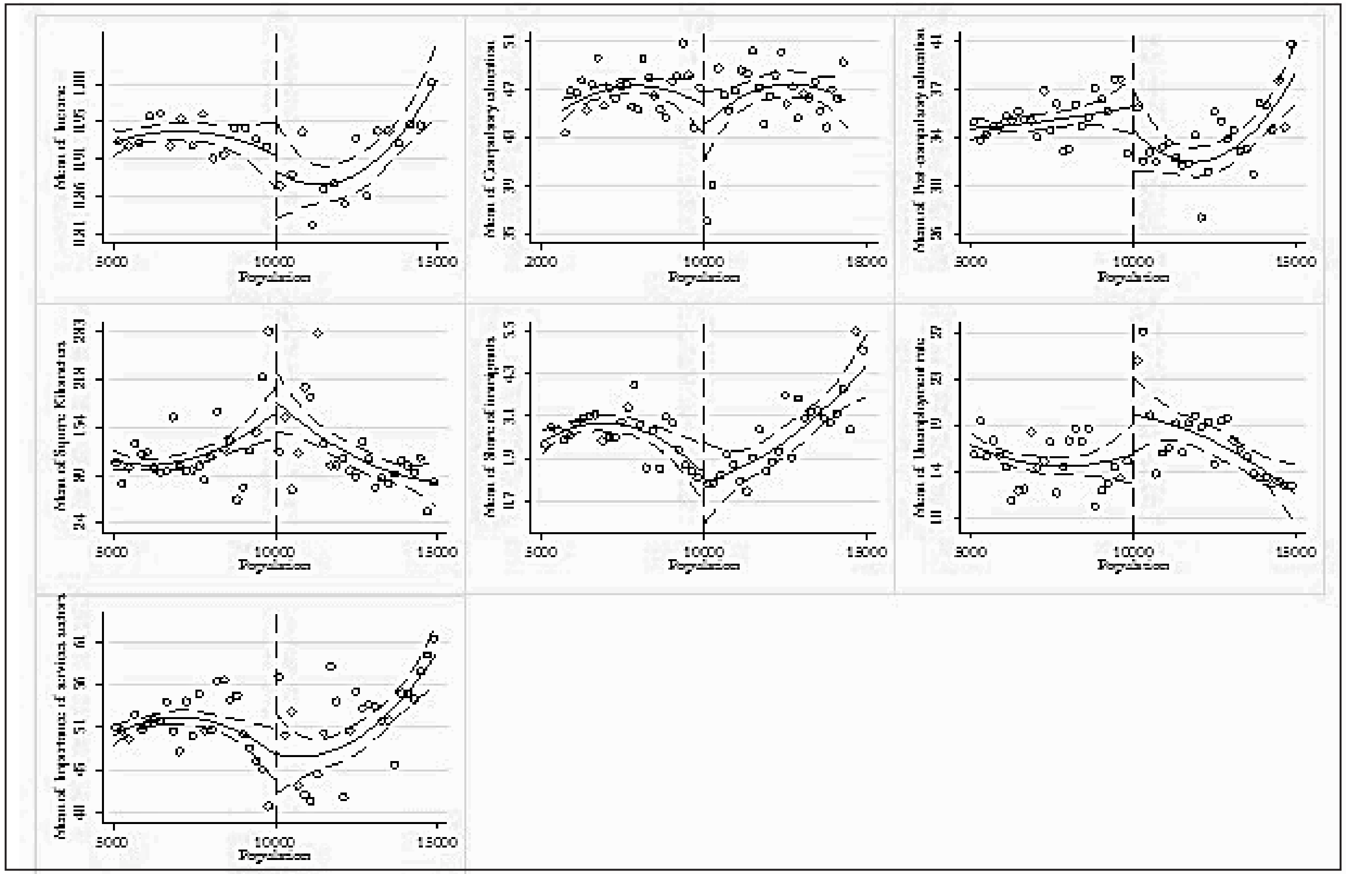
McCrary (2008). Figure 2 presents the results of both methods for examining the continuity of the forcing variable at the threshold. Panel (a) shows the histogram of the population using different bin widths: the largest width is 1,000 inhabitants, the mid-scale is 400 inhabitants and the smallest is 200 inhabitants. Panel (b) shows the results of the McCrary test. In both cases, we observe that the forcing variable is not discontinuous at the threshold. Interestingly, Foremny *et al.* (2015), in a study of Spanish local government manipulation of reported population levels to obtain higher transfers, conclude that municipalities around the 10,000 threshold do not misreport their population numbers as grants do not change at this threshold.

Figure 2: Continuity of the forcing variable at the threshold



Note: Panel (a) shows the histogram for three different bin widths: 1,000, 400 and 200 inhabitants. Panel (b) presents the results of the McCrary test, consisting on running local linear regressions at both sides of the threshold. The circles represent bins of the population density.

Figure 3: Continuity of the control variables



Note: All graphs present local linear regressions of the control variables on each side of the threshold. Starting from the top left corner the variables shown are economic activity, compulsory education, post-compulsory education, surface (in km²), share of immigrants, unemployment rate and importance of the services sector.

A further assumption that must be met in order for an RDD to work is that no other variable at the municipality level should experience a jump at the threshold, because if this were not the case, the coefficient would also be identifying this jump. In order to test that this does not occur in this setting, at least for the observables, I examine the continuity of the control variables used in the regression (i.e. those reported in Table 2) at the threshold. I adjust local linear regressions on each side of the threshold for each of the control variables and plot them. Figure 3 shows the results. We observe that none of the control variables presents a jump at the threshold and, therefore, the coefficient previously estimated is only capturing the effect of the regulation on big-box openings.

In order to test the robustness of these first stage results, I estimate equation (1) again, but instead of using the sample of post-regulation municipalities, I perform the analysis using the non-regulated municipalities in each year, i.e. the pre-regulation sample. If this placebo exercise works, there should be no difference in the number of big-box openings around the threshold.

Table 5 reports the results of this placebo test. The structure of the table is the same as that in Table 4, with the first four columns presenting the results for polynomial regressions with and without control variables and the last three columns showing the results for local linear regressions. All the estimations show that there is no difference between municipalities around the threshold prior to the regulation. In fact, if anything, according to columns (1) and (5), it would be negative. Thus, we conclude that the difference in the number of big-box openings at the threshold identified in Table 4 is due to the commercial regulation.

Table 5. Placebo test - The effect of commercial regulations on big-box openings in non-regulated municipalities

	Dependent variable: Number of big-box openings						
	Polynomial Regressions				Local Linear Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T_{it}	-0.163*	-0.005	-0.060	0.016	-0.03***	0.000	-0.009
	(0.088)	(0.059)	(0.072)	(0.053)	(0.011)	(0.017)	(0.020)
Polynomials	1	2	1	2	--	--	--
Bandwidth	--	--	--	--	Optimal	-50%	+50%
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Obs.	2,641	2,641	2,641	2,641	2,495	531	2,581

Notes: (1) Robust standard errors in parentheses, clustered at the municipality level (2) The sample used in all regressions consist on the pool of the non-regulated municipalities in each year. (3) The independent variable is a dummy that takes a value equal to one if the municipality is above the 10,000 inhabitant threshold and zero otherwise. (3) All regressions include region and time fixed effects in order to control for region specific time invariant characteristics and countrywide time shocks. (4) Columns (3) to (7) also include the pre-regulation levels of population, economic activity and education levels, size of the municipality in square kilometres, immigration level, unemployment and importance of the services sector in order to control for trends. (5) *** p<0.01, ** p<0.05, * p<0.1

5. Results

In this section, the results of the 2SLS regressions estimating the effects of big-box openings on grocery store closures are presented and interpreted. In addition, a number of robustness tests are presented. Finally, the potentially heterogeneous effects of the location and type of big-box opened are evaluated.

5.1. The impact of big-box openings on grocery store closures

This section presents the results of evaluating the effect of big-box openings on grocery store closures. To address this question, I estimate the following 2SLS equation, where the key variable regarding the opening of big-box stores is instrumented with the treatment variable from the first stage (T_{it}) obtained when estimating equation (1):

$$\Delta\, grocery\ stores_{it} = \theta_{it} + \varphi_{it} \cdot big\text{-}box\ openings_{it} + \sigma_{it} \cdot g\,(P_{i,t-4}) + \varrho_t + \pi_r + X'_{it}\vartheta + \epsilon_{it} \quad (2)$$

where Δ *grocery stores*$_{it}$ is the change in the number of grocery stores between *t* and *t-n* (where *n* is between 1 and 5) aggregated at the municipality level. This equation is also estimated for the two different degrees of polynomial fit: a first-degree and a second-degree fit. As before, *big-box openings*$_{it}$ is the number of big-box openings in municipality *i* up to time *t,* so it also represents the change in the stock of big-box stores. The regression also includes the same control variables as in the first stage, (X'_{it}) as well as region and time fixed effects. The coefficient of interest is φ_{it}, which can be interpreted as the ratio between two "sharp" RDDs. The "intent-to-treat" estimation, i.e. a reduced form of the effect of T_{it} on *grocery stores*$_{it}$, is divided by β_{it} obtained from equation (1).

Table 6 presents the results of estimating the effects of big-box openings on grocery store closures. The first four columns show the results of estimating polynomial regressions, while the fifth reports the results of estimating a local linear regression using the optimal bandwidth. In columns (1) and (2) the control variables are not included, while in columns (3) and (4) they are. To test whether there are any effects of big-box openings on grocery store closures, equation (2) is estimated using the change between *t* and *t-2*, *t* and *t-3*, *t* and *t-4* and *t* and *t-5*. Specifically, I estimate the equation separately for each of these four time spans, their results being presented in each row of Table 6. As in Table 4, the preferred estimation is the one in the fourth column. Examining the results in Table 6, it can be seen that the opening of big-box stores has some effects on the number of grocery stores, these effects being manifest two to four years after the opening. Indeed, the opening of a big-box store in a given municipality results in the gradual closure of grocery stores. Around ten grocery stores have shut down two years after a big-box opening and the number of closures increases to between 14 and 20 stores by the end of the fourth year. Note that the regressions representing the effects five years after the opening present very similar coefficients, showing that the impact seems to be concentrated within the first four years following the opening. To put these numbers into perspective, they should be compared with the means around the threshold reported in Table 2. Thus, losing between 14 and 20 grocery stores in the four-year period represents a loss of between 20 and 30% of the existing grocery stores in an area where a big-box store has opened. If we examine the last column, which shows the local linear regression, we observe that, although the point estimates are the same as before, the conventional errors are larger and the coefficients are no longer significant.

Table 6. The effect of big-box openings on grocery store closures

		Dependent variable: Change in the number of grocery stores				
		Polynomial regressions				LLR
		(1)	(2)	(3)	(4)	(5)
Big-Box openings t,t-2	Coef.	-6.35	-5.42	-10.44*	-9.21**	-13.67
	s.e.	(6.25)	(4.12)	(6.11)	(4.45)	(8.91)
	Obs.	5,814	5,814	5,814	5,814	4,247
Big-Box openings t,t-3	Coef.	-13.80	-9.11*	-16.17*	-12.87**	-16.49
	s.e.	(9.38)	(5.52)	(8.37)	(5.75)	(10.62)
	Obs.	4,533	4,533	4,533	4,533	4,062
Big-Box openings t,t-4	Coef.	-20.28	-10.72	-20.33*	-13.82**	-10.47
	s.e.	(12.78)	(6.98)	(10.77)	(6.96)	(8.66)
	Obs.	3,252	3,252	3,252	3,252	1,708
Big-Box openings t,t-5	Coef.	-23.78*	-11.86	-20.92**	-13.01*	-8.73
	s.e.	(13.03)	(8.07)	(10.57)	(7.48)	(8.53)
	Obs.	2,139	2,139	2,139	2,139	1,355
Polynomials		1	2	1	2	--
Bandwidth		--	--	--	--	Optimal
Controls		No	No	Yes	Yes	Yes

Notes: (1) Robust standard errors in parentheses, clustered at the municipality level (2) The independent variable is the number of big-box openings between t and t-n at the municipality level, instrumented by a dummy that captures the change in the probability of treatment due to the commercial regulation. Each row represents a different regression. (3) All regressions include region and time fixed effects in order to control for region specific time invariant characteristics and countrywide time shocks. (4) Columns (3) to (5) also include the pre-regulation levels of population, economic activity and education levels, size of the municipality in square kilometres, immigration level, unemployment and importance of the services sector in order to control for trends. (5) *** p<0.01, ** p<0.05, * p<0.1

These results are robust to different tests. Table 7 shows the results of estimating equation (2) in three different settings. It also presents the first stage results for each of the three tests. Only the results for the preferred estimations are presented in each setting, i.e. the second-degree polynomial regression and the local linear regression using the optimal bandwidth. The first two columns present the results of estimating the effects of big-box openings on grocery stores when the municipalities that experienced a big-box opening before the regional law was passed are also included. In this case, we observe a very similar first stage and a slightly smaller, but qualitatively similar, second stage. This is a reasonable result given that the municipalities affected by a big-box opening prior to the introduction of the regulation may have already experienced grocery store closures. As such, their inclusion is offsetting to some extent the previous results. Columns (3) to (4) and (5) to (6) present the results when using as the running variable the population lagged one year more and one year less than in the original regression, i.e. using the three-year lagged population and the four-year lagged population, respectively. In both cases, the first stage remains the same as in Table 4 and the second stage is the same as that shown in Table 6. This test shows that the results are not sensitive to the lags of the running variable.

Table 7. The effect of big-box openings on grocery store closures – Robustness checks

		Dependent variable: Change in the number of grocery stores					
		Openings before the law		3-years-lagged population		5-years-lagged population	
		PR	LLR	PR	LLR	PR	LLR
		(1)	(2)	(3)	(4)	(5)	(6)
Big-Box openings t,t-2	Coef.	-7.03*	-8.81	-9.36**	-8.53	-8.586*	-10.05
	s.e.	(3.91)	(8.89)	(4.74)	(7.83)	(4.40)	(6.75)
	Obs.	6,321	5,708	5,844	5,513	5,814	5,517
Big-Box openings t,t-3	Coef.	-10.89**	-11.63	-12.26**	-11.14	-12.03**	-16.52*
	s.e.	(4.94)	(9.60)	(6.14)	(9.83)	(5.62)	(10.04)
	Obs.	4,929	4,478	4,558	4,288	4,533	3,353
Big-Box openings t,t-4	Coef.	-10.85*	-11.68	-11.9	-9.06	-13.24*	-13.19*
	s.e.	(6.05)	(9.23)	(7.76)	(9.60)	(6.83)	(7.88)
	Obs.	3,537	3,200	3,272	3,042	3,252	1,934
First stage	Coef.	0.324***	0.355***	0.302***	0.393***	0.327***	0.443***
	s.e.	(0.092)	(0.09)	(0.105)	(0.112)	(0.107)	(0.106)
	Obs.	7,713	7,066	7,130	6,707	7,095	6,720
Polynomial		2	--	2	--	2	--
Bandwidth		--	Optimal	--	Optimal	--	Optimal
Controls		Yes	Yes	Yes	Yes	Yes	Yes

Notes: (1) Robust standard errors in parentheses, clustered at the municipality level (2) The independent variable is the number of big-box openings between t and t-n at the municipality level, instrumented by a dummy that captures the change in the probability of treatment due to the commercial regulation. Each row represents a different regression. (3) Columns (1) and (2) present the results when including all the municipalities that experienced a big-box opening before the regional law was implemented. Columns (3) and (4) show the results of including the municipalities that changed from one side of the threshold to the other during the period of analysis. Columns (5) and (6) and (7) and (8) report the results when using the 3-year lagged population and the 5-year lagged population as running variables respectively. (4) All regressions include region and time fixed effects in order to control for region specific time invariant characteristics and countrywide time shocks. They also include the pre-regulation levels of population, economic activity and education levels, size of the municipality in square kilometres, immigration level, unemployment and importance of the services sector in order to control for trends. (5) *** $p<0.01$, ** $p<0.05$, * $p<0.1$

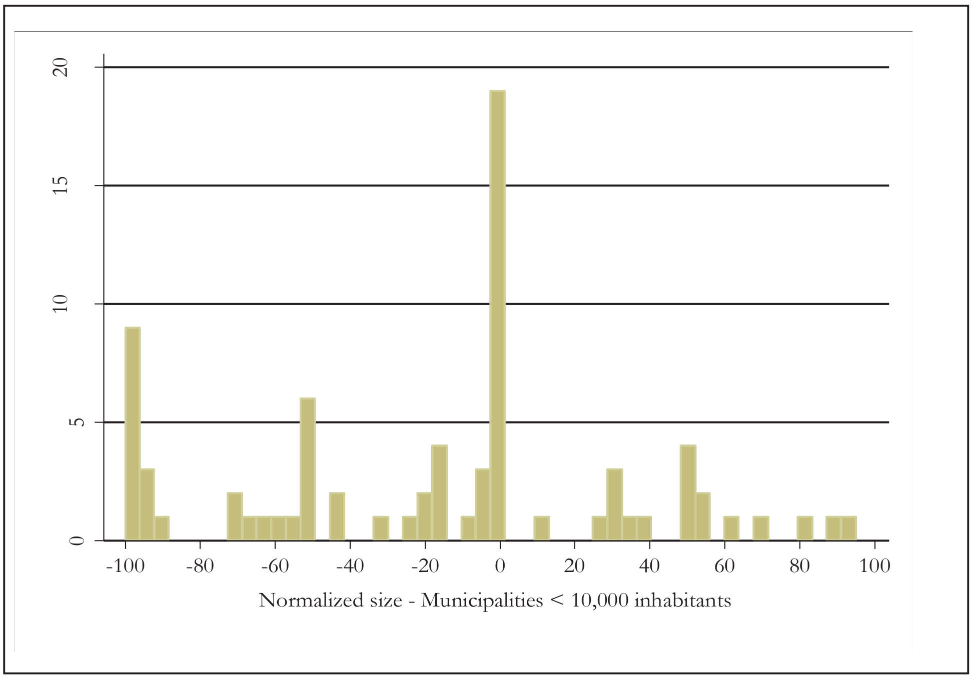
The previous results confirm the negative effect of big-box openings on the number of pre-existing grocery stores. This implies that the commercial regulation restricting the opening of big-box stores may be fulfilling its main goal, namely, the protection of grocery stores. However, we need to evaluate any other indirect effects that this regulation may have. The most straightforward is the impact that the entry of big-boxes could have on employment in the municipality. Typically, grocery stores in Spain are family-owned business that do not usually hire any extra staff. On average the size of such stores is 0.98 employees plus the owner giving an average total of 1.98 jobs per grocery store.[10] Thus, for every grocery store forced to pull down its shutters, 1.98 jobs are lost. If we take the coefficients from our preferred estimation in Table 6, about 14 grocery stores were found to shut down in the four-year period after a big-box opening, which means a municipality loses 27.72 jobs. However, this number needs to be put into perspective, as we have to consider the number of jobs created when a big-box store is opened. On average, a big-box store employs 42 employees.[11] Therefore, the net employment effect would be an increase of around 14.28 jobs. So, even if the commercial regulation is preventing the disappearance of grocery stores, it may also have an indirect negative net effect on *local* employment. These results are consistent with the theoretical predictions and the policy recommendations made in Ushchev *et al.* (2015) where it is claimed that big-box openings tend to hollow out city centres but that the regulation should only be implemented when malls are not efficient enough to capture the whole market.

However, it is important to note that the above results also depend on the exact definition (size in square metres) given to a big-box store. In fact, each region, as observed in Table 1, sets its own limits on what it considers a big-box store to be. Thus, it might be the case that chains seek to bypass the regulation by building stores just below the threshold (in order for the store not to be considered a big-box) and so they can avoid having to apply for a second licence. Indeed, in the case of the UK, Sadun (2015) reports evidence of this actually happening, thus undermining the regulation. This paper has shown that the regulation is positively affecting the regulated municipalities, at least in terms of grocery store closures. Therefore, were we to observe a bunching of stores just

[10] Extracted from the Spanish Ministry of Agriculture's database.
[11] This average is computed using data available in the 2011 Census of Chain Supermarkets, which reports (in some instances) the number of employees in big-box stores. The number has been corroborated by examining information available on the websites of the main chains of big-box stores in Spain.

below the threshold in those municipalities, this would indicate that the previous results are downward-biased. Figure 4 presents the size distribution of chain stores computed using the 2011 Census of Chain Supermarkets dataset. It reports this distribution for municipalities below the 10,000 inhabitant threshold. Given that the regions included in the study have different size definitions for a big-box store, the size axis has been normalised. We observe that, in the regulated municipalities there is, indeed, evidence of bunching just below the threshold, indicating that some chains have tried to avoid the regulation. Thus, this graph presents evidence that, while the previous findings indicate an impact of big-box openings on grocery stores, it may be an underestimate of the real effect, in terms of store closures.

Figure 4: Bunching around the threshold



Note: This figure shows a frequency histogram of the number of big-box openings around the threshold for municipalities smaller than 10,000 inhabitants. The size (in square meters) is normalized according to the criterion of each region in order to consider a store a 'big-box'.

5.2. Heterogeneous effects of big-box openings on grocery store closures

The results reported above describe the average impact of all big-box openings on grocery store closures within the period analysed, regardless of the specific characteristics of the big-box store. In this section, I evaluate whether the effects are driven by the location of the big-box – in the city centre or in the suburbs – or the typology of big-box opened – conventional supermarkets versus discount supermarkets. Note that the total number of big-box openings is 317 (Table 3). Of these, 88 were opened in city centres while 229 were located in the suburbs. Likewise, by typology, 129 correspond to discount supermarkets and 188 to conventional chain stores. The reason for exploring any (possible) geographical effects of big-box openings is that big-box stores opening in locations close to existing grocery stores, i.e., in city centres, might be competing more directly with these small shops and harming them more (Sadun, 2015). On the other hand, it might also be the case that certain complementarities are created between big-box and grocery stores, stimulating demand for non-substitutable products. To this end, I estimate the following equation:

$$\Delta\, \textit{grocery stores}_{it} = \theta_{it} + \varphi_{\mathrm{it}} \cdot \textit{big-box openings}_{it} +$$
$$\mu_{it} \cdot \textit{big-box openings}_{it} \cdot \textit{location}_s + \tau \cdot \textit{location}_s + \sigma_{it} \cdot g\,(P_{i,t\text{-}4}) + + \varrho_t + \pi_r + X'_{it}\vartheta + \epsilon_{it} \quad (3)$$

where $\Delta\, \textit{grocery stores}_{it}$ is the change in the number of grocery stores between *t* and *t-4* aggregated at the municipality level, indicating only the cumulative effect four years after the big-box opening. The variable $\textit{location}_s$ indicates the location of the big-box store. It takes a value equal to one if the big-box opens near the city centre and a value equal to zero if it locates in the suburbs. In the regression, this indicator is interacted with the main explanatory variable and, thus, I can estimate the opening effect allowing for some geographical differences in how big-box openings may affect grocery store closures. The results are presented in the first two columns of Table 8. We observe that there are negative effects of big-box openings in both the city centre and the suburbs on grocery store closures, but that there is no significant difference between the two locations. Thus, it does not seem to be the case that the city centre big-box stores affect grocery stores any differently to the way in which out-of-town big-boxes affect them.

Table 8. The effect of big-box openings on grocery store closures – Heterogeneous effects

		Dependent variable: Change in the number of grocery stores			
		Polynomial regressions			
		(1)	(2)	(3)	(4)
Big-Box openings t,t-4	City Centre (*Location=1*)	-19.22*** (7.04)	-16.09** (6.43)		
	Suburbs (*Location=0*)	-27.09** (12.75)	-20.33* (11.53)		
	Conventional (*Type=1*)			-27.33** (10.71)	-24.42** (10.16)
	Discount (*Type=0*)			-3.50 (8.84)	-1.50 (8.86)
Polynomials		1	2	1	2
Controls		Yes	Yes	Yes	Yes
Observations		4,407	4,407	4,407	4,407

Notes: (1) Robust standard errors in parentheses, clustered at the municipality level (2) The independent variable is the number of big-box openings between t and t-4, instrumented by a dummy that captures the change in the probability of treatment due to the commercial regulation. In columns (1) and (2), this variable is interacted with a dummy variable equal to one if the big-box is opened in (or next to) the city centre and zero if it is opened in the suburbs. In columns (3) and (4) the dummy variable is interacted with a dummy equal to one if the big-box is considered to be a conventional supermarket, i.e. selling all brands and equal to zero if it is a discount big-box, i.e. typically selling their own, lower price brands. (3) All regressions include region and time fixed effects in order to control for region specific time invariant characteristics and countrywide time shocks. They also include the pre-regulation levels of population, economic activity and education levels, size of the municipality in square kilometres, immigration level, unemployment and importance of the services sector in order to control for trends. (4) *** p<0.01, ** p<0.05, * p<0.1

Additionally, I evaluate whether the effects from Table 6 differ depending on the typology of the big-box opened. I divided the sample into two different types of big-boxes: conventional and discount stores. The former are those chains that sell well-known brands, whereas the latter typically sell their own, lower price brands. To evaluate whether there is any differential effect between these two types, the following equation is estimated:

$$\Delta\, grocery\ stores_{it} = \theta_{it} + \varphi_{it} \cdot big\text{-}box\ openings_{it} + \mu_{it} \cdot big\text{-}box\ openings_{it} \cdot type_s + \tau \cdot type_s + \\ + \sigma_{it} \cdot g\,(P_{i,t-4}) + \varrho_t + \pi_r + X'_{it}\vartheta + \epsilon_{it} \quad (4)$$

where $\Delta\,grocery\ stores_{it}$ is again the change in the number of grocery stores between t and $t\text{-}4$. The variable $type_s$ indicates the typology of the big-box store, taking a value equal to one if the big-box is conventional and zero if it is a discount one. The results of interacting this indicator with the variable capturing the big-box opening are presented in the last two columns of Table 8. We see that there is a clear negative and significant effect of big-box openings on grocery store closures when the big-box is conventional. In contrast, discount big-boxes do not seem to have any impact on grocery store closures. These results may be indicating a persistence of consumer preferences. It could be that consumers are used to certain kinds of products and brands and do not easily switch to unknown products even if they can be purchased relatively cheaper in discount big-box stores. Thus, conventional big-box stores may be competing more directly with grocery stores. They sell the same products but in a one-stop shop, which could be more convenient for consumers than having to make the two or more stops typically needed when buying food from grocery stores.

6. Conclusions

The opening of big-box stores has become a political concern in many countries over the last few decades. Their critics claim they create enormous negative externalities in pre-existing market and city structures, exacerbating pollution levels and contributing to the hollowing out of city centres, as grocery stores are forced into closure. Yet, there are those who argue that these stores tend to push prices down and, so, consumers are better off when big-box stores locate to their municipalities. In this paper, I exploit a commercial regulation in Spain, aimed at restricting the entry of big-box stores, to evaluate the extent to which these openings cause grocery stores to close. More specifically, this regulation requires developers seeking to build a big-box store in a municipality with less than 10,000 inhabitants to obtain a second licence from the regional government, in addition to the municipal licence.

Using an RDD analysis, I first tested whether this regulation does in fact prevent developers from establishing big-box stores in regulated municipalities. The findings show that, indeed, non-regulated municipalities experienced 0.3 more openings than regulated municipalities. I then used this jump around the threshold to instrument the effect of big-box openings on grocery store closures. The results suggest that, following the opening of a big-box, the affected municipality gradually loses grocery stores, typically from the city centre, showing some evidence of downtown hollowing out. In fact, four years after the opening, between 20 and 30% of the pre-existing grocery stores have closed down. When evaluating the heterogeneity of these effects, the results seem to show that there are no significant short-run differences between big-box store openings in the city centre and those out-of-town. This may show, at least in the short run, that both downtown and suburb big-boxes act as direct competitors of grocery stores. I performed an additional heterogeneity analysis in which I examined conventional and discount big-box stores separately, where the former are chain stores selling all well-known brands at market prices while the latter typically sell their own, low-price brands. In this case, all the effect could be attributed to the conventional stores, offering some evidence that these shops, which sell the same kind of products as grocery stores but in a one-stop shop, may match consumer preferences better and may also be more convenient, at least in the short run.

The findings reported herein have a number of policy implications. First, the regulation introduced was designed to restrict the entry of big-boxes and as such to prevent grocery stores from closing. This paper has shown that this aim has indeed been met, given that non-regulated municipalities suffered more closures than regulated municipalities. In fact, some bunching of stores below the size threshold was also observed, suggesting that the results may even be underestimating the effects. However, while the regulation may have served its purpose, there may be other indirect effects that need to be taken into consideration but, unfortunately, due to problems of data availability, this paper has been unable to do so. The main concern associated with this policy is the (possible) negative impact it has on employment. However, if the loss of jobs generated by the closure of grocery stores is offset by the employment created by big-box opening, the net employment effect would be positive. Thus, the regulation may be undermining *local* employment instead of protecting it.

References:

Basker, E. (2005): "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion", *Review of Economics and Statistics,* 87(1), 174–183.

Basker, E. (2005): "Selling a cheaper mousetrap: Wal-Mart's effect on retail prices", *Journal of Urban Economics*, 58(2), 203–229.

Basker, E. (2007): "The Causes and Consequences of Wal-Mart's Growth". *The Journal of Economic Perspectives*, 21(3), 177–198.

Basker, E. and Noel, M. (2009): "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry", *Journal of Economics and Management Strategy*, 18(4), 977-1009.

Bertrand, M. and Kramarz F. (2002): "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry," *Quarterly Journal of Economics*, 117(4), 369-413.

Cheshire, P., Hilber, C. and Kaplanis, I. (2015): "Land use regulation and productivity—land matters: evidence from a UK supermarket chain", *Journal of Economic Geography*, 15, 43–73.

Ellickson, P.B. and Grieco, P.L.E. (2013): "Wal-Mart and the geography of grocery retailing", *Journal of Urban Economics*, 75, 1-14.

Forenmy, D., Jofre-Monseny, J. and Sole-Olle, A. (2015): "'Hold that Ghost': Using Notches to Identify Manipulation of Population-Based Grants", CESifo Working Paper No. 5577.

Griffith, R. and H. Harmgart (2008): "Supermarkets and Planning Regulation," *CEPR Discussion Papers No.* 6713.

Haltiwanger, J., Jarmin,R. and Krizan, C.J. (2010): "Mom-and-Pop Meet Big-Box: Complements or Substitutes?" *Journal of Urban Economics* 67(1), 116–134.

Haskel, J., Sadun, R. (2012): "Regulation and UK retailing productivity: evidence from microdata", *Economica*, 79, 425-448.

Imbens, G. and K. Kalyanaraman (2012): "Optimal bandwidth choice for the Regression Discontinuity estimator," *Review of Economic Studies,* 79(3), 933-959.

Imbens, G. and T. Lemieux (2008): "Regression discontinuity designs: a guide to practice," *Journal of Econometrics,* 142, 615–635.

Jia, P. (2008): "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry",*Econometrica* 76(6), 1263–1316.

Lee, D. and Lemieux, T. (2010): "Regression Discontinuity Designs in economics," *Journal of Economic Literature* 48(2) 281-355.

Matea, M.L. and Mora-Sanguinetti, J.S. (2009): “Developments in retail trade regulation in Spain and their macroeconomic implications”, Bank of Spain Working Paper 0908.

McCrary, J. (2008): “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test”. *Journal of Econometrics*, 142(2).

Neumark, D, Zhang, J. and Ciccarella, S. (2008): “The Effects of Wal-Mart on Local Labor Markets”,*Journal of Urban Economics* 63(2), 405–430.

Sadun, R. (2015): "Does Planning Regulation Protect Independent Retailers?". *Review of Economics and Statistics,* 97(5), 983-1001.

Schivardi, F. and Viviano, E. (2011): “Entry Barriers in Retail Trade,” *Economic Journal*, 121(551), 145-170.

Uschev, P., Sloev, I. and Thisse, J.F. (2015): “Do we go shopping downtown or in the ‘burbs?, *Journal of Urban Economics*, 85, 1-15.

Chapter 5

Concluding remarks

Cities present high levels of worker and business productivity thanks to their agglomeration economies, which are usually capitalised in higher wages. Cities are, moreover, the perfect environment for consumption, thanks to their large supply of amenities. However, the density of cities is at the same time responsible for rising congestion costs and higher housing prices. Thus, and in line with the urban economics literature, the equilibrium city size depends on the trade-off between the benefits accrued from these agglomeration forces and the costs associated with larger cities. This thesis contributes to this literature by providing three interesting findings about the economics of city formation and city evolution. First, it reports that city growth is non-random when cities are in their early stages of development, while parallel growth emerges as they become older (chapter 2). Second, it shows how agglomeration economies, and labour market pooling in particular, play an important role in the adaptation of a city's labour market to firm-specific shocks (chapter 3). And, finally, it provides evidence of the fact that the opening of big-box stores affects the number of grocery stores in city centres, which can make them less attractive due to the loss of some of their consumption amenities (chapter 4). This concluding chapter summarizes these main findings and discusses several implications and further extensions.

The second chapter of this thesis inspects one of the mechanisms driving the existence of different cities of different sizes. Using data from US cities, it studies the evolution of city growth throughout the twentieth century. More specifically, the analysis focuses on the role played by the new-born cities created during the decades between 1900 and 2000. By means of parametric and nonparametric methods two main results are obtained. The first finding is that there are differences in city growth rates according to the age of the city. In general, when a city is born it presents a very high growth rate but, as the decades pass, it matures and its growth rate stabilises or even declines.

Second, nonparametric regressions are performed to examine the relationship between the time dimension of growth (the city's age) and the city's initial size. The results confirm that there are deviations from Gibrat's law for the smallest cities of all ages but they are especially important for the youngest ones. In fact, as cities become older, Gibrat's law tends to hold. These results suggest that most of the growth differential across cities is driven by their first decade of existence, which is generally in line with the parametric results. Thus, this chapter contributes to the urban economics literature by shedding light on the mechanisms and timings of urban growth, revealing non-random growth for new-born cities and parallel growth as they become older.

The third chapter of this thesis estimates the real net local employment responses to large manufacturing plant closures as a result of their international relocations. Specifically, it estimates the employment effects of the closure of 45 large manufacturing plants in Spain, which relocated to (mainly) developing countries between 2001 and 2006. Each municipality experiencing a closure is matched to a small set of comparable municipalities in terms of employment level and industry mix in the year 2000. It is found that treatments and controls do not differ in their 1990-2000 (pre-treatment) employment trends either, thereby lending credence to the identification assumption underpinning the differences-in-differences estimates used in this chapter.

The results show that when a plant closes, for each job directly lost in the plant closure, only between 0.3 and 0.6 jobs are actually lost in the local economy, with the adjustment being concentrated in local incumbent firms in the industry having suffered the closure. This finding is, thus, showing evidence of labour market pooling effects. Another implication of these findings is that traditional input-output analyses tend to overstate the net employment losses of large plant closures. In this particular application, the input-output prediction overestimates the negative employment consequences by an order of three. This chapter, contributes to the urban economics literature by being the first attempt to quantify the impact of plant closures on local employment. In addition, it also shows how the existence of agglomeration economies helps the affected areas to partially overcome a firm-specific shock thanks to labour pooling effects.

The fourth chapter of this thesis studies the effects of big-box store openings, usually located in out-of-town sites, on grocery stores, which are typically identified as city centre consumption amenities. Using an RDD analysis and focusing on the food sector, this chapter makes use of a regulation aimed at restricting the entry of big-box stores as the source of exogenous variation. It first tests if this regulation actually prevents developers from opening big-box stores in regulated municipalities. The findings show that, indeed, non-regulated municipalities experience 0.3 more openings than their regulated counterparts. Then, this jump around the threshold is used to instrument the effect of big-box openings on grocery store closures.

The results indicate that, after a big-box opens, the affected municipality gradually loses grocery stores, typically from the city centre, showing evidence of downtown hollowing out. In fact, four years after the opening, between 20% and 30% of pre-existing grocery stores have closed down. Moreover, when evaluating the heterogeneity of these effects, the results seem to show that there are no significant short-run differences between big-box store openings in the city centre and those out-of-town. This indicates that, at least in the short-run, both downtown and suburb big-boxes act as direct competitors of grocery stores. An additional heterogeneity analysis is also performed by splitting the results between conventional and discount big-box stores, where the former are chains selling well-known brands whereas the latter typically sell their own brands at lower prices. In this case, all the effect on grocery stores can be attributed to conventional stores, showing evidence that these shops, which sell the same kind of products as grocery stores but in a one-stop shop, may match consumer preferences better and may also be more convenient for them.

The main contribution of this chapter is to present evidence on the implications and effects of public policies on cities' retail sectors. If the existence of shopping areas in city centres is understood as a consumption amenity for the residents, the fact that these amenities disappear following the implementation of certain planning policies can make the affected city less competitive when it comes to attracting new residents. Therefore, although it might be more beneficial for a city's productivity to open a big-box in terms of local consumption amenities, it is likely to have a negative impact on the citizens' quality of life.

Some final considerations are worth making regarding the external validity and policy implications of the findings reported in this thesis. First, the second chapter focuses on events in the twentieth century, a period when the US underwent a marked process of urbanisation. As such, these findings could serve as interesting input for policy makers in developing countries, which are now experiencing their own processes of urbanisation. Indeed, if there is a statistical regularity driving the population growth of cities, dependent on their initial size or age, some investment in these developing countries could be performed strategically.

However, this is not the case of the third and fourth chapters, where the analyses focus on specific shocks to the local economy in Spain in the 2000s. This being the case, the lessons learnt from these studies can only be applied to countries with similar institutions. For instance, in the case of the third chapter, the fact that only between 0.3 and 0.6 jobs are lost in the local economy when a big plant closes is the result of a thick labour market with relatively immobile workers. Therefore, in a country with considerably more labour mobility, the results may well differ.

In the same line, the results of the fourth chapter may also be very different in a country with less tradition of small grocery stores and whose retail sector is dominated by big malls outside the city centres. Indeed, even in countries with similar market structures but different consumer preferences, the regulations preventing the emergence of large store formats could have different implications for the local economy than those reported in this chapter. Therefore, what is required is a general picture of a country's retail sector and consumer preferences before we can make any assumptions about the policy implications of the regulations examined in this study.

References

Abadie, A. and Gardeazabal, J. (2003), "The economic costs of conflict: A case study of the Basque Country", *The American Economic Review*, 93(1), 113-132.

Ahlfeldt, G. (2013): "Urbanity". Discussion Paper No' 0136. Spatial Economics Research Centre, London School of Economics.

Ahlfeldt, G. and Holman, N. (2015): "Distinctively different: A new approach to estimating architectural amenities". Discussion Paper No' 0171. Spatial Economics Research Centre, London School of Economics.

Albouy, D. (2015): "What are cities worth? Land rents, local productivity, and the total value of amenities". *Review of Economics and Statistics*, forthcoming.

Arzaghi, M. and Henderson, J.V. (2008), "Networking Off Madison Avenue", *Review of Economic Studies*, 75(4), 1011-1038.

Basker, E. (2005): "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion", *Review of Economics and Statistics,* 87(1), 174–183.

Basker, E. (2005): "Selling a cheaper mousetrap: Wal-Mart's effect on retail prices", *Journal of Urban Economics*, 58(2), 203–229.

Basker, E. (2007): "The Causes and Consequences of Wal-Mart's Growth". *The Journal of Economic Perspectives*, 21(3), 177–198.

Basker, E. and Noel, M. (2009): "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry", *Journal of Economics and Management Strategy*, 18(4), 977-1009.

Behrens, K. and Robert-Nicoud, F. (2015): "Agglomeration theory with heterogeneous agents". In Duranton, G., Henderson, J.V. and Strange, W. (Eds.), Handbook of Regional and Urban Economics, Vol.5. North-Holland, Amsterdam.

Bertrand, M. and Kramarz F. (2002): "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry," *Quarterly Journal of Economics*, 117(4), 369-413.

Black, D. and Henderson, V. (2003): "Urban Evolution in the USA", *Journal of Economic Geography*, 3, 343-372.

Blanchard, F.S. (1960): "Ghost towns of New England: their ups and downs", ed. Dodd, Mead & Company, New York.

Bleakley, H. and Lin, J. (2012): "Portage and path dependence", *Quarterly Journal of Economics*, 127(2), 587-644.

Bogue, D. (1953): "Population Growth in Standard Metropolitan Areas: 1900-1950", Washington, DC: Housing and Home Finance Agency.

Bosker, M., Brakman S., Garretsen H. and Schramm, M. (2008): "A century of shocks: the evolution of the German city size distribution 1925–1999", *Regional Science and Urban Economics*, 38, 330–347.

Carlino, G.A. and Saiz, A. (2008): "City beautiful". Working Paper 08-22, Federal Reserve Bank of Philadelphia.

Cheshire, P., Hilber, C. and Kaplanis, I. (2015): "Land use regulation and productivity—land matters: evidence from a UK supermarket chain", *Journal of Economic Geography*, 15, 43–73.

Cheshire, P., Hilber, C. and Sanchis-Guarner, R. (2014): "Do English Planning Policies Make Shopping More Sustainable?", London School of Economics, mimeo.

Combes, P.P, Duranton, G. and Gobillon, L. (2011): "The Identification of Agglomeration Economies" *Journal of Economic Geography*, 11, 253-266.

Combes, P.P. and Gobillon, L. (2015): "The empirics of agglomeration economies". In Duranton, G., Henderson, J.V. and Strange, W. (Eds.), Handbook of Regional and Urban Economics, Vol.5. North-Holland, Amsterdam.

Couture, V. (2015): "Valuing the Consumption benefits of Urban Density", University of California, Berkeley, mimeo.

Cuberes, D. (2009): "A model of sequential city growth", *The BE Journal of Macroeconomics,* 9 (1) (Contributions), Article 18.

Cuberes, D. (2011): "Sequential City Growth: Empirical evidence", *Journal of Urban Economics*, 69, 229-239.

D’Costa, S., Overman, H. (2014): “The urban wage growth premium: sorting or learning?”. *Regional Science and Urban Economics*, 48, 168–179.

De la Roca, J., and Puga, D. (2012): “Learning by working in big cities”. Discussion Paper 9243, Centre for Economic Policy Research.

Desmet, K. and Rappaport, J. (2015): “The settlement of the United States, 1800 to 2000: The Long Transition to Gibrat’s Law”. *Journal of Urban Economics*, forthcoming.

Dobkins, L. and Ioannides, Y.M. (2000): “Dynamic evolution of the US city size distribution”, in: Huriot, J.M., Thisse, J.F. (Eds.), The Economics of Cities. Cambridge University Press, Cambridge, pp. 217-260.

Dobkins, L. and Ioannides, Y.M. (2001): “Spatial Interactions among US cities: 1900-1990”, *Regional Science and Urban Economics*, 31, 701-731.

Duranton, G. (2008): “Viewpoint: from cities to productivity and growth in developing countries”. *Canadian Journal of Economics*, 41(3), 689-736.

Duranton, G. (2014): “Growing through cities in developing countries”. World Bank Research Observer.

Duranton, G. and Puga, D. (2004): "Micro-foundations of urban agglomeration economies," Handbook of Regional and Urban Economics, in: J. V. Henderson & J. F. Thisse (ed.), Handbook of Regional and Urban Economics, edition 1, volume 4, chapter 48, 2063-2117.

Duranton, G. (2007): “Urban evolutions: the fast, the slow and the still”, *American Economic Review*, 97(1), 197-221.

Duranton, G. and Puga, D. (2015): “The Growth of Cities”. In Aghion, P. and Durlauf, S.N. (Eds.), Handbook of Economic Growth, vol.2, Elsevier, Amsterdam, 781-853.

Edmiston, K.D. (2004), “The net effects of large plant locations and expansions on county employment”, *Journal of Regional Science*, 44(2), 289-320.

Eeckhout, J. (2004): “Gibrat’s law for (All) Cities”, *American Economic Review*, 94 (5), 1429-1451.

Ellickson, P.B. and Grieco, P.L.E. (2013): “Wal-Mart and the geography of grocery retailing”, *Journal of Urban Economics*, 75, 1-14.

Ellison, G., Glaeser, E.L. and Kerr, W.R. (2010): "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100(3), 1195-1213.

Encuesta sobre Estrategias Empresariales. 2008. Las empresas industriales en 2006. Madrid. Fundación SEPI.

Faggio, G. and Overman, H. (2014), "The effect of public sector employment on local labour markets", *Journal of Urban Economics*, 79, 91-107.

Faggio, G., Silva, O. and Strange, W.C. (2015), "Heterogeneous agglomeration", *Review of Economics and Statistics*, forthcoming.

Forenmy, D., Jofre-Monseny, J. and Sole-Olle, A. (2015): "'Hold that Ghost': Using Notches to Identify Manipulation of Population-Based Grants", CESifo Working Paper No. 5577.

Fox, W.F. and Murray M.N. (2004), "Do economic effects justify the use of fiscal incentives?", *Southern Economic Journal*, 71(1), 78-92.

Fujita, M. and Thisse, J.F. (2002): "Economics of agglomeration: Cities, Industrial Location and Regional Growth". Cambridge University Press (Cambridge).

Giesen K. and Südekum J. (2011): "Zipf's Law for cities in the regions and the country", *Journal of Economic Geography*. 11(4), 667–686.

Giesen K. and Südekum J. (2014): "City Age and City Size", *European Economic Review*, 71,193-208.

Glaeser, E.L (2001), "The economics of location-based tax incentives", Discussion Paper No. 1932. Cambridge, MA: Harvard Institute of Economic Research.

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

Glaeser, E.L. and Shapiro, J. (2003): "Urban Growth in the 1990s: Is City Living Back?", *Journal of Regional Science*, 43 (1), 139-165.

González-Val, R. (2010): "The Evolution of the US City Size Distribution from a long term perspective (1900-2000)", *Journal of Regional Science*, 50 (5), 952-972.

Greenstone, M. and Moretti, E., (2004), "Bidding for industrial plants: Does winning a 'million dollar plant' increase welfare?", *mimeo*.

Greenstone, M., Hornbeck, R. and Moretti, E., (2010): "Identifying agglomeration spillovers: evidence from winners and losers of large plant openings", *Journal of Political Economy*, 118 (3), 536-598.

Griffith, R. and H. Harmgart (2008): "Supermarkets and Planning Regulation," *CEPR Discussion Papers No.* 6713.

Haltiwanger, J., Jarmin,R. and Krizan, C.J. (2010): "Mom-and-Pop Meet Big-Box: Complements or Substitutes?" *Journal of Urban Economics* 67(1), 116–134.

Härdle, W. (1990): "Applied Nonparametric Regression". Econometric Society Monographs. Cambridge, New York and Melbourne: Cambridge University Press.

Haskel, J. and Sadun, R. (2012): "Regulation and UK retailing productivity: evidence from microdata", *Economica*, 79, 425-448.

Henderson, J.V. and Venables, A.J. (2009): "Dynamics of city formation", *Review of Economic Dynamics*, 2, 233-254.

Henderson, J.V. and Wang, H.G. (2007): "Urbanization and city growth: the role of institutions", *Regional Science and Urban Economics*, 37 (3), 283-313.

Hooker, M.A. and Knetter, M.M. (2001), "Measuring the economic effects of military base closures", *Economic Inquiry*, 39(4), 583-598.

Imbens, G. and K. Kalyanaraman (2012): "Optimal bandwidth choice for the Regression Discontinuity estimator," *Review of Economic Studies,* 79(3), 933-959.

Imbens, G. and T. Lemieux (2008): "Regression discontinuity designs: a guide to practice," *Journal of Econometrics,* 142, 615–635.

Ioannides, Y.M. and Overman, H.G. (2003): "Zipf's Law for cities: An empirical examination", *Regional Science and Urban Economics*, 33, 127-137.

Ioannides, Y. and Skouras, S. (2013): "US city size distribution: Robustly Pareto, but only in the tail". *Journal of Urban Economics*, 73(1), 18-29.

Jia, P. (2008): “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry”,*Econometrica* 76(6), 1263–1316.

Jofre-Monseny, J. (2009), “The scope of agglomeration economies: Evidence from Catalonia”, *Papers in Regional Science*, vol. 88(3), 575-590.

Jofre-Monseny, J., Marín-López, R. and Viladecans-Marsal, E. (2011): “The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms”, *Journal of Urban Economics*, 70, 61-74.

Jofre-Monseny, J., Marín-López, R. and Viladecans-Marsal, E. (2014): “The determinants of localization and urbanization economies: Evidence from the location of new firms in Spain”, *Journal of Regional Science,* vol. 54, 313-337.

Kim, S. and Margo, R.A. (2004): “Historical perspectives on US economic geography”, in Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Urban and Regional Economics, Vol 4. Amsterdam: Elsevier Science, North-Holland, Chapter 66, pp.2982-3019.

Krugman, P. (1991), *Geography and Trade*, MIT Press, Boston.

Lee, D. and Lemieux, T. (2010): “Regression Discontinuity Designs in economics,” *Journal of Economic Literature* 48(2) 281-355.

Moretti, E. (2010), “Local Multipliers”, *American Economic Review*, 100(2), 373-377.

Marshall, A., (1890): “Principles of Economics”. Macmillan, London.

Matea, M.L. and Mora-Sanguinetti, J.S. (2009): “Developments in retail trade regulation in Spain and their macroeconomic implications”, Bank of Spain Working Paper 0908.

McCrary, J. (2008): “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test”. *Journal of Econometrics*, 142(2).

Michaels, G., Rauch, F. and Redding, S.J. (2012): “Urbanization and structural transformation”, *Quarterly Journal of Economics*, 127(2), 535-586.

Mitchener, K. and McLean, J. (2003), The productivity of US states since 1880, *Journal of Economic Growth*, 8 (1), 73-114.

Myro, R. and Fernández-Otheo, C.M. (2008), "Profitability of Spanish foreign direct investment", *Journal GCG Georgetown University-Universia*, 2(3), 76-99.

Neumark, D, Zhang, J. and Ciccarella, S. (2008): "The Effects of Wal-Mart on Local Labor Markets",*Journal of Urban Economics* 63(2), 405–430.

Overman, H. and Puga, D. (2010), "Labor pooling as a source of agglomeration: An empirical investigation", in Glaeser, E. (Ed.), *Agglomeration Economics*, Chicago University Press, Chicago.

Poppert, P.E. and Herzog Jr., H.W. (2003), "Force reduction, base closure, and the indirect effects of military installations on local employment growth", *Journal of Regional Science*, 43(3), 459-482.

Rappaport, J. (2007): "Moving to nice weather", *Regional Science and Urban Economics*, 37(3), 375-398.

Rappaport, J. and Sachs, J. (2001): "The US as a coastal nation", *Journal of Economic Growth*, 8 (1), 5-46.

Roback, J., (1982): "Wages, rents and the quality of life", *Journal of Political Economy*, 90, 1257–1278.

Rosenthal, S.S. and Strange, W.C. (2003): "Geography, industrial organization, and agglomeration", *Review of Economics and Statistics*, 85(2), 377-393.

Rossi-Hansberg, E. and Wright, M.L.J. (2007): "Urban structure and growth". *Review of Economic Studies*, 74(2), 597-624.

Sadun, R. (2015): "Does Planning Regulation Protect Independent Retailers?". *Review of Economics and Statistics,* 97(5), 983-1001.

Schivardi, F. and Viviano, E. (2011): "Entry Barriers in Retail Trade," *Economic Journal*, 121(551), 145-170.

Shapiro, J. M. (2006): "Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital", *Review of Economics and Statistics*, 88(2), 324-335.

Uschev, P., Sloev, I. and Thisse, J.F. (2015): "Do we go shopping downtown or in the 'burbs?, *Journal of Urban Economics*, 85, 1-15.

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

Wilson, J.D. (1999), "Theories of tax competition", *National Tax Journal*, 52(2), 269-304.